SORT 46 (2) July-December 2022, 189-216 DOI: 10.2436/20.8080.02.123

Compositional combination and selection of forecasters

Antonio Martín Arroyo* and Aránzazu de Juan Fernández†

Abstract

The Split-Then-Combine approach has previously been used to generate the weights of forecasts in a combination in the Euclidean space. This paper extends this approach to combine forecasts inside the simplex space, the sample space of positive weights adding up to one. As it turns out, the simplicial statistic given by the sample centre compares favourably against the fixed-weight, average forecast. Besides, we also develop a Combination-After-Selection method to get rid of redundant forecasters. We apply these approaches to make out-of-sample one-step ahead combinations and subcombinations of forecasts for several economic variables. This methodology is particularly useful when the sample size is smaller than the number of forecasts, a case where other methods (e.g., ordinary least squares or principal component analysis) are not applicable.

MSC: *54C65, 91B84, 62P20.*

Keywords: *Aitchison geometry, Combination-After-Selection, Dimensionality problem, Simplex, Split-Then-Combine.*

Competing interest: The authors do not present any conflicts of interests.

Declarations

Funding: The second author acknowledges financial support from the Spanish Government Project PID2019-108079GBC22.

* a.martin@uam.es. Universidad Autónoma de Madrid (UAM): Economía Cuantitativa, E-III-306 , Avda. Francisco Tomás y Valiente 5, 28049 Madrid (Spain). ORCID ID: 0000-0002-9071-5231

† Corresponding author: aranzazu.dejuan@uam.es. Universidad Autónoma de Madrid (UAM): Economía Cuantitativa, E-III-307, Avda. Francisco Tomás y Valiente 5, 28049 Madrid (Spain). ORCID ID: 0000-0002-7171-7434

Received: October 2021
Accepted: October 2022

Availability of data and material: The data used in this paper are obtained from a public source and are available from the authors upon request.
Code availability: The application developed in this paper was estimated using R code. The codes are available from the authors upon request.
Authors's contribution: Both authors have worked in the manuscript in a similar way. Both have prepared the technical part of the manuscript. Both authors have developed the application and both authors have contributed to the writing of the paper.

1. Introduction

There is a vast body of literature advocating the usefulness of forecast combination methods, both theoretically and empirically. A simple and widely used one consists of simply attributing equal weights to the individual predictions (neutral element of a weight combination in the simplex.) However, the idea of determining the optimal weight combination that minimize some objective criterion (e.g., the mean square forecast error) is more appealing (Conflitti, De Mol and Giannone, 2015). This is the case of the varying-weight sample centre g (Aitchison, 1982). It is the central tendency of our weight combinational sample and it is defined as the weight combination whose components are the sample geometric mean of the weights of each forecaster. Unlike ordinary least squares (OLS) and principal component analysis (PCA), this is a viable strategy even when the number of forecasters to be combined gets large, provided that we constrain weights to be positive and add up to one. Hence, the optimal combination problem reduces to a (possibly high-dimensional) constrained least-squares regression problem, where the complete covariance structure between weights is taken into account. Indeed, this enforces an implicit shrinkage on weights which ensures a reasonable out-of-sample performance of the combined forecasters. This problem turns out to be analogous to the determination of no-short minimum variance Markowitz portfolios, which are a special case of a larger family of sparse and stable portfolios that are derived through a constrained "lasso" regression problem (Tibshirani, 1996), where the weight vector has a unit L1-norm. This type of constraint is known to enforce sparsity, namely the presence of zeros in the weight vector, which means that only a small number of forecasters will be selected (subcombinations in our Combination After Selection (CAS) approach).

Forecasters have access to a wide variety of information and forecasting techniques, thus leading to a considerable degree of heterogeneity or redundancy among them. A weighted average forecast is expected to perform better than individual ones because this way we can diversify away idyosincratic forecast misspecifications, thus reducing the variance of the forecast. The simplest example is the (fixed weight) arithmetic average. More sophisticated methods that make use of varying weights usually do not improve the average in empirical applications because of the instability of the estimated weights (a problem known as *forecast combination puzzle,* Stock and Watson, 2004); in particular, when an increasing number of forecasters requires us to estimate an increasing number of weights (a problem known as *the curse of dimensionality*). The forecast combination puzzle has been considered by Smith and Wallis (2009), who pointed out that the failure of more sophisticated combination methods is due to the estimation of the combining weights.

With forecast (or model)-specific combinations, forecasting is often based on predicting the same variable independently by forecasters. However, analysts who are interested in forecasting a variable from a specific source should not ignore the forecasts from other competing sources. A forecast combination is in fact influenced by all the forecasts; hence, the relationship among individual forecasters is lost when forecasts are independently analysed. Only a few methods have been suggested that incorporate dependence between forecasters. Multivariate models could incorporate dependence between forecasters if we knew such a dependence. Alternatively, we can engage straightaway with weight distributions based on given individual forecast errors, as dependence between weights can be incorporated directly, thus increasing forecast accuracy.

One important difference between modeling forecast-specific combinations and weight distributions is that weights are directly dependent on each other on an aggregated level. The awareness of problems, however, arising from the use of standard statistical methods with proportions (weights) dates back to Pearson, (1897); that is, spurious effects on their covariance structure. In particular, each row or column of the variance matrix of a vector of weights sums up to zero. Given that the variances are always positive, this implies that some covariances are forced towards negative values (Chayes, 1960).

Independent modelling and forecasting with forecast-specific combinations are not only unattractive since they ignore dependence patterns among (relative) weights, but also because weights often fail to be coherent in the sense of the erratic way in which the covariance associated with two specific weights can fluctuate in sign as we move from a full combination to lower and lower dimensional subcombinations. In fact, there is no relationship between the variance matrix of a subcombination and that of the full combination. Besides, variances may display different rank orderings as we form subcombinations, which could lead to implausible forecasters.

Also, avoided forecasts in a subcombination will result in an increase of weights for some other forecasters. By definition of a weight combination, not only is there a common element in the numerator and denominator of each weight, but also all weights have a common denominator. Avoided forecasters in a subcombination thus affect both the numerator and the denominator, and the dependence between forecasters is therefore not as easy to predict.

Moreover, all combinations are subcombinations of a larger one. Since the covariance between two weights depends on which other forecasters are reported in the dataset, there is no guarantee that a plot of a subcombination exhibits similar or even compatible patterns with the plot of the original dataset, even if the forecasters not included in the subcombination are irrelevant (redundant).

There is thus incoherence of the correlation between weights as a measure of dependence. Note, however, that the ratio of two weights remains unchanged when we move from a full combination to a subcombination. Therefore, as long as we work with scale invariant functions (i.e., ratios), we shall be subcombinationally coherent (Aitchison, 1986).

Since standard descriptive statistics (e.g., arithmetic mean and standard deviation) are not informative with combinations, in this paper we propose a time-varying method to combine, select, and recombine forecasters based on Aitchison (1982, 1986), who characterized compositions as vectors having a relative scale and identified its sample space with the simplex. More crucial than the constraining property of compositional data is the scale-invariant property of this kind of data. Indeed, when we are considering only few forecasters of a full combination we are not working with constrained data but our data are still compositional. This approach has been successfully applied to various fields; see, for instance, Aitchison (1986) Billheimer, Guttorp and Fagan (2001), Egozcue and Pawlowsky-Glahn (2005, 2019), Coenders and Ferrer-Rosell (2020) and Greenacre (2021). Software packages available now to deal with compositional data are, for example, Van den Boogaart and Tolosana-Delgado (2013) and Filzmoser, Hron and Templ (2018). To our knowledge, it has not been applied to combinations of forecasts. Compositional Data Analysis (CoDA) is a well-established set of statistical methods for the analyses of compositional data, that enables coherent modelling of weight combinations where dependences between weights are explicitly modeled, so a relative improvement in the weight for one forecaster leads to a decline in the relative weight for the remaining ones.

Any statement about weight combinations can be reformulated in terms of (centred) logratios and viceversa (one-to-one transformation). Data are projected into multivariate real space, opening up all available standard multivariate techniques. Moreover, weight combinations may be represented by orthonormal coordinates (Mateu-Figueras, Pawlowsky-Glahn and Egozcue (2011); Pawlowsky-Glahn and Buccianti (2011), Pawlowsky-Glahn, Egozcue and Tolosana-Delgado (2015)) in a real Euclidean space that can be interpreted in themselves or from their representation in the simplex (Aitchison geometry).

The analysis that is presented in this paper uses the Split-Then-Combine (STC) approach of Arroyo and de Juan Fernández (2014) to generate the weights of a combination. Because they are restricted to be positive and sum up to one, we propose the sample centre g of our weight combinational sample as our basic simplicial combination vector. To get a subcombination, we develop a Combination-After-Selection (CAS) procedure to recombine the best subset of forecasters.

The paper is organized as follows: the next section describes the STC approach both in the Euclidean and simplex spaces. Then, we explain the CAS strategy. In the empirical application, in Section 4, we pull out information provided by panels of quarterly periodicity from a pool of expert forecasters for the US macroeconomy over the period 1991–2018. Forecast accuracy of simplicial combinations are compared with the uniform benchmark arithmetic average.The results obtained with CAS are clearly better than the obtained with the other combinations. Finally, some concluding remarks complete the paper.

2. The Split-Then-Combine (STC) approach

Arroyo and de Juan (2014) proposed the Split-Then-Combine approach to generate combinations for panel m across J forecasters $\left(\widehat{Y}_{t,j}\right)$, $j = 1,2,...,J$, along $t = 1,2,...,T$ periods using the expression:

$$\widehat{Y}_t^{(m)} = \omega_{t,1}^{(m)}\widehat{Y}_{t,1}^{(m)} + \omega_{t,2}^{(m)}\widehat{Y}_{t,2}^{(m)} + ... + \omega_{t,J}^{(m)}\widehat{Y}_{t,J}^{(m)},$$

where the weights $\omega_{t,j}^{(m)}$ vary in two dimensions: (1) from one period to the next; and (2) from one panel to another. We have one panel for each season. Each panel is a tableau of T rows (years) and J columns (forecasters). Each row is then closed to a positive weight combination with weights adding up to one. Finally, this weight combination is used to weight forecasters in out-of-sample forecasting exercises. For example, if we are working with monthly data, we will have 12 panels, one for each month; if we work with quarterly data, we will have four panels, one for each quarter. Panels take into account the different behaviour of the time series among seasons, but STC can also be applied to time series with lower frequency than quarterly or monthly data.[1]

The weights of the STC approach must satisfy two restrictions: be positive and sum up to one; the latter is to avoid biased combinations if individual forecasts are unbiased. Arroyo and de Juan (2014) developed the STC in the Euclidean Space. Here, we also study the STC in the so-called Aitchison geometry (Billheimer et al., 2001, and Pawlowsky-Glahn and Egozcue, 2001).

In order to see the differences between both methods, we first briefly review the STC approach in the Eucidean space; then, we expand the STC approach to the simplex space.

2.1. The STC approach in the Euclidean Space

Table 1 shows how the STC approach works in the Euclidean space. Columns 2 to 5 show the forecasts of the variable of interest for panel m. Each element of this column represents the forecast of each forecaster for a given period. For instance, $\widehat{Y}_{2,1}^{(m)}$ is the forecast of a variable of interest Y from forecaster 2 for period 1 in panel m. The 6th column shows the cross average by period for the J forecasters; that is, $\overline{\widehat{Y}}_{J,1}^{(m)}$ is the average of the J forecasters for the first forecasting period. The 6th row shows the time average by forecaster, that is, $\overline{\widehat{Y}}_{1,T_1}^{(m)}$ is the average over time of all the forecasts from the first forecaster. Column 7 reports the actual, observed data of the variable and the 7th row shows the precision of each forecast average with respect to the overall average $\left(\overline{\overline{\widehat{Y}}}_{J,T_1}^{(m)}\right)$. This measure is used to construct the weights ω that will be assigned to each forecast in the STC approach in the Euclidean space.

[1] See Bujosa-Brun et al. (2020) for an application of the STC approach to annual data with only one panel.

Table 1. *STC approach in the Euclidean Space.*

Panel m	1	2	...	J	$\overline{\widehat{Y}}^{(m)}_{J,t}$	Real data
1	$\widehat{Y}^{(m)}_{1,1}$	$\widehat{Y}^{(m)}_{2,1}$	...	$\widehat{Y}^{(m)}_{J,1}$	$\overline{\widehat{Y}}^{(m)}_{J,1}$	$Y^{(m)}_{1}$
2	$\widehat{Y}^{(m)}_{1,2}$	$\widehat{Y}^{(m)}_{2,2}$	...	$\widehat{Y}^{(m)}_{J,2}$	$\overline{\widehat{Y}}^{(m)}_{J,2}$	$Y^{(m)}_{2}$
...	...	...	...	...	...	...
T_1	$\widehat{Y}^{(m)}_{1,T_1}$	$\widehat{Y}^{(m)}_{2,T_1}$	...	$\widehat{Y}^{(m)}_{J,T_1}$	$\overline{\widehat{Y}}^{(m)}_{J,T_1}$	$Y^{(m)}_{T_1}$
$\overline{\widehat{Y}}_{j,T_1}$	$\overline{\widehat{Y}}_{1,T_1}$	$\overline{\widehat{Y}}_{2,T_1}$	...	$\overline{\widehat{Y}}_{J,T_1}$	$\overline{\overline{\widehat{Y}}}^{(m)}_{J,T_1}$	$\overline{Y}^{(m)}_{T_1}$
Fixed	$\left(\overline{\widehat{Y}}^{(m)}_{1,T_1} - \overline{\overline{\widehat{Y}}}^{(m)}_{J,T_1}\right)^{-2}$	$\left(\overline{\widehat{Y}}^{(m)}_{2,T_1} - \overline{\overline{\widehat{Y}}}^{(m)}_{J,T_1}\right)^{-2}$	...	$\left(\overline{\widehat{Y}}^{(m)}_{J,T_1} - \overline{\overline{\widehat{Y}}}^{(m)}_{J,T_1}\right)^{-2}$		

The STC weights ω are then computed with the information up to time T_1 for each panel using the precision accuracy of each forecaster based on the normalized average squared forecast error:

$$\omega^{(m)}_{j,T_1} = \frac{\left(\overline{\widehat{Y}}^{(m)}_{j,T_1} - \overline{\overline{\widehat{Y}}}^{(m)}_{J,T_1}\right)^{-2}}{\sum_{j=1}^{J}\left(\overline{\widehat{Y}}^{(m)}_{j,T_1} - \overline{\overline{\widehat{Y}}}^{(m)}_{J,T_1}\right)^{-2}}.$$

From these weights, we then form the STC combination in $T_1 + 1$ for panel m:

$$\widehat{Y}^{(m)}_{T_1+1} = \omega^{(m)}_{1,T_1}\widehat{Y}^{(m)}_{1,T_1+1} + \omega^{(m)}_{2,T_1}\widehat{Y}^{(m)}_{2,T_1+1} + ... + \omega^{(m)}_{J,T_1}\widehat{Y}^{(m)}_{J,T_1+1}.$$

This expression must be computed for each panel, $m = 1, 2, ..., M$. These weights satisfy two restrictions: they are positive and add up to one. Once we get forecasts at $T_1 + 1$, we re-compute the weights by rolling over another one-step-ahead combination for $T_1 + 2$, and so on, always keeping the same two restrictions.

2.2. Difficulties with the weight combinations

Standard descriptive statistics are not informative with weight combinations. In particular, the arithmetic mean and the variance of individual weights do not fit the Aitchison geometry as value of central tendency and measure of dispersion. These statistics are defined in the framework of Euclidean geometry in real space, which is not a sensible geometry for weights. Therefore, it is necessary to introduce alternatives. They are found in the concept of sample centre (Aitchison, 1997), variation matrix, and total variance (Aitchison, 1986).

The constraints of constant unit sum and relative meaning of the forecasters' weights have important implications for their statistical analysis, thus rendering direct application of multivariate statistical methods misleading or spurious when applied to combinations for various reasons: (see Chayes, 1960 and Barceló-Vidal and Martín-Fernández, 2016).

1. **Nonnormality:** due to the bounded range of values between 0 and 1, instead of $-\infty$ and $+\infty$.

2. **Spurious correlation**

3. **Singularity**: Euclidean (i.e., raw) variance matrices of random weights are always singular due to the constant sum constraint. A classical way to get rid of singularity is to erase one weight, but results will depend on which one is erased, not being an operation that is permutation invariant.

4. **Negative bias**: Some of the Euclidean covariances are forced towards negative values. Hence Euclidean correlations are not free to range over the usual interval $(-1, 1)$ subject only to the non-negative definiteness of the variance matrix.

5. **Null-correlation:** With negative bias, what is the meaning of zero correlation between two components of a combination?

6. **Subcombinational Incoherence:** There is no relationship between the Euclidean variance matrix of a subcombination and that of the full combination. Besides, variances may display different and unrelatable rank orderings as we form subcombinations. Note, however, that the ratio of two components remains unchanged when we move from full combination to a subcombination so that as long as we work with scale invariant functions (i.e., ratios), we shall be subcombinationally coherent.

7. **Nonsense of scatterplots** for pairs of forecasters: Since the raw covariance between two weights depends on which other forecasters are reported in the dataset (all combinations are subcombinations of a larger one), there is no guarantee that the Euclidean plot of a closed subcomposition of forecasters exhibits similar or even compatible patterns with the Euclidean plot of the original dataset, even if the forecasters not included in the subcomposition are irrelevant. Thus, a regression line drawn in such a plot cannot be trusted.

8. Finally, the construction of a combination from a vector of Euclidean amounts is a constraining closing operation similar to that of the construction of a vector of subcombinations from the related combination. We may therefore expect the same difficulty in relating variance-covariance matrices of weights in the simplex and those in the the Euclidean space.

Weight combinations are multivariate observations carrying relative information: those following the principle of scale invariance, typically being represented in proportions and percentages. In other words, for combinations the relevant information is contained in (log-)ratios. Combinations thus need an own set of statistical methods and should not be treated with statistical methods made for interval scale data. Instead, combinations should be always treated in a log-ratio-transformed scale. It is quite evident that our dataset can only be combinational if it has at least two forecasters. Otherwise, we cannot speak of a weight in a unit total. That implies a substantial difference between combinational data and other multivariate datasets. Most multivariate analysis begin with a univariate analysis of the individual variables (the marginals), whereas each marginal forecaster of a combinational dataset has no meaning on itself, isolated from the remaining forecasters. One combinational dataset should only use proportional weight values. Therefore, results on a subset of forecasters (subcombination) do not depend on the presence or absence of other irrelevant forecasters in the dataset (subcompositional coherence).

3. The STC approach in the Simplex Space and Combination-after-Selection (CAS)

Traditional decomposition techniques provide inconsistent results when applied to compositional data as they do not recognize the implicit constraints of summing to a constant (Aitchison, 1982, 1986): mathematically, compositional data lie in the bounded space of the simplex while traditional decomposition techniques are defined for data in the real space. Aitchison (1986, pp.79) showed that by making log-ratio transformations it is possible to express compositional data in the real space where the data can be analysed with conventional models and then transformed back into the simplex. For instance, the Aitchison inner product, defined in tems of logratios, turns out to be equivalent to the Euclidean inner product in terms of centred logratios. We make use of the centred log-ratio transformation to express the weights in the real space. The clr transformation takes the logarithm of the ratio of each weight divided by the geometric mean of all weights. This transformation maintains the initial constraint in the weights as its elements sum to 0 by construction but resulting values are real. The inverse clr transformation takes the data back to the simplex with the closure operator $\mathcal{C}$ that divides the exponential of each clr entry by the sum of all entries.

Consider a $T \times J$ panel $\widehat{\mathbb{Y}}$ of T out-of-sample forecasts $\widehat{Y}_{t,j}$ produced over time by J forecasters on some variable of interest Y_t, and $\mathbb{A}$ be its related panel of prediction accuracies $a_{t,j} \equiv \left(\widehat{Y}_{t,j} - Y_t\right)^{-2} \in \mathbb{R}_+$. Then, the matrix

$$\mathbb{W} \equiv \begin{pmatrix} w_{1,1} & \dots & w_{1,J} \\ \dots & \dots & \dots \\ w_{t,1} & \dots & w_{t,J} \\ \dots & \dots & \dots \\ w_{T,1} & \dots & w_{T,J} \end{pmatrix} \equiv \begin{pmatrix} w'_{1\bullet} \\ \dots \\ w'_{t\bullet} \\ \dots \\ w'_{T\bullet} \end{pmatrix} \equiv \begin{pmatrix} w_{\bullet 1} & \dots & w_{\bullet j} & \dots & w_{\bullet J} \end{pmatrix},$$

with weights $w_{t,j} \equiv a_{t,j} / \sum\limits_{j=1}^{J} a_{tj}$ represents T combination vectors $w_{1\bullet}$, ..., $w_{T\bullet}$ such that $w_{t,j} > 0$ for all t and j, and $\sum\limits_{j=1}^{J} w_{t,j} = 1$ for all t. Thus, $w'_{t\bullet}$ is just a $1 \times J$ point in a simplex space $\mathbb{S}^{J-1}$ of positive weights adding up to one of dimension $J - 1$ [2]. The function $\mathcal{C} : \mathbb{R}_{+}^{J} \mapsto \mathbb{S}^{J-1}$ that transforms a vector of precisions $a_{t\bullet} \in \mathbb{R}_{+}^{J}$ into a vector of weights $w_{t\bullet} \in \mathbb{S}^{J-1}$ is called a closure transformation $w_{t\bullet} = \mathcal{C}(a_{t\bullet})$. Since this operator cancels out any constant, $\mathcal{C}(ca_{t\bullet}) = \mathcal{C}(a_{t\bullet})$, it is scale invariant. Hence, we just need to work with scale invariant functions (e.g., ratios or logratios). The ratio of two weights remains unchanged when we move from a full combination to a subcombination; that is, $a_{t,i}/a_{t,j} = w_{t,i}/w_{t,j} = s_{t,i}/s_{t,j}$ for all t. Hence, as long as we work with ratios or logratios, we guarantee scale invariance. Therefore, we only consider relative precision among forecasters: each weight in a combination vector has no meaning on itself isolated from the others. Every statement about vectors in $\mathbb{S}^{J-1}$ will be fully expressed in terms of logratios in $\mathbb{R}_{+}^{J-1}$ with inferences transformed back from $\mathbb{R}_{+}^{J-1}$ into combinational statements in $\mathbb{S}^{J-1}$. The STC sample centre g of $\mathbb{W}$ is defined as:

$$g = \mathcal{C}\left(\prod\nolimits_{t=1}^{T} w_{t,1}^{1/T}, ..., \prod\nolimits_{t=1}^{T} w_{t,J}^{1/T}\right) \equiv \mathcal{C}\left(g\left(w_{\bullet 1}\right), ..., g\left(w_{\bullet J}\right)\right). \quad (1)$$

That is, the point in the simplex given by the closure of the geometric averages of weights over time. It can also be viewed as the inverse function of the clr isomorphic transformation applied to the time average of the sample forecasters' weights (Pawlowsky-Glahn et al., 2015). Note that in this definition, the geometric mean is considered column-wise (i.e., by forecasters), while in the clr transformation the geometric mean is considered row-wise (i.e., by samples).

The centred logratio transformation $\texttt{clr} : \mathbb{S}^{J-1} \mapsto \mathbb{R}$, for each $t = 1, ..., T$,

$$x_{t,j} = \texttt{clr}\left(w_{t,j}\right) := \ln w_{t,j} - \frac{1}{J} \sum_{j=1}^{J} \ln w_{t,j} = \ln \frac{w_{t,j}}{\prod_{j=1}^{J} w_{t,j}^{1/J}} \equiv \ln \frac{w_{t,j}}{g\left(w_{t\bullet}\right)}, j = 1, ..., J, \quad (2)$$

where $g(w_{t\bullet})$ is the geometric average of the J weights for the t^{th} observation. This function may be interpreted as a bijection $\mathbb{S}^{J-1} \leftrightarrow \mathbb{H}^{J-1}$ between $\mathbb{S}^{J-1}$ and a vector subspace of $\mathbb{R}_{+}^{J}$ defined by the expression $\mathbb{H}^{J-1} := \left\{ x_{t\bullet} \in \mathbb{R}_{+}^{J} : \sum_{j=1}^{J} x_{t,j} = 0 \right\}$, orthogonal to the vector of ones. The inverse *clr* transformation is then defined by

$$\texttt{clrInv}\left(x_{t\bullet}\right) := \mathcal{C}\left(\exp x_{t\bullet}\right) = \mathcal{C}\left(w_{t\bullet}/g\left(w_{t\bullet}\right)\right) = \mathcal{C}\left(w_{t\bullet}\right) = w_{t\bullet} \in \mathbb{S}^{J-1}, \quad (3)$$

that is, $\texttt{clrInv}$ allows us to go from $\mathbb{R}_{+}^{J-1}$ back to $\mathbb{S}^{J-1}$.

[2]Although in most CoDa papers the superscript of the simplex space is the number of parts, we prefer to emphasize its dimension which, due to the constraint, is $J - 1$. This is in line with the dimension of an isomorphic subspace of the real space isometric with the simplex.

The CAS subcombination is defined as $\mathcal{C}(w_1,...,w_I) = (s_1,...,s_I) \in \mathbb{S}^{I-1}$ inside a simplex of a lower dimension $I-1$ so that $s_1 > 0,...,s_I > 0$ and $s_1 + ... + s_I = 1$. Sometimes, especially when $J >> T$, we perform another subsequent selection by choosing those forecasters inside the previous CAS selection.

A selected CAS subcombination $\mathcal{CS} : \mathbb{S}^{J-1} \mapsto \mathbb{S}^{I-1}$ will be viewed as taking place in two stages: a selection of $I < J$ forecasters by a selecting $I \times J$ matrix $\mathbb{S}$, followed by its closure,

$$\mathcal{CS}(g) = \mathcal{C}(\mathbb{S}g) := \frac{(w_1,...,w_I)'}{w_1 + ... + w_I} = (s_1,...,s_I)'. \tag{4}$$

For $I = 3$, the CAS subcombination can be represented in a ternary diagram by barycentric coordinates (height of the point over the side of the triangle opposite to it). Similarly, for $I = 4$, it can be represented by a tetrahedron where each possible 3-forecast subcombination vector is found by projecting every 4-forecast vector onto the side opposite to the vertex corresponding to the removed forecasters.

The performance of CAS is good just because we get rid of redundant forecasters (curse of dimensionality), thus increasing the forecast accuracy of simplicial statistics in a simplex of a lower dimension (sometimes just a tetrahedron $J = 4$).

We have also carried out Q-mode clustering (Filzmoser et al., 2018) and biplot (Gabriel, 1971) analyses. The main goal is to achieve highly homogeneous clusters of forecasters' weights; i.e., the weights within a cluster should be very similar to each other. On the other hand, different clusters should be dissimilar, because otherwise they should have been merged into one cluster. The variation matrix Υ with elements given by the sample variance over time, $\Upsilon_{i,j} \equiv var\left(\ln \frac{w_{\bullet i}}{w_{\bullet j}}\right)$, with diagonal elements all 0, will be used to define the total variation in $\mathbb{W}$ as $\upsilon^2 := \sum_{i=1}^{J-1} \sum_{j=i+1}^{J} \Upsilon_{i,j}$. Then, υ will be a proper measure of distance among forecasters in cluster analysis, with limit cases of perfect association ($\upsilon = 0$) to perfect independence ($\upsilon = +\infty$). The variation matrix (Aitchison, 1986, or its normalized version) is suitable to express the association between weights. Low values express a high association, and all ratios in a sample are nearly perfectly proportional to each other, while large values express that the ratios are very different from each other. A measure of global dispersion of the weight combinational sample is the total variance (sum of all components of the variation matrix divided by 2J), which turns out to be the time average squared Aitchison distance of each weight combination to the sample centre, also called metric variance (Pawlowsky-Glahn & Egozcue, 2001).

The CAS approach that selects forecasters from the sample centre g of $\mathbb{W}$ can be summarized in the following steps:

1. Given a $T \times J$ table $\widehat{\mathbb{Y}}$ of J forecasters over T time periods in a given season (month or quarter in our cases), compute the related $T \times J$ table $\mathbb{A}$ of $1 \times J$ vectors $a'_{t\bullet}$ of prediction accuracies for each time period $t \in [1,T]$.

2. Convert $\mathbb{A}$ into a $T \times J$ table $\mathbb{W}$ of combination vectors $w'_{t\bullet}$ of weights inside the simplex; that is, weights in each row of $\mathbb{W}$ are positive and add up to one.

3. Calculate the sample centre g of $\mathbb{W}$.

4. Select the CAS subcombination of those forecasters with simplicial weights larger than $1/J$. [3].

5. Repeat steps 1-4 for all panels.

6. Add the next row of out-of-sample accuracy forecasts to the tableau, re-compute the matrix of weights, and update the sample centre and CAS subcombination. Continue this way until the end of the forecast period.

4. Empirical application

We apply the *STC* in the simplex and *CAS* to the variables defined in Table 2, where we include their definition and the samples used to form the combinations of forecasts. Here, we deal with forecasters obtained from the Survey of Professional Forecasters (SPF) from the Federal Reserve Bank of Philadelphia (2018). Blanks in the Survey due to the entry and exit of forecasters are fulfilled following the same strategy as in Poncela et al. (2011), that is, we only consider one-step-ahead forecasts and select only those forecasters without missing data. When there is a missing datum, we use the two-steps-ahead forecast to fill it. Forecasters with more than four consecutive missing data are excluded. For each sample, we only take into account balanced panels. This strategy is also used in Lahiri, Peng and Zhao (2017). Because of the entry and exit of forecasters in the survey, we also analyse different sample sizes, depending on the number of included forecasters. In Table 3, we show, for each variable, the number of forecasters chosen in each subsample. The combinations of forecasters are computed for the periods 2015 to 2018. Note that, in some samples, the number of forecasters is larger than the number of observations, a fact that cannot be treated with other methods (e.g., regression and PCA).

[3] When $J >> T$, we made a first subselection by applying cluster and biplot analyses. Redundant forecasts were defined, with the former, as those whose weights belong to the same cluster; and, with the latter, as those lying on a common line. The sample centre of the remaining weights were then chosen prior to using the CAS strategy.

Table 2. *Definition of the main variables used in the application. Source: Survey of Professional Forecasters documentation. SA = Seasonal Adjusted.*

Variable	Definition	Sample
NGDP	Forecasts for the quarterly level of nominal GDP. SA. billions $	1991 Q1 - 2018 Q4
PGDP	Forecasts for the quaterly level of the chian-weighted GDP price index. SA. Index. Base year 1992	1991 Q1 - 2018 Q4
UNEMP	Forecasts for the quarterly average unemployment rate. SA. % points	1991 Q1 - 2018 Q4
EMP	Forecasts for the quarterly average level of nonfarm payroll employment. SA. Thousands of jobs.	2004 Q1 - 2018 Q4
INDPROD	Forecasts for the quarterly average level of the index of industrial prod. SA. Index.	1991 Q1 - 2018 Q4
HOUSING	Forecasts for the quarterly average level of housing starts. SA. millions.	1991 Q1 - 2018 Q4
TBILL	Forecasts for the quarterly average 3-months Treasury Bill rates. % points	1991 Q1 - 2018 Q4
BOND	Forecasts for the quarterly average level of Moody's Aaa corporate. Bond yield. % points	1991 Q1 - 2018 Q4
RGDP	Forecasts for the quarterly chain-weighted real GDP. SA. annual rate. Base years 1992 - 1995, fixed weighted real GDP	1991 Q1 - 2018 Q4
RCONSUM	Forecasts for the quarterly chain-weighted real personal consumption expenditures. SA, annual rate, base years 1992 - 1995.	1991 Q1 - 2018 Q4
RNRESIN	Forecasts for the quarterly chain-weighted real nonresidential fixed investment. SA. annual rate, base years 1992 - 1995.	1991 Q1 - 2018 Q4
RRESINV	Forecasts for the quarterly chain-weighted real residential fixed investment. SA., annual rate, base years 1992 - 1995	1991 Q1 - 2018 Q4
RFEDGOV	Forecasts for the quarterly chain-weighted real federal government consumption and gross investment. SA, annual rate, base years 1992-95	1991 Q1 - 2018 Q4
RLSGOV	Forecasts for the quarterly level of chain-weighted real state and local government consumption and gross investment. SA. annual rate. base years 1992 - 1995	1991 Q1 - 2018 Q4
CPI	Forecasts for the headline CPI inflation rate. SA, annual rate, % points. Quarterly forecasts are annualized quarter-overquarter percent changes of the quarterly average price index level	1991 Q1 - 2018 Q4

Table 3. *Variables, samples and number of forecasters.*

Variable	Sample (1)		Sample (2)		Sample (3)		Sample (4)		Sample (5)	
	T	J	T	J	T	J	T	J	T	J
NGDP	24	3	20[a)]	6	15[d)]	10	9[g)]	18	5	22
PGDP	24	3	20[a)]	6	15[d)]	10	9[g)]	20	5	25
UNEMP	24	4	20[a)]	6	15[d)]	12	9[g)]	22	5	27
EMP			11	16	10	20	8	22	5	28
INDPROD	24	4	19[b)]	8	15[d)]	12	9[g)]	21	5	26
HOUSING	24	4	19[b)]	10	15[d)]	15	10[f)]	19	5	26
TBILL	24	5	19[b)]	8	15[d)]	11	9[g)]	19	5	24
BOND	24	3	19[b)]	5	14[e)]	7	9[g)]	13	5	17
RRESINV	24	5	20[a)]	9	15[d)]	13	9[g)]	19	5	28
RGDP	24	5	20[a)]	9	15[d)]	14	9[g)]	25	5	31
RCONSUM	24	5	20[a)]	9	16[c)]	13	10[f)]	20	5	29
RNREIN	24	5	20[a)]	9	16[c)]	13	10[f)]	20	5	29
RFEDGOV	24	5	20[a)]	9	16[c)]	13	10[f)]	19	5	28
RLSGOV	24	5	20[a)]	9	16[c)]	13	10[f)]	19	5	28
CPI	24	5	20[a)]	8	16[c)]	12	10[f)]	19	5	29

T = number of periods, J = number of forecasters, Sample (1): 1991 - 2014; Sample (2) a) 1995 - 2014; b) 1996 - 2014; Sample (3) c) 1999 - 2014; d) 2000 - 2014; e) 2001 - 2014; Sample (4) f) 2005 - 2014; g) 2006 - 2014; Sample (5) 2010 - 2014; For the EMP variable the samples are: (1) 2004-2014; (2) 2005-2014; (3) 2007-2014 and (4) 2010-2014

To analyse the prediction accuracy of combinations, we look at four well-known measures: Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Median Absolute Percentage Error (MdAPE). The definitions of the accuracy measures are:

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right);\ RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{n}};$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|;\ MdAPE = Median\left(\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|\right).$$

Although in general these measures produce similar results, there are some differences depending on the type of the combination considered.[4]

[4]In previous studies, we also used Mean Absolute Scaled Error (MASE), and it made no difference with MAPE as to which method generates better results.

We compute four kinds of combinations, three with varying weights: Euclidean space, E_STC, simplex space, S_STC and CAS and the fixed-weight arithmetic average, AVE.

4.1. General results

Table 4. *Summary depending on J and T.*

	AVE	E_STC	S_STC	CAS	$EUCLIDEAN$	$SIMPLEX$	$TOTAL$
$J < T$	215	171	73	289	386	362	758
(%)	(28.3)	(22.6)	(9.6)	(39.5)	(50.4)	(49.1)	(59.9)
$J > T$	115	102	65	227	217	292	509
(%)	(22.6)	(20.0)	(12.8)	(44.6)	(42.6)	(57.4)	(40.2)
$TOTAL$	330	273	138	526	603	664	1266
(%)	(26.0)	(21.6)	(10.9)	(41.5)	(47.6)	(52.4)	

Number of times that accuracy measures favored a combination procedure.Percentages in parenthesis.

We have analyzed 1266 values of accuracy measures. General results are shown in Table 4. According to the type of weights, they favored fixed weights in 330 cases (26.0%) and varying weights in 937 (74.0%). With respect to the latter, 273 (21.6%) favored E_STC, 138 (10.9%) S_STC and 526 (41.5%) CAS. Although the CAS procedure is clearly favored, there is not a clear difference when we compare the results between Euclidean and simplex spaces. In fact, when the combinations are done in a sample with more observations than forecasters, the Euclidean combinations (AVE and E_STC) generate results as good as those obtained with the simplex (50.4% vs 49.1%); but clearly CAS is the best, with a 39.5% of the cases.

When we focus on the results for $J > T$, simplex is better (57.4% vs 42.6%), CAS works very well precisely when some other methods have little to say.

Table 5. *Results for each combination procedure by variable and accuracy criteria. Percentages of beats*

	Mean Error				RMSE				MAPE				MdAPE			
	AVE	*E_STC*	*S_STC*	*CAS*	*AVE*	*E_STC*	*S_STC*	*CAS*	*AVE*	*E_STC*	*S_STC*	*CAS*	*AVE*	*E_STC*	*S_STC*	*CAS*
NGDP	40.0	0.0	10.0	50.0	52.4	0.0	4.8	42.9	30.0	45.0	5.0	20.0	30.0	45.0	5.0	20.0
PGDP	13.6	0.0	45.5	40.9	40.0	0.0	30.0	30.0	5.0	30.0	35.0	30.0	10.0	30.0	30.0	30.0
UNEMP	35.0	0.0	5.0	60.0	40.0	0.0	20.0	40.0	19.0	42.9	0.0	38.1	25.0	40.0	0.0	35.0
EMP	75.0	0.0	6.3	18.8	81.3	0.0	0.0	18.8	68.8	6.3	0.0	25.0	68.8	25.0	0.0	6.3
INDPROD	30.0	0.0	5.0	65.0	25.0	0.0	5.0	70.0	20.0	10.0	0.0	70.0	10.5	0.0	10.5	78.9
HOUSING	0.0	0.0	19.0	76.2	5.0	0.0	20.0	75.0	0.0	10.0	5.0	85.0	0.0	10.0	5.0	85.0
TBILL	10.0	90.0	0.0	0.0	5.0	90.0	0.0	5.0	28.6	42.9	9.5	19.0	15.0	60.0	10.0	15.0
BOND	10.0	0.0	5.0	85.0	18.2	0.0	9.1	72.7	15.0	5.0	15.0	65.0	15.0	0.0	10.0	75.0
RRESIN	30.0	0.0	5.0	35.0	55.0	0.0	15.0	30.0	30.0	45.0	5.0	20.0	23.8	61.9	0.0	14.3
RGDP	15.0	0.0	0.0	85.0	10.0	0.0	20.0	70.0	10.1	35.0	5.0	50.0	10.1	40.0	5.0	45.0
RCONSUM	15.0	0.0	15.0	70.0	30.0	0.0	10.0	60.0	35.0	0.0	5.0	60.0	30.0	0.0	0.0	70.0
RNRESIN	20.0	0.0	25.0	55.0	25.0	0.0	25.0	50.0	5.0	90.0	0.0	5.0	4.8	85.7	4.8	4.8
RFEDGOV	40.0	0.0	25.0	356.0	60.0	5.0	25.0	10.0	57.1	14.3	19.0	9.5	50.0	30.0	15.0	5.0
RLSGOV	40.0	0.0	10.0	50.0	50.0	0.0	20.0	30.0	45.0	25.0	10.0	20.0	15.0	35.0	25.0	25.0
CPI	0.0	30.	15.0	25.0	15.0	35.0	0.0	50.0	0.0	50.0	10.0	40.0	10.0	55.0	5.0	30.0
MEAN	26.9	10.3	12.7	50.1	34.1	8.7	13.6	43.6	24.6	30.1	8.2	37.1	21.2	34.5	8.4	35.9

4.2. Results by method of combination, variable and accuracy criteria

Table 5 shows the percentage of beats by variable and accuracy criteria for each combination procedure. The following comments are worth mentioning:

1. Results about Euclidean and simplex spaces vary depending on the accuracy measure considered. Whereas combinations in the former are clearly better with MAPE and MdAPE, those in the latter are better with MAE and RMSE.
2. When we analyse combinations according to the type of weights, fixed weights are always the worst, therefore it is worthwile to use varying weights.
3. *CAS* is on average the best, reaching 50% of the cases with ME.
4. S_STC is only the best for the *PGDP* considering ME, MAPE and MdAPE, whereas E_STC is the best for several variables when we consider MAPE and MdAPE.
5. *AVE*'s best results occur with RMSE.

4.3. Results by number of forecasters and accuracy criteria

Tables 6 and 7 show the results of each combination by the number of forecasters and accuracy criteria.

Table 6. *Number of beats of each combination by accuracy criteria and number of forecasts.*

	Mean Error				*RMSE*				*MAPE*				*MdAPE*			
	AVE	*E_STC*	*S_STC*	*CAS*	*AVE*	*E_STC*	*S_STC*	*CAS*	*AVE*	*E_STC*	*S_STC*	*CAS*	*AVE*	*E_STC*	*S_STC*	*CAS*
$J < T$	52	21	23	95	66	18	26	75	51	65	12	66	46	67	12	63
(%)	(27.3)	(10.8)	(11.9)	(50.0)	(35.7)	(9.9)	(14.0)	(40.4)	(26.3)	(33.5)	(6.25)	(34.1)	(24.3)	(35.8)	(6.4)	(33.5)
$J > T$	31	13	17	63	37	10	20	61	27	36	15	50	20	43	13	53
(%)	(25.2)	(10.4)	(13.9)	(50.4)	(29.1)	(7.7)	(15.4)	(47.9)	(21.2)	(28.0)	(11.9)	(39.0)	(15.1)	(33.6)	(10.1)	(41.18)
TOTAL	83	34	40	158	103	28	46	135	78	101	27	116	65	111	25	116
(%)	(26.4)	(10.7)	(12.7)	(50.2)	(33.0)	(9.3)	(14.6)	(43.4)	(24.2)	(31.3)	(8.4)	(36.0)	(20.6)	(34.9	(7.9)	(36.6)

Table 7. *Number of beats of EUCLIDEAN and SIMPLEX combinations by accuracy criteria and number of forecasts.*

	Mean Error		*RMSE*		*MAPE*		*MdAPE*	
	EUCLIDEAN	*SIMPLEX*	*EUCLIDEAN*	*SIMPLEX*	*EUCLIDEAN*	*SIMPLEX*	*EUCLIDEAN*	*SIMPLEX*
$J < T$	73	118	85	101	116	78	113	75
(%)	(38.1)	(61.9)	(45.6)	(54.4)	(59.8)	(40.2)	(60.1)	(39.9)
$J > T$	44	80	47	80	63	65	63	66
(%)	(35.7)	(64.4)	(36.8)	(63.3)	(49.2)	(50.9)	(48.7)	(51.3)
TOTAL	117	198	131	181	179	143	176	141
(%)	(37.1)	(62.9)	(42.0)	(58.0)	(55.6)	(44.4)	(55.5)	(44.5)

In Table 6, we present the number of beats of each combination according to the accuracy criteria and the number of forecasters. Only in 2 cases, E_STC beats *CAS* and always when the number of forecasters is lower than that of observations. In the rest of the cases, *CAS* is always the best reaching 50% of ME, almost twice of AVE combination.

In Table 7, we show the results attending to the space where the combination is formed. As much as $J > T$ simplex is always the best, reaching more than 60% of the cases for ME and RMSE. When $J < T$, for MAPE and MdAPE, Euclidean combinations are better. These good results are obtained because E_STC works very well depending on these measures.

4.4. Results according to the variability of the forecasts

The basic idea under this section is the following: a fixed-weight combination assigns the same weight to forecasts, so if variability among them is small, then the average will work well in the same direction, however wrong it may be ('precisely' wrong) unless they are unbiased. On the other hand, when variability is high, it is better to assign different weights. This is in line with the results obtained by Jose and Winkler (2008) by comparing the accuracy of the average with trimmed and Winsorized averages and the results by Genre et al. (2013) by using the European Central Bank (ECB) survey of professional forecasters. In this latter paper, they find that some combination methods outperform the simple average of forecasts in variables with heterogeneity of forecasters and apparent bias.

In order to verify this hypothesis, we compute the variation coefficient (VC) of each variable for each combination and forecast period from 2015 to 2018. We also plotted the forecasts for each period [5]. In fact, this issue forms part of the selection procedure presented in this paper, i.e. to select those forecasters that do not share common information. In this empirical application, the forecasters come from the Survey of Professional Forecasters (SPF) and may have common information in forming their forecasts. This is the reason why we expect some forecasts to be highly correlated (even redundant) and others with low correlation. Then, *CAS* takes advantage of this situation and usually generates better results.

The main comments that can be pointed out are the following:

1. When all the forecasts included in the sample are highly correlated and their plots show a similar behaviour, AVE is usually the best combination. A clear example of this situation is shown in Figure 1 where we plot the forecasts for *NGDP* for all the samples.

[5] In order to save space, these results are available upon request.

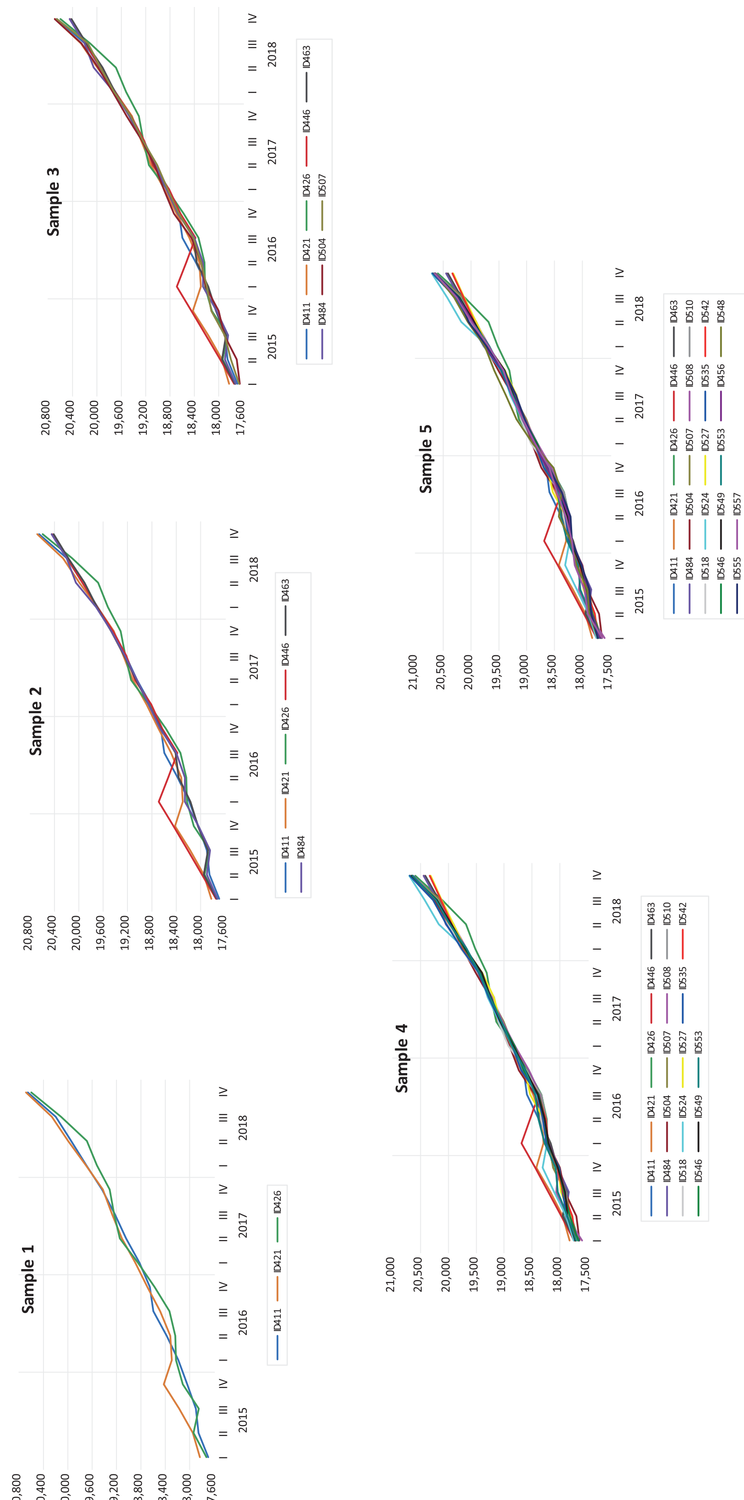

Figure 1. *NGDP forecasts by samples.*

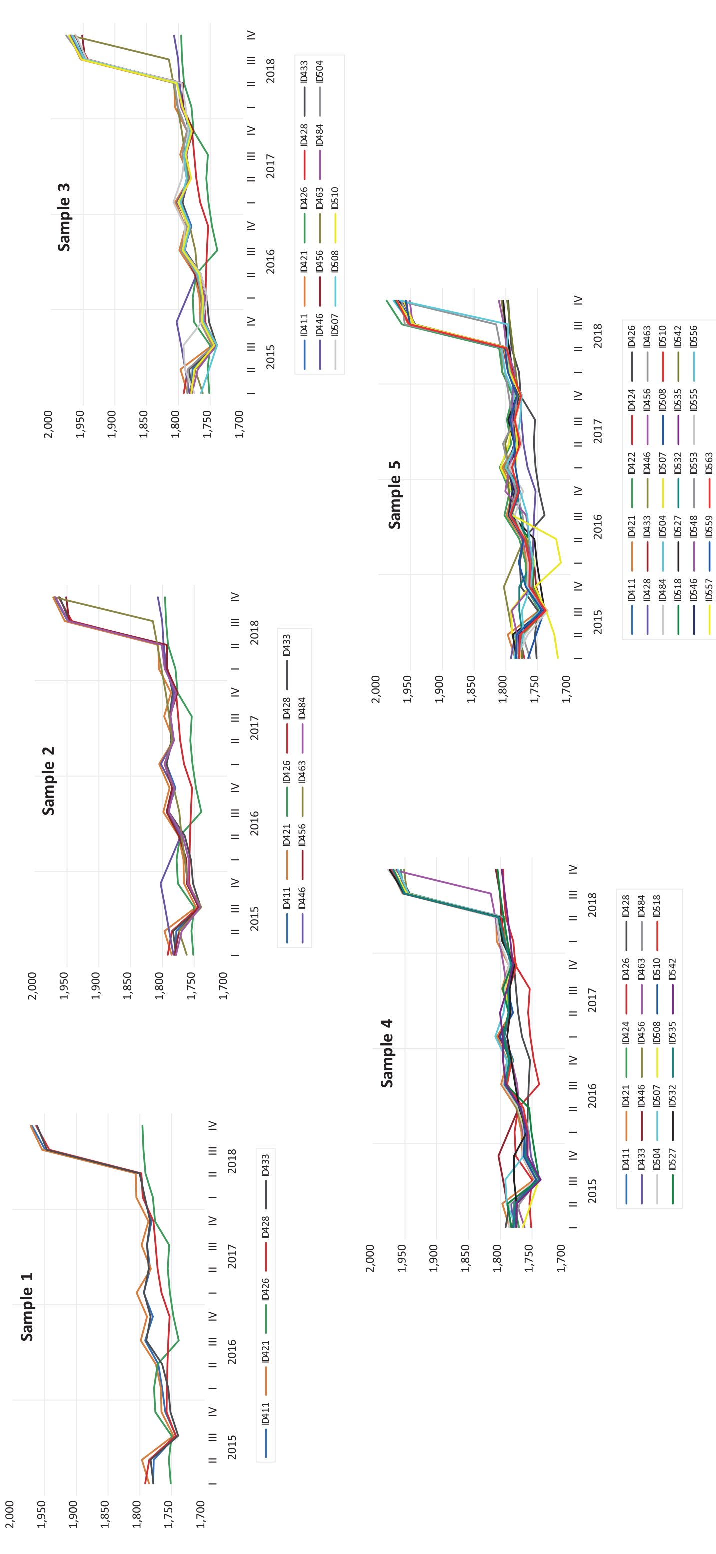

Figure 2. *RLSGOV forecasts by samples.*

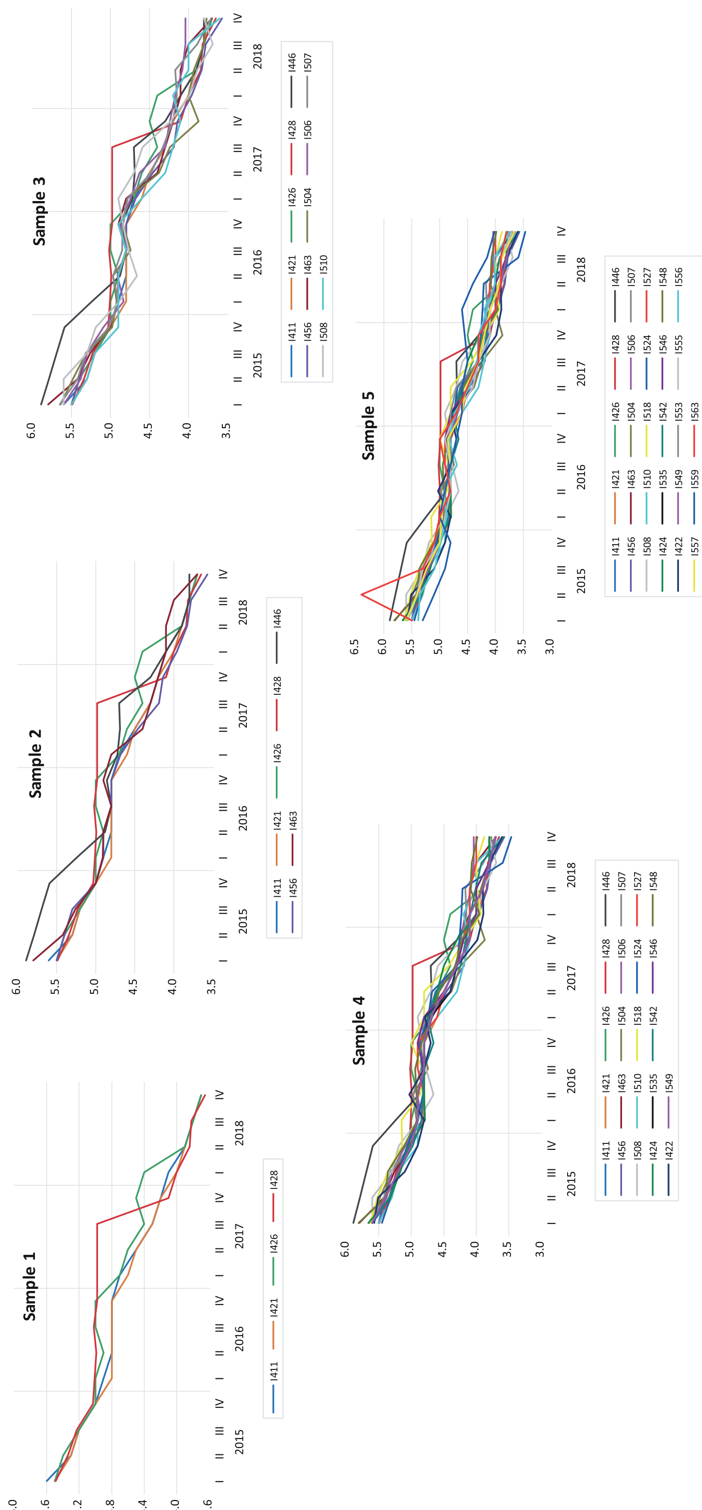

Figure 3. *UNEMP forecasts by samples.*

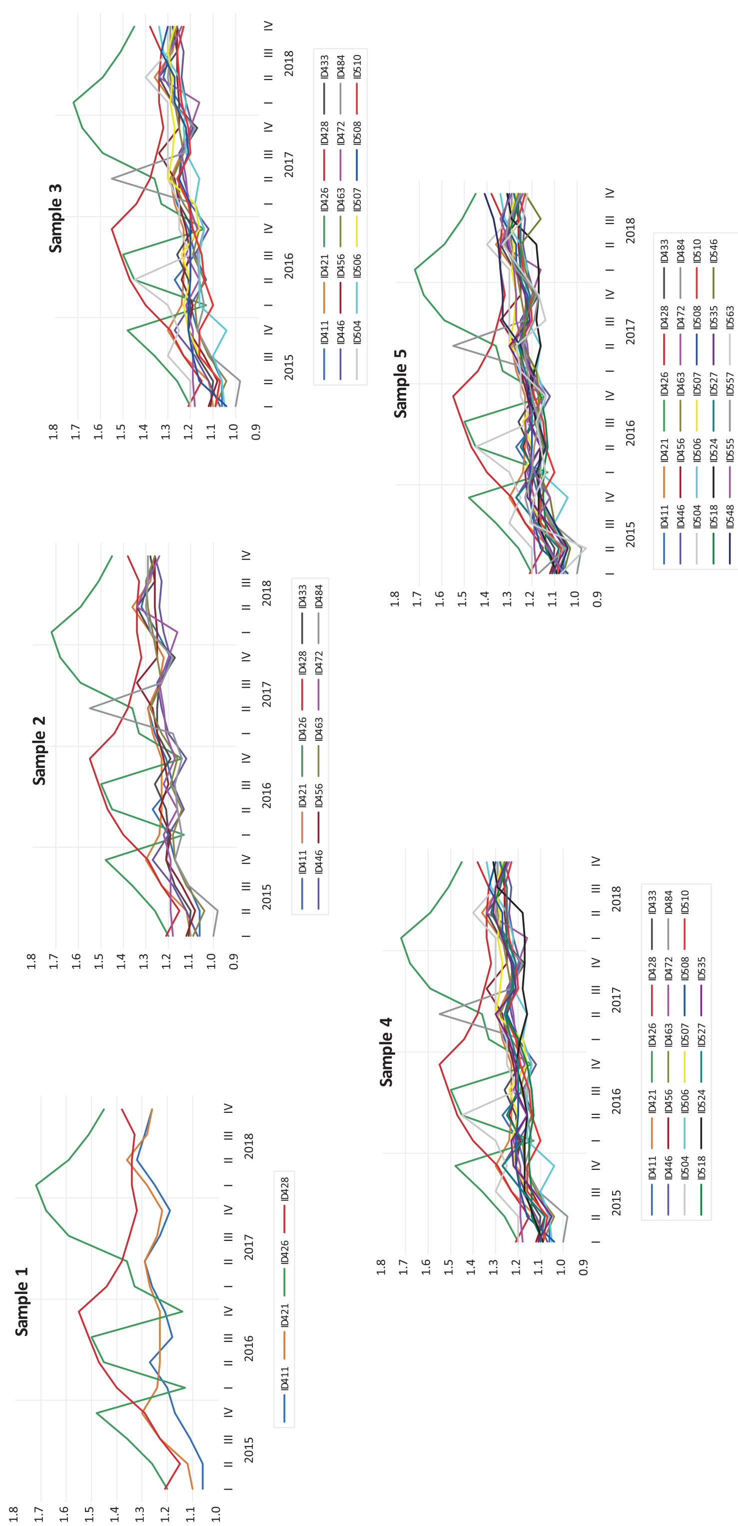

Figure 4. *HOUSING forecasts by samples.*

2. When some of the forecasts are correlated but their plots differ somewhat, AVE is better because of its varying-weight allocation. Figure 2 shows this situation for $RLSGOV$.

3. In a mixed situation with some forecasts highly correlated and some others not so, *CAS* is the best because it only selects non-redundant forecasters. In Figure 3 we show this situation for $UNEMP$.

4. In general, with low correlated forecasts, varying-weight combinations generate better results: the E_STC and S_STC, when the forecasts show a similar behaviour, and *CAS*, when they don't. Figure 4 shows a clear example of this situation for $HOUSING$.

Table 8 shows the variation coefficient (VC) and results for the aforementioned variables[6]. The analysis of the VC will be done jointly with Figures 1 to 4.

1. **NGDP**: All the graphs in Figure 1 show very little variation between forecasts. The VC in each sample is very low, suggesting that AVE should be used. Looking at the combination results, AVE is the winner in all the samples with the exception of sample 4. In this case, *CAS* generates the best forecasts for all the forecasting periods. Notice that in the graph for sample 4, although the forecasts follow a similar behaviour, there are some of them with different patterns that can be used to improve the forecast combination through *CAS*.

2. **RLSGOV**: The behaviour of the forecasts for this variable is different from the one observed before. In this case, the forecasts seem to have a similar behaviour, but the correlation between them is not too high. Then, assigning different weights generates better combinations. Looking at Figure 2, we can see that S_STC obtains very good results in 2017 and perhaps in 2016. Our perception from the graph is confirmed in Table 7: varying-weight combinations outperform the fixed-weight one. This situation is also supported by the VC, which shows higher values than the observed for *NGDP*. So, in this case, the fact that not all the forecasts show the same pattern leads to better forecasting results with varying-weight methods.

3. **UNEMP**: The VC of this variable in Table 7 clearly shows higher values than the observed for the previous variables. This fact can indicate that the average forecast may not be the best combination in this case. Looking at Figure 6, not all the forecasts have the same pattern. This favors the varying-weight combinations, E_STC, S_STC and *CAS*, the latter being the one that beats more times. Therefore, in this case, selection is better than a full combination either fixed AVE or varying E_STC.

[6]The VC, figures and results for the other variables are available upon request. They have been omitted to save space.

Table 8. *Coefficients of variation for selected variables and number of beats of the combination procedures by samples.*

	NGDP					*UNEMP*					*RLSGOV*					*HOUSING*				
	CV	*AVE*	*E_STC*	*S_STC*	*CAS*	*CV*	*AVE*	*E_STC*	*S_STC*	*CAS*	*CV*	*AVE*	*E_STC*	*S_STC*	*CAS*	*CV*	*AVE*	*E_STC*	*S_STC*	*CAS*
Sample (1)		9	2	3	2		5	4	2	5		9	5	0	2		0	0	0	16
2015	0.667	2	2	0	0	0.603	2	2	0	0	0.588	3	1	0	0	7.970	0	0	0	4
2016	0.405	0	0	3	1	2.150	0	2	2	0	0.724	3	1	0	0	11.287	0	0	0	4
2017	0.313	3	0	0	1	4.927	0	0	0	4	0.695	0	2	0	2	9.604	0	0	0	4
2018	0.409	4	0	0	0	1.575	3	0	0	1	0.324	3	1	0	0	9.850	0	0	0	4
Sample (2)		8	1	3	4		7	3	1	5		4	1	4	7		1	2	4	9
2015	0.830	4	0	0	0	3.447	1	1	0	2	0.834	0	0	0	4	6.813	0	0	0	4
2016	0.365	1	0	2	1	2.011	3	0	1	0	0.627	3	0	1	0	9.079	0	0	1	3
2017	0.245	0	1	1	2	4.080	0	2	0	2	0.566	1	0	3	0	8.518	0	0	2	2
2018	0.465	3	0	0	1	2.434	3	0	0	1	1.666	0	1	0	3	7.486	1	2	1	0
Sample (3)		6	6	1	3		5	4	1	6		2	4	3	7		0	1	0	11
2015	0.741	2	2	0	0	2.743	0	2	0	2	0.807	0	1	0	3	6.575	0	0	0	4
2016	0.342	1	0	1	2	1.829	3	1	0	0	0.550	2	1	1	0	8.528	0	0	0	4
2017	0.217	1	2	0	1	3.804	0	1	0	3	0.504	0	0	0	4	7.485	0	1	0	3
2018	0.392	2	2	0	0	2.951	2	0	1	1	1.418	0	2	2	0	6.476	0	0	0	4
Sample (4)		2	5	0	10		6	4	1	5		9	3	2	2		0	3	1	13
2015	0.590	0	2	0	2	2.253	0	0	0	4	0.725	3	0	0	1	6.188	0	0	0	4
2016	0.306	2	2	0	1	1.766	2	2	0	0	0.499	3	1	0	0	7.665	0	0	1	3
2017	0.248	0	1	0	3	3.104	1	1	1	1	0.446	3	0	1	0	6.814	0	2	0	2
2018	0.502	0	0	0	4	3.132	3	1	0	0	1.751	0	2	1	1	6.247	0	0	0	4
Sample (5)		6	5	0	5		1	2	0	13		6	1	4	5		0	0	6	10
2015	0.546	1	2	0	1	2.384	0	0	0	4	0.803	3	1	0	0	5.869	0	0	1	3
2016	0.360	1	2	0	1	1.358	0	0	0	4	1.920	1	0	3	0	6.403	0	0	1	3
2017	0.233	2	0	0	3	3.043	0	0	0	4	0.429	0	0	0	4	6.320	0	0	0	4
2018	0.459	3	1	0	0	3.463	1	2	0	1	1.708	2	0	1	1	5.825	0	0	4	0

4. **HOUSING**: Figure 4 is a clear example for *CAS* to form a combination. Different behaviour of some forecasts and high VC are the clues to select forecasters to obtain better forecasting results. Although there is a common behaviour of some forecasts, the selection of orthogonalized forecasts improves the results.

Similar results are confirmed for the other variables analysed in the empirical application. As a matter of fact, high VC and different behaviour might be the clues to consider CAS as the best subcombination to forecast a variable.

4.5. Results according to the forecast ability

When the Diebold and Mariano (1995) or Giacomini and White (2006) tests are not appropriate, it might be interesting to break down the Mean Squared Forecast Error (MSFE) into three components (bias, variance, and covariance) to assess which of them holds sway over a given MSFE:

$$MSFE := \frac{1}{H}\sum_{h=1}^{H}\left(\widehat{Y}_{T+h} - Y_{T+h}\right)^2 \equiv \left(\overline{\widehat{Y}}_H - \overline{Y}_H\right)^2 + \left(sd(\widehat{Y}_H) - sd(Y_H)\right)^2 \tag{5}$$
$$+2\left(sd(Y_H)\right)\left(sd(Y_H)\right)\left(1 - corr[\widehat{Y}_H, Y_H]\right),$$

where $\overline{\widehat{Y}}_H$ is an H-period average forecast, $\overline{Y}_H$ is the corresponding average for the realized values (Y_H), $sd(\widehat{Y}_H)$ is the standard deviation of the forecasts, $sd(Y_H)$ is the standard deviation of the realized values for the forecast period, and $corr[\widehat{Y}_H, Y_H]$ is the correlation between forecasts and realized values. Then, proportions are defined as follow:

$$\text{Bias proportion: } \frac{\left(\overline{\widehat{Y}}_H - \overline{Y}_H\right)^2}{MSFE},$$

$$\text{Variance proportion: } \frac{\left(sd(\widehat{Y}_H) - sd(Y_H)\right)^2}{MSFE},$$

$$\text{Covariance proportion: } \frac{2\left(sd(Y_H)\right)\left(sd(Y_H)\right)\left(1 - corr[\widehat{Y}_H, Y_H]\right)}{MSFE},$$

We study which one constributes more to the MSFE. A ranking of preferences may be given by the following four situations:

1. CASE 1: The best will be when there are little bias and variance (hence, high covariance proportion).

2. CASE 2: The next one will be when there is little bias, but high variance (hence, low covariance proportion).

3. CASE 3: Bad situations happen when the bias is high: either with high variance,

4. CASE 4: Or the worst, with low variance (‘precisely’ wrong).

Using this classification, we show in Table 9 the bias, variance, and covariance proportions for the combination procedures with lowest MSFE[7] and in Table 10 we summarize this information according to Euclidean and simplex combinations.

Table 9. *Classification according to their forecast ability.*

	AVE		E_STC		S_STC		CAS		$TOTAL$	
	#	%	#	%	#	%	#	%	#	%
Case 1	10	13.3	3	4.0	1	1.4	3	4.0	17	5.67
Case 2	0	0.0	18	24.0	28	37.8	18	24.0	65	21.67
Case 3	8	10.7	34	45.3	39	52.7	44	58.7	125	41.67
Case 4	57	76.0	20	26.7	6	8.1	10	13.3	93	31.0

#: Proportions of the best MSFE procedure included in specific cases.

Table 10. *Classification according to Euclidean or Simplex.*

	$Euclidean$		$Simplex$	
	#	%	#	%
Case 1	13	8.7	4	2.7
Case 2	18	12.0	46	30.9
Case 3	42	28.0	83	55.7
Case 4	77	51.3	16	10.7

#: Proportions of the best MSFE procedure included in specific cases.

From Table 9, we can conclude that AVE is mainly classified in the worst situation: high bias and low variance (76% of the cases), but it is also the first method classified in the best situation (13.3% of the cases).

In general, the other methods are classified most of the times in cases 2 and 3 (low bias and hig variance or high bias and high variance).

From Table 10, the case 3 is the most often with the simplex representing more than 50% of the cases, being case 2 the second best situation that happens almost 31%.

Considering the different methods of combination, we obtain that for AVE, case 4 is the most often whith MSFE. For all the others, case 3 is the one that happens most often.

[7]The specific values for the bias, variance, and covariance proportions for each variable, each sample, and each combination procedure are available upon request.

5. Conclusions

In this paper, we have used the Split-Then-Combine (STC) approach to build positive weights that sum up to one. Because of these two restrictions, most methods from multivariate statistics are inapplicable for combinational datasets, giving rise to a number of issues that make inappropiate the Euclidean geometry. Instead, the Aitchison geometry considers combinations of forecasters inside the simplex, the sampling space of positive weights adding up to one. A one-to-one transformation between the simplex and real spaces allows us to use the sample centre of the simplex, with time-varying weights, to find a Combinations after Selection (CAS) simplicial subcombinations that selectis those forecasters in a full combination that assign higher weights than the one allocated by the benchmark average.

The methodology can be summarized in these steps: first, we split experts' forecasts by seasons to assess their relative forecast performance that periodically evolves over time. Second, we choose as a combination vector the sample centre of the simplex. Then, we select forecasters inside a simplex of lower dimension by means of a centred logratio transformation. Finally, we make rolling, truly out-of-sample, one-step-ahead combinations of forecasts, even in cases where the sample size is smaller than the number of forecasters. Once a new observation is known, we recalculate the weights that we then keep one-step-ahead to form a new out-of-sample combination.

We present experimental results with a pool of expert forecasters of the US macroeconomy over the period 1991–2018. In most cases, the Combination after Selection strategy improves the average (neutral combination in the simplex space) with different criteria of forecasting accuracy, and works very well even when the number of forecasters is greater than the number of observations. Forecast combination can improve forecasting accuracy, provided that the sets of forecasters contain some independent information.

As a general rule, we can conclude that when there are a high number of heterogeneous forecasters to be combined, the best way to form a combination is by selecting a CAS simplicial subcombination formed by the most weighted, non-redundant forecasters.

For combinations of forecasts, the relevant information is contained in the clr coefficients between forecasts. This by itself might also be interesting to symmetrize possible right-skewed distributions of forecaster's precisions. Further research, therefore, will focus on pivot (or more general orthonormal) coordinates that aim to extract all relative information about a particular forecast in the combination. Moreover, exploratory and preprocessing issues may also be discussed: visualization, outlier detection, missing values, and zeros form a touchstone of the logratio analysis. Finally, many popular statistical methods, such as principal component analysis, cluster analysis, classification and regression analysis, may be adapted for dealing with combinations carrying relative information.

Ackowledgements The second author acknowledges financial support from the Spanish Government Project PID2019-108079GBC22.

References

Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Sociesty, B*, 44, pp. 139-177.

Aitchison, J. (1986). *The statistical analysis of compositional data*, Monographs on Statistics and Applied Probability. London: Chapman & Hall, (Reprinted in 2003 with additional material by The Blackburn Press).

Aitchison, J. (1992). On criteria for measures of compositional difference, *Mathematical Geology*, 24, 365-379.

Aitchison, J. (1997). The one-hour course in compositional data analysis or compositional data analysis is simple. In V. Pawlowsky-Glahn (Ed.), Proceedings of IAMG'97—The third annual conference of the International Association for Mathematical Geology, Volume I, II and addendum (pp. 3-35). Barcelona: International Center for Numerical Methods in Engineering (CIMNE), pp. 1100.

Arroyo, A.S.M. and de Juan Fernández, A. (2014). Split-then-Combine method for out-of-sample combinations of forecasts, *Journal of Business Administration Research*, 3 (1), pp. 19-37.

Barceló-Vidal, C. and Martín-Fernández, J.A. (2016). The mathematics of Compositional Analysis, Austrian *Journal of Statistics,* 45, pp. 57-71, doi: 10.17713/ajs.v45i4.142.

Barrow, D.K. and Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: implications for inventory management. *International Journal of Production Economics,* 177, pp. 24-33.

Billheimer, D., Guttorp, P. and Fagan, W. (2001). Statistical interpretation of species composition, *Journal of the American Statitictical Asociation*, 96 (456), 2001, pp. 1205-1214.

Van den Boogaart, K.G. and Tolosana-Delgado, R. (2013). *Analyzing Compositional Data with R*, Berlin: Springer.

Bujosa-Brun, M., García-Ferrer, A., de Juan, A. and Martín Arroyo, A. (2020). Evaluating early warning and coincident indicators of business cycles using smooth trends, *Journal of Forecasting, 39, pp. 1-17. DOI: 10.1002/for.2601.*

Chayes, F. (1960). On correlation between variables of constant sum, *Journal of Geophysical Research* 65 (12), pp. 4185-4193.

Coenders G, and Ferrer-Rosell, B. (2020). Compositional data analysis in tourism. Review and future directions. *Tourism Anal* 25: pp. 153-168.

Conflitti, C., De Mol, C. and Giannone, D. (2015). Optimal combination of survey forecasts, *International Journal of Forecasting* 31 pp. 1096-1103.

Diebold, F.X. and Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, vol 13(3). pp. 253-263.

Egozcue, J.J. and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, 37 (7), 2005, pp. 795-828.

Egozcue, J.J. and Pawlowsky-Glahn, V. (2019). Compositional data: the sample space and its struture. *Test*, 28(3), pp. 599-638.

Federal Reserve Bank of Philadelphia (2018). Survey of Professional Forecasters documentation. August 29, 2018.

Filzmoser, P., Hron, K., and Templ, M. (2018). *Applied Comopositional Data Analysis*, Springer.

Gabriel, K.R. (1971). The biplot—graphic display of matrices with application to principal component analysis. *Biometrika*, 58 (3), pp. 453-467.

Genre, V., Kenny, G., Meyler, A. and Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average?. *International Journal of Forecasting,* 29, pp. 108-121.

Giacomini, R. and White, H. (2006). Test of conditional predictive ability, *Econometrica*, 74(6), pp. 1545-1578.

Greenacre, M. (2021). Compositional data analysis. *Annual Review of Statistics and its Application*, 8, pp. 271-299.

Lahiri, K., Peng, H. and Zhao, Y. (2017). Online learnig and forecast combination in unbalanced panels. *Econometric Reviews,*vol. 36, pp. 257-288.

Jose, V.R.R. and Winkler, R.L. (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, 24, pp. 163-169.

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Egozcue, J.J. (2011). The Principle of Working on Coordinates, in *Compositional data analysis: Theory and Applications*, eds. V. Pawlowsky-Glahn and A. Buccianti, 29-42, Wiley and sons.

Pawlowsky-Glahn, V. and Egozcue, J.J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 15(5), pp. 384-398.

Pawlowsky-Glahn, V., Egozcue, J.J. and Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data*. Wiley.

Pawlowsky-Glahn, V. and Buccianti, A. (2011). *Compositional Data Analysis: Theory and Applications*. Chichester: Wiley.

Pearson, K. (1897). Mathematical contributions to the theory of evolution. On form of spurious correlation which may arise when indices are used in the measurement of rorgans. *Proceedings of the Royal Society of London,* LX, pp. 489-502.

Poncela, P., Rodríguez, J., Sánchez-Mangas, R. and Senra, E. (2011). Forecast combination through dimension reduction techniques. *International Journal of Forecasting,* 27, pp. 224-237.

Smith, J. and Wallis, K.F. (2009). A simple explanation of the forecast combination puzzle. *Oxford bulletin of economics and statistics*, 71(3), pp. 331-355.

Stock, J.H. and Watson, M.W. (2004). Combination forecasts of output growth in a seven country dataset, *Journal of Forecasting,* vol. 23, pp. 405-430.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of Royal Statistical Society, Series B*, 58, pp. 267-288.