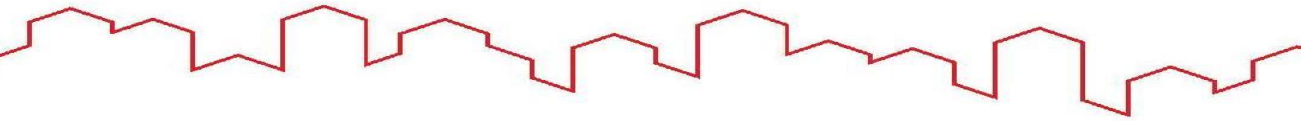
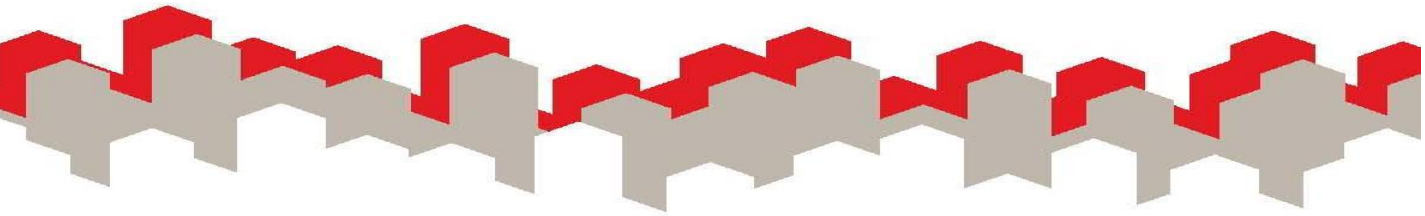


Department of Economics

Econometric Analysis with Compositional and Non-Compositional Covariates

Michael Ben-Gad[1]
City, University of London

Department of Economics
Discussion Paper Series
No. 22/01



[1] Corresponding author: Department of Economics, City, University of London, Northampton Square, London EC1V 0HB, UK. E-mail: Michael.Ben-Gad.1@city.ac.uk

Econometric Analysis with Compositional and Non-Compositional Covariates

Michael Ben-Gad*
Department of Economics
City, University of London
Northampton Square, London EC1V 0HB, UK

October 2, 2022

Abstract

In this paper I consider how best to incorporate compositional data (shares of a whole which can be represented as points on a simplex) together with noncompositional data as covariates in a linear regression. The standard method for incorporating compositional data in regressions is to omit one share to overcome the problem of singularity. I demonstrate that doing so ignores the compositional nature of the data and the resulting models are not objects in a vector space, which in turn reduces their usefulness. In terms of Aitchison geometry—the only geometry that can generate a vector space on a simplex—I show how this method also grossly distorts the relationship between points in the compositional data set. Futhermore, the regression coefficients that result are not permutation invariant, so unless there is an obvious baseline category to be omitted with which the other variables in the composition ought naturally to be compared, this approach gives researchers latitude to choose the permutation of the model that supports a particular hypothesis or appears most convincing in terms of p-values. The alternatives in this paper build on work by Aitchison (1982, 1986) on additive logarithmic ratio (ALR) transformations and Egozcue et al. (2003) on isometric logarithmic ratio (ILR) transformations. Transforming the compositional data using ALRs generates regressions that are permutation invariant and hyperplanes in a vector space. However, ALRs translate the points in the simplex into coordinates relative to an oblique basis, so the angles and distances between the data points remain somewhat distorted—though this distortion is inversely related to the number of shares in the composition. By contrast, ILRs eliminate the distortion by translating the points into coordinates relative to an orthogonal basis. However, the resulting regressions are no longer permutation invariant and are difficult to interpret. To overcome these shortcomings, Hron et al. (2012) suggest using ILRs, but combining the coefficient estimates across all the different permutations to produce one statistical model. I demonstrate that estimating a separate regression for each permutation is unnecessary—estimating either a single regression using ALR coordinates or a constrained regression and then multiplying the resulting regression coefficients and standard errors associated with the compositional variables by a simple factor is sufficient. Though log-ratios incorporate more information about the nature of compositional data as coordinates in a simplex, I demonstrate that it does not exacerbate the inherent multicollinearity present in compositional datasets. Throughout, I use economic growth regressions with compositional data on ten religious categories, similar to Barro and McCleary (2003) and McCleary and Barro (2006), to demonstrate and contrast all these different approaches.

JEL classification: C50, O47

Keywords: Compositional Data, Aitchison Geometry, Isometric Logarithmic Ratios, Economic Growth Regressions

*mbengad@city.ac.uk. I would like to thank John Luke Gallup, seminar participants at City, University of London, particularly Mireia Jofre-Bonet, Giulia Faggio and Agne Suziedelyte, and participants of the 2022 Conference of the Money, Macro and Finance Society at the University of Kent, for their advice and helpful comments.

1 Introduction

Empirical models in economics and other social sciences often incorporate compositional data—variables that together constitute shares of a whole and can be represented as points on a simplex.[1] Linear regression models usually include an intercept term, meaning that a matrix of explanatory variables which includes compositional data is perfectly multicollinear—a subset of variables can be expressed as a linear combination of the others. To ensure the matrix has full column rank it is common in empirical work to omit one variable within the composition.

There are several problems with this approach. First, though the underlying statistical model is unaltered, the estimated coefficients and associated standard deviations that correspond to the remaining compositional variables are not permutation invariant, and can change a great deal depending on which variable is chosen for omission. Hence, unless there is a baseline category against which the other shares should naturally be compared, researchers are free to choose whichever permutation generates the most desired or persuasive-looking results. Second, regression analysis typically involves analysing data that can be expressed as coordinates in Euclidean space. Indeed, the squared errors whose sum is being minimised are Euclidean distances. Integrating compositional data means including coordinates in a different space—simplex space—for which Euclidean geometry is inappropriate. The regression models estimated in the usual manner are not objects in vector space, and the standard manner in which they are interpreted is inappropriate. When measured using Aitchison (1982, 1986) geometry, which does generate a vector space on a simplex, the manner in which Euclidean geometry distorts the relationship between the coordinates is readily apparent.

To address these problems, I present a new method for estimating regressions with compositional data using isometric logarithmic ratio (ILR) transformations, first developed by Egozcue et al. (2003). My analysis focuses on models where not all the explanatory variable are compositional. The resulting regression is permutation invariant and ensures that the subset of variables which are compositional enter the regression as coordinates of a vector space with an orthonormal basis. Throughout I present the methods used to estimate models in the language of matrix transformations and projections familiar to economists. This allows me to demonstrate the close relationship between constrained regression, additive log-ratios (ALR) developed by Aitchison (1986) and ILRs. ILRs incorporate more information about the underlying nature of the data, yet as I demonstrate, this does not exacerbate the inherent multicollinearity of compositional data.

To demonstrate the practical implications of using compositional and noncompositional data together, Section 2 introduces a simple cross-country economic growth regression as first developed by Barro (1991) and Mankiw et al. (1992). These types of models often incorporate a variety of variables associated with neoclassical growth theory, each of which can be expressed as coordinates on the real line, alongside additional variables that may be composi-

1. Sets of dichotomous variables such as seasonal dummies represent special cases where the points are restricted to the vertices of the simplex.

tional, representing aggregate expenditure shares or demographic attributes of the population in the different countries in the sample. This paper follows Barro and McCleary (2003) and McCleary and Barro (2006) and includes shares of each country's population that adhere to different religious denominations.

In Section 3, I demonstrate that at a fundamental level, the basic statistical properties of a regression that incorporates raw compositional data are permutation invariant—i.e., changing the component within the composition that is omitted has no effect on either the error terms of the regression or the properties of the coefficients corresponding to the covariates outside the composition. Yet these coefficients can only be interpreted in reference to the omitted baseline variable. That may be defensible if there is a natural baseline category. For example, when including in a growth regression the population shares that obtain primary, secondary, or tertiary education alongside those with no education at all, it might seem sensible to omit that last category and treat it as a baseline. In the present context, when we include in our example the religious composition of the population, any baseline religion we might choose will have little justification (Barro and McCleary (2003) and McCleary and Barro (2006) choose Catholics as the omitted category).

There is an extensive literature on a range of problems associated with statistical analysis across many disciplines, particularly regarding power, bias and p-hacking (in the context of economics see Brodeur et al. (2016) and Ioannidis et al. (2017)). I demonstrate that if a regression includes compositional data, these problems are compounded by the way the sizes and signs of coefficients, and associated p-values, can change depending on which variable is omitted.

Moreover, even if there is a natural category to omit, this approach still creates a number of problems. As I demonstrate in Section 4, interpretation of the resulting regression model is undermined by the fact that the application of standard Euclidean geometry to coordinates in a simplex does not generate a vector space. Furthermore, the distances and angles between these coordinates, when measured using the more appropriate Aitchison geometry (Figure 1b), will appear distorted if treated as coordinates with respect to a canonical basis using the Euclidean geometry (Figure 1a). This is especially salient as the relationship between the two distance measures is not monotonic (see Figure 1c).

In Section 5, I demonstrate how models can be made permutation invariant by first transforming the compositional portion of the data using additive logarithmic ratios developed by Aitchison and Shen (1980) and Aitchison (1982, 1986). As demonstrated by Aitchison and Bacon-Shone (1984), the resulting model can also be interpreted as a constrained regression. Once transformed, the compositional data are also coordinates in a vector space, though with an oblique basis, so some distortion of angles and distance remains. To correct this I demonstrate in Section 6 that it is very simple to translate the coefficients and standard errors from additive logarithmic ratio (ALR) transformations to the ILRs developed by Egozcue et al. (2003). In Section 8 I demonstrate that using log-ratios does not exacerbate the problem of multicollinearity, which is inherent to any set of compositional data. Section 9 concludes.

Unlike ALRs, regressions that use ILRs are not permutation invariant, however Hron et al. (2012) demonstrate how to combine the different permutations into a single model. The methodology presented in Section 6 extends their work in several ways. First, unlike in geophysics, where compositional data is often analysed in isolation using log-ratios, in social sciences (or medicine) we often have compositional data alongside noncompositional data. Hence, the compositional data in the models I consider are not the only independent variables but a subset of a larger dataset that has noncompositional components as well.[2] Therefore, I characterise all our results in terms of the two different parts of the partitioned matrix which encompass both compositional and noncompositional data. Second, the method for estimating the model in Hron et al. (2012) requires obtaining the coefficient and standard error corresponding to the log deviation of one variable in the compositional data set, with the respect to the geometric mean of all the others, one at a time. This requires first transforming the data into a particular set of pivot log-ratios, then estimating the regression, and then generating a new set of pivot log-ratios and estimating the regression again for each variable. By contrast, I demonstrate a simple relationship between ALR and ILR transformations which means a model with the latter can be obtained by simply multiplying the coefficients and standard errors from the former by a small scale factor, which only varies by the number of components in the composition.

There is a long-running debate about the relative merits of using ALR and ILR. Critics of the ILR approach, particularly Aitchison (2008), Greenacre et al. (2021) and Greenacre et al. (2022), argue that isometry is not critically important, the results can be hard to interpret, and the method for calculating them is cumbersome. In this application, I demonstrate how the last consideration is not relevant for the case of linear regression. Indeed the two yield similar results, particularly if the number of variables in the composition is large.

2 The Problem of Linear Regression and Compositional Data

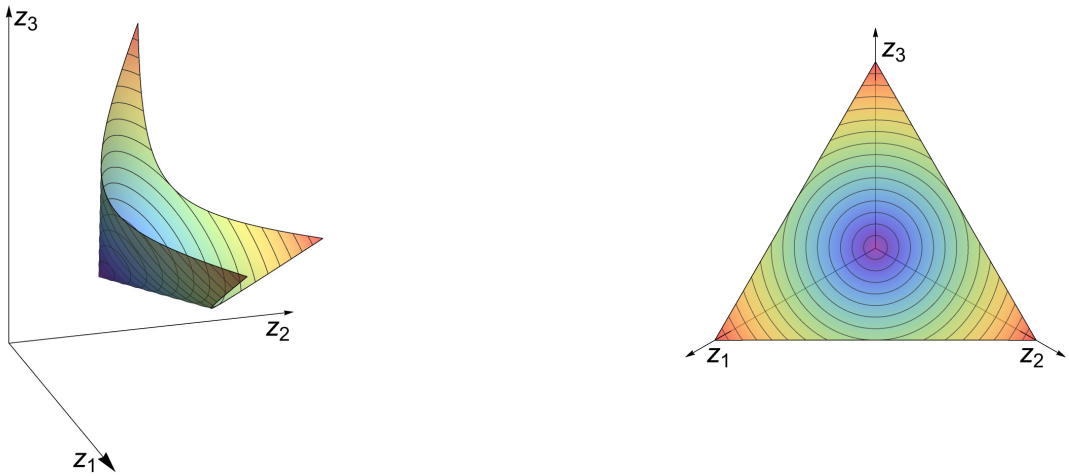
Consider the following specification of the cross-country growth regression adopted from Mankiw et al. (1992):

$$\gamma_h = g - \eta \ln A + \eta \ln y_{h,0} - \frac{\eta\alpha}{1-\alpha} \ln s_{K,h} + \frac{\eta\alpha}{1-\alpha} \ln (p_h + g + \delta) + \pi\Omega_h + \varepsilon_h, \tag{1}$$

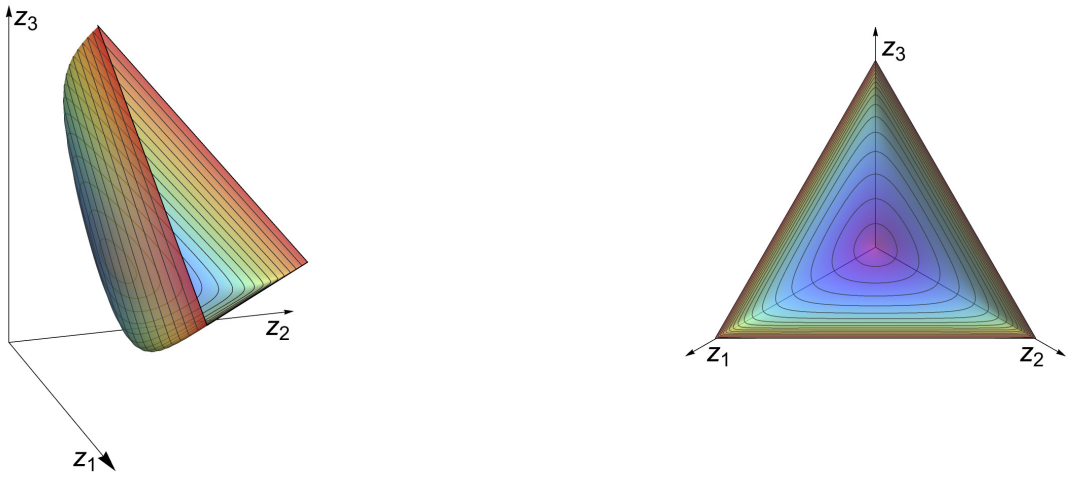
where for each country $h \in \{1, .., n\}$: γ_h represents per-capita output growth, $y_{h,0}$ the initial level of per-capita output, $s_{K,h}$ the share of output devoted to the accumulation of physical capital, and p_h the rate of population growth. α, g, A and δ represent factor shares of physical capital, the rate of change in total factor productivity, the level of labour augmenting technological progress, and the rate of depreciation. These parameter are all assumed the same across the sample of countries, as is the value of η, which the theory posits to be negative, and which measures the rate of convergence. Finally, Ω_h represents an extra vector of additional

2. Chen et al. (2017) considers the case where all the variables, dependent and independent are compositional.

(a) Euclidean Distances



(b) Aitchison Distances



(c) Euclidean/Aitchison Distances

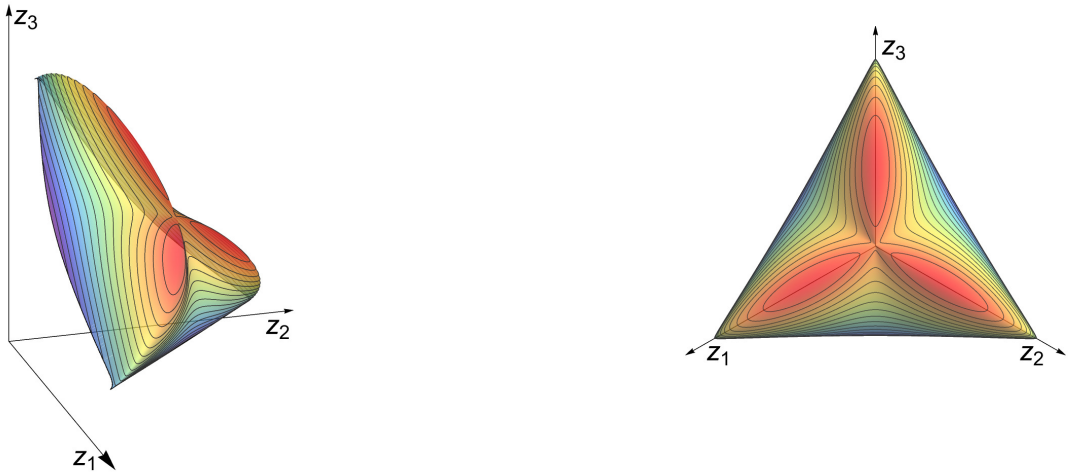


Figure 1: The left hand plots measure distances from the Barycentre $\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$ in a three-part composition for $0 < z_i < z$ and $0 < z$ using a) the standard Euclidean distance measure, b) the Aitchison distance measure in Appendix A.1, and c) the ratio of Euclidean to Aitchison distances. The right-hand plots are the projections on a flat surface where $z_1 + z_2 + z_3 = z$ matches the notion of closure and generates the ternary diagrams of the simplex.

covariates, as for example in Sala-i-Martin (1997), or controls to capture further cross-country heterogeneity.

Suppose all, or some, of the additional variables represented by Ω_h are compositional, so that they represent shares that sum to a constant. This naturally arises if, for example, one wants to know how growth might be related to the age composition of the population, the shares of the population with different levels of educational attainment, the shares who speak particular languages, or a country's religious composition. If we define all the right-hand side variables, along with a column of ones that generate the intercept term, and at least one set of compositional variables as $\mathbf{X}$, then clearly we cannot estimate (1), as $\mathbf{X}$ does not have full column rank and therefore $\mathbf{X}'\mathbf{X}$ is not invertible.

Instead we partition the covariate matrix in two, $\mathbf{X}$=($\mathbf{N}$,$\mathbf{C}$), where $\mathbf{N}$ is an $n \times K$ matrix that incorporates the variables associated with the Solow model and any variables in Ω_h that are not compositional, and $\mathbf{C}$ is an $n \times (D+1)$ matrix where the last column is the vector $\mathbf{i}_n$, that generates the constant term in the regression, and the other D columns are any of the covariates in $\mathbf{X}$ that together are represented as n points on $D-1$ unit simplex:

$$\mathbb{S}^D = \left\{ [c_1, c_2, ..., c_D], \quad c_i > 0, \quad \sum_{i=1}^{D} c_i = 1 \right\}. \tag{2}$$

This means that the rank of matrix $\mathbf{C}$ is less than $D+1$ as in each row the entries in the first D columns sum to one and equal the ones in column $D+1$. To overcome this problem we could remove the last column in $\mathbf{C}$, and estimate the regression without the intercept term. However, forcing the regression hyperplane through the origin is inappropriate for most applications. For the case of the economic growth equation (1), it would mean imposing the restriction that the rate of change in total factor productivity is (globally) zero. Instead, the most common practice is to remove one of the compositional covariates, with the idea that the missing share is "absorbed by the constant."

Define the $(D+1) \times D$ matrix:

$$\mathbf{Q}_{D,d} = \left[\begin{array}{c:c} \mathbf{I}_{d-1,d-1} & \mathbf{0}_{D-d+2,D-d+1} \\ \hdashline \mathbf{0}_{d,D-d+1} & \mathbf{I}_{D-d+1,D-d+1} \end{array}\right].$$

Then the $n \times D$ matrix $\mathbf{C}_{\setminus d} = \mathbf{C}\mathbf{Q}_{D,d}$ is the matrix $\mathbf{C}$ with the $d \in \{1, ..., D+1\}$ column removed and $\mathbf{C}_{\setminus D+1} = \mathbf{C}\mathbf{Q}_{D,D+1}$ represents the coordinates in the simplex (2).

Now the equation can be estimated as:

$$\boldsymbol{y} = \mathbf{N}\boldsymbol{\beta}_{N\setminus d} + \mathbf{C}_{\setminus d}\boldsymbol{\beta}_{C\setminus d} + \boldsymbol{\varepsilon}_{\setminus \boldsymbol{d}}, \tag{3}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of growth rates for the n different countries in the sample, $\boldsymbol{\beta}_{\setminus d}$ the length D vector of coefficients associated with the compositional data after the d^{th} column has

been removed along with the intercept term, and $\boldsymbol{\varepsilon}_{\setminus d}$ is the vector of error terms. The term $\boldsymbol{\beta}_{N\setminus d}$ represents the length K vector of coefficients that correspond to the noncompositional data when the equation is estimated with the d^{th} column removed from $\mathbf{C}$. Note that the special case $d = D + 1$ will generate a regression without an intercept.

Define the projection matrices $\mathbf{P}_N \equiv \mathbf{N}\left(\mathbf{N}'\mathbf{N}\right)^{-1}\mathbf{N}'$ and $\mathbf{P}_{C\setminus d} \equiv \mathbf{C}_{\setminus d}\left(\mathbf{C}'_{\setminus d}\mathbf{C}_{\setminus d}\right)^{-1}\mathbf{C}'_{\setminus d}$. Furthermore define the $n \times (K + D)$ partitioned matrix which contains both the variables in $\mathbf{N}$ and the variables in $\mathbf{C}_{\setminus d}$: $\mathbf{X}_{\setminus d} = [\mathbf{N} \vdots \mathbf{C}_{\setminus d}]$ and the corresponding projection matrix $\mathbf{P}_{X\setminus d} \equiv \mathbf{X}_{\setminus d}\left(\mathbf{X}'_{\setminus d}\mathbf{X}_{\setminus d}\right)^{-1}\mathbf{X}'_{\setminus d}$.

We solve the normal equations associated with (3) for the vector of coefficients and corresponding variances:

$$\begin{bmatrix} \boldsymbol{\beta}_{N\setminus d} \\ \boldsymbol{\beta}_{C\setminus d} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{N}'\left(\mathbf{I}_n - \mathbf{P}_{C\setminus d}\right)\mathbf{N}\right)^{-1}\mathbf{N}'\left(\mathbf{I} - \mathbf{P}_{C\setminus d}\right)\boldsymbol{y} \\ \left(\mathbf{C}'_{\setminus d}\left(\mathbf{I}_n - \mathbf{P}_N\mathbf{N}\right)\mathbf{C}_{\setminus d}\right)^{-1}\mathbf{C}'_{\setminus d}\left(\mathbf{I} - \mathbf{P}_N\right)\boldsymbol{y} \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} \mathrm{Var}\left(\boldsymbol{\beta}_{N\setminus d}\right) \\ \mathrm{Var}\left(\boldsymbol{\beta}_{C\setminus d}\right) \end{bmatrix} = \frac{1}{n-K}\begin{bmatrix} \boldsymbol{\varepsilon}'_{\setminus d}\boldsymbol{\varepsilon}_{\setminus d}\left(\mathbf{N}'\left(\mathbf{I}_n - \mathbf{P}_{C\setminus d}\right)\mathbf{N}\right)^{-1} \\ \boldsymbol{\varepsilon}'_{\setminus d}\boldsymbol{\varepsilon}_{\setminus d}\left(\mathbf{C}'_{\setminus d}\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\right)^{-1} \end{bmatrix} \tag{5}$$

where $\mathbf{I}_n$ is an $n \times n$ identity matrix and the error term is:

$$\boldsymbol{\varepsilon}_{\setminus d} = \left(\mathbf{I}_n - \mathbf{P}_{X\setminus d}\right)\boldsymbol{y}. \tag{6}$$

A number of studies follow the procedure outlined above, where the compositional data represented by $\mathbf{C}_{\setminus d}$ include educational attainment: (Petrakis and Stamatakis (2002)), age structure: (Lindh (1999)), shares of GDP or government spending: (Devarajan et al. (1996), Voigt et al. (2015), Bose et al. (2007), Cavallo and Daude (2011), and Voigt et al. (2015)), languages spoken: (Hall and Jones (1999) and Rodrik et al. (2004)), and ancestory: (Putterman and Weil (2010)). In the next two sections, I examine some of the deficiencies of this approach, before exploring alternative methods that use log-ratios. Throughout, I employ as an example a cross-country regression that captures the main features of the Solow growth model in 1, augmented with data that divides each country's population according to religious affiliation as in Barro (1996), Sala-i-Martin (1997), Hall and Jones (1999), Barro and McCleary (2003), Sala-i-Martin et al. (2004), Noland (2005), and McCleary and Barro (2006).

3 A Lack of Permutation Invariance

It might seem that every permutation of (3) generates a completely different model in (4)–(6). In fact, it is more appropriate to think of each permutation as generating a different perspective of the same statistical model. Most attributes will remain unchanged, but some will differ as we change the direction from which it is viewed.

Define the $D \times D$ matrix for $d \neq f$:

$$\mathbf{L}_{d,f} = \begin{cases} \mathbf{A}_d\mathbf{S}_{d+1}\mathbf{S}_d\mathbf{S}_{d-1}...\mathbf{S}_{f-1} & \text{for all } d < f \\ \mathbf{A}_{d-1}\mathbf{S}_{d-1}\mathbf{S}_{d-2}...\mathbf{S}_{f+1} & \text{for all } d > f \end{cases}$$

where:

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{I}_{i-1} & -\mathbf{i}_{D-1} & \mathbf{0}_{i,D-i} \\ \mathbf{0}_{D-i-1,i-1} & 1 & \mathbf{I}_{D-i} \end{bmatrix},$$

$$\mathbf{S}_j = \begin{bmatrix} \mathbf{I}_{j-1} & \mathbf{0}_{j-1,2} & \mathbf{0}_{j-1,D-j-1} \\ \mathbf{0}_{2,j-1} & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} & \mathbf{0}_{2,D-j-1} \\ \mathbf{0}_{D-j-1,j-1} & \mathbf{0}_{D-j-1,2} & \mathbf{I}_{D-j-1} \end{bmatrix},$$

where $\mathbf{i}_m$ a column vector of ones of length m, and $\mathbf{0}_{m,n}$ an $m \times n$ matrix of zeros.

Postmultiplying any $\mathbf{C}_{\setminus d}$ by $\mathbf{A}_i$ replaces the remaining i^{th} column in that matrix with the i^{th} column in the complete matrix $\mathbf{C}$, and postmultipling the result by the permutation matrix $\mathbf{S}_j$ exchanges the the j and j-1 columns in the new matrix. Hence for all $d, f \in \{1, ..., D\}$ where $d \neq f$: $\mathbf{C}_{\setminus d} = \mathbf{C}_{\setminus f}\mathbf{L}_{d,f}$, $\mathbf{C}_{2\setminus f} = \mathbf{C}_{\setminus d}\mathbf{L}_{f,d}$ and $\mathbf{L}_{d,f} = \left(\mathbf{L}_{f,d}\right)^{-1}$.

Lemma 1. *The projection matrices associated with the matrices* $\mathbf{C}$ *and* $\mathbf{X}$*, as well as the error terms, are all permutation invariant. Hence* $\mathbf{P}_{C\setminus d} = \mathbf{P}_{C\setminus f} = \mathbf{P}_C$, $\mathbf{P}_{X\setminus d} = \mathbf{P}_{X\setminus f} = \mathbf{P}_X$ *and* $\boldsymbol{\varepsilon}_{\setminus d} = \boldsymbol{\varepsilon}_{\setminus f} = \boldsymbol{\varepsilon}$ *for all* $d, f \in \{1, ..., D\}$.

Proof.

$$\begin{aligned} \mathbf{P}_{C\setminus d} &= \mathbf{C}_{\setminus d}\left(\mathbf{C}'_{\setminus d}\mathbf{C}_{\setminus d}\right)^{-1}\mathbf{C}'_{\setminus d} \\ &= \mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\left(\left(\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\right)'\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\right)^{-1}\left(\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\right)' \\ &= \mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\mathbf{L}_{d,f}^{-1}\left(\mathbf{C}'_{\setminus f}\mathbf{C}_{\setminus f}\right)^{-1}\left(\mathbf{L}'_{d,f}\right)^{-1}\mathbf{L}'_{d,f}\mathbf{C}'_{\setminus f} \\ &= \mathbf{C}_{\setminus f}\left(\mathbf{C}'_{\setminus f}\mathbf{C}_{\setminus f}\right)^{-1}\mathbf{C}'_{\setminus f} = \mathbf{P}_{C\setminus f} = \mathbf{P}_C \end{aligned}$$

$$\begin{aligned} \mathbf{P}_{X\setminus d} &= \mathbf{X}_{\setminus d}\left(\mathbf{X}'_{\setminus d}\mathbf{X}_{\setminus d}\right)^{-1}\mathbf{X}'_{\setminus d} \\ &= \mathbf{P}_N + \left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\left\{\left[\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\right]'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\right\}^{-1}\left[\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\right]' \\ &= \mathbf{P}_N + \left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\left\{\left[\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\right]'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\right\}^{-1}\mathbf{L}'_{d,f}\mathbf{C}'_{\setminus f}\left(\mathbf{I}_n - \mathbf{P}_N\right)' \\ &= \mathbf{P}_N + \left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\left(\mathbf{L}_{d,f}\right)^{-1}\left\{\left[\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\right]'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\right\}^{-1}\left(\mathbf{L}'_{d,f}\right)^{-1}\mathbf{L}'_{d,f}\mathbf{C}'_{\setminus f}\left(\mathbf{I}_n - \mathbf{P}_N\right)' \\ &= \mathbf{P}_N + \left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\left\{\left[\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\right]'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\right\}^{-1}\left[\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\right]' \\ &= \mathbf{X}_{\setminus f}\left(\mathbf{X}'_{\setminus f}\mathbf{X}_{\setminus f}\right)^{-1}\mathbf{X}'_{\setminus f} = \mathbf{P}_{X\setminus f} = \mathbf{P}_X \end{aligned}$$

Replacing $\mathbf{P}_{X\setminus d}$ with $\mathbf{P}_X$ in (6): $\boldsymbol{\varepsilon}_{\setminus d} = \boldsymbol{\varepsilon}_{\setminus f} = \boldsymbol{\varepsilon}$. □

Theorem 1. *The estimated coefficients and variances associated with the non-compositional data* $\mathbf{N}$ *in (3) are permutation invariant:* $\boldsymbol{\beta}_{N\setminus d} = \boldsymbol{\beta}_{N\setminus f} = \boldsymbol{\beta}_N$ *and* $Var\left(\boldsymbol{\beta}_{N\setminus d}\right) = Var\left(\boldsymbol{\beta}_{N\setminus f}\right) = Var\left(\boldsymbol{\beta}_N\right)$ *for all* $d, f \in \{1, ..., D\}$.

Proof. Follows directly from (4) and (15) and Lemma 1. □

In the context of a cross-country growth model, all the estimated coefficients and t-statistics for the covariates associated with the Solow model (but not the intercept term) along with any other noncompositional covariates are invariant to which compositional variable is omitted from the regression.

Theorem 2. *The* R^2 *and* F*-test for the regression are invariant to which compositional variable is omitted from the regression.*

Proof. From Lemma 1 the error terms are permutation invariant:

$$R^2 = 1 - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\boldsymbol{y}'\left(\mathbf{I}_n - \mathbf{i}_n\left(\mathbf{i}_n'\mathbf{i}_n\right)^{-1}\mathbf{i}_n'\right)\boldsymbol{y}}$$

and:

$$F\left[K+D, n-K-D+1\right] = \frac{R^2/\left(K+D\right)}{\left(1-R^2\right)/\left(n-K-D+1\right)}$$

□

Theorem 3. *The F-test for joint hypothesis that for any choice* $d \in \{1,...,D\}$, $\boldsymbol{\beta}_{C\setminus d} = 0$ *is permutation invariant.*

Proof. $F\left[D+1, n-K-D+1\right]$

$$\begin{aligned}
&= \frac{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\left(\mathbf{C}_{\setminus d}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\right)^{-1}\mathbf{C}_{\setminus d}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\boldsymbol{\varepsilon}/\left(D+1\right)}{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_X\right)\boldsymbol{\varepsilon}/\left(n-K-D+1\right)} \\
&= \frac{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\left(\mathbf{L}_{d,f}'\mathbf{C}_{\setminus f}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\right)^{-1}\mathbf{L}_{d,f}'\mathbf{C}_{\setminus f}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\boldsymbol{\varepsilon}/\left(D+1\right)}{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_X\right)\boldsymbol{\varepsilon}/\left(n-K-D+1\right)} \\
&= \frac{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\left(\mathbf{L}_{d,f}\right)^{-1}\left(\mathbf{C}_{\setminus f}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\right)^{-1}\left(\mathbf{L}_{d,f}'\right)^{-1}\mathbf{L}_{d,f}'\mathbf{C}_{\setminus f}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\boldsymbol{\varepsilon}/\left(D+1\right)}{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_X\right)\boldsymbol{\varepsilon}/\left(n-K-D+1\right)} \\
&= \frac{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\left(\mathbf{C}_{\setminus f}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\right)^{-1}\mathbf{C}_{\setminus f}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\boldsymbol{\varepsilon}/\left(D+1\right)}{\boldsymbol{\varepsilon}'\left(\mathbf{I}_n - \mathbf{P}_X\right)\boldsymbol{\varepsilon}/\left(n-K-D+1\right)}
\end{aligned}$$

□

The implication of Theorems 1 through 3 is that the choice of which particular share to omit, among the set of compositional variables, has no bearing on many features of the model, at least as they relate to the coefficients associated with the noncompositional data and the overall goodness of fit. Moreover, Theorem 1 tells us that the explanatory power of the compositional variables as a set is not altered by which $D-1$ of the D variables we choose to include. However, the same cannot be said for the coefficients associated with the compositional data themselves, or the intercept term.

Theorem 4. *The estimated coefficients and variances associated with the compositional data* $\mathbf{C}$ *are generally not permutation invariant. Specifically,* $\boldsymbol{\beta}_{C\setminus d} = \mathbf{L}_{f,d}\boldsymbol{\beta}_{C\setminus f}$ *and* $Var\left(\boldsymbol{\beta}_{C\setminus d}\right) = \mathbf{L}_{f,d}\,Var\left(\boldsymbol{\beta}_{C\setminus f}\right)\mathbf{L}_{f,d}'$

Proof. The permutation invariance of $\boldsymbol{\beta}_N$ and $\boldsymbol{\varepsilon}$ together with equation (3) implies $\mathbf{C}_{\setminus d}\boldsymbol{\beta}_{C\setminus d} = \mathbf{C}_{\setminus f}\boldsymbol{\beta}_{C\setminus f}$. Premultiplying both sides by $\mathbf{C}_{\setminus d}'$ and solving for $\boldsymbol{\beta}_{C\setminus d}$:

$$\begin{aligned}
\boldsymbol{\beta}_{C\setminus d} &= \left(\mathbf{C}_{\setminus d}'\mathbf{C}_{\setminus d}\right)^{-1}\mathbf{C}_{\setminus d}'\mathbf{C}_{\setminus f}\boldsymbol{\beta}_{C\setminus f} \\
&= \left(\mathbf{C}_{\setminus d}'\mathbf{C}_{\setminus d}\right)^{-1}\mathbf{C}_{\setminus d}'\mathbf{C}_{\setminus d}\mathbf{L}_{f,d}\boldsymbol{\beta}_{C\setminus f} \\
&= \mathbf{L}_{f,d}\boldsymbol{\beta}_{C\setminus f}
\end{aligned}$$

Similarly:

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{\beta}_{C\setminus d}\right) &= \frac{1}{n-K}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\left(\mathbf{C}_{2\setminus d}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus d}\right)^{-1} \\
&= \frac{1}{n-K}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\left(\mathbf{L}_{d,f}'\mathbf{C}_{\setminus f}'\left(\mathbf{I}_n - \mathbf{P}_N\right)\mathbf{C}_{\setminus f}\mathbf{L}_{d,f}\right)^{-1} \\
&= \mathbf{L}_{d,f}^{-1}\mathrm{Var}\left(\boldsymbol{\beta}_{C\setminus f}\right)\left(\mathbf{L}_{d,f}'\right)^{-1}
\end{aligned}$$

and since $\mathbf{L}_{d,f}^{-1} = \mathbf{L}_{f,d}$:

$$\text{Var}\left(\boldsymbol{\beta}_{C\setminus d}\right) \quad = \quad \mathbf{L}_{f,d}\text{Var}\left(\boldsymbol{\beta}_{C\setminus f}\right)\mathbf{L}'_{f,d}$$

□

To understand the implications of the above, define for $i < d$ $[i > d]$, $\beta^i_{C\setminus d}$ as element i $[i-1]$ in $\boldsymbol{\beta}_{C\setminus d}$ corresponding in each case to variable $i \neq d$ in $\mathbf{C}$. For any $i \neq d$:

$$\left\{\beta^i_{C\setminus d}, \text{Var}\left(\beta^i_{C\setminus d}\right)\right\} = \begin{cases} \left\{\beta^i_{C\setminus f} - \beta^d_{C\setminus f}, \text{Var}\left(\beta^i_{C\setminus f} - \beta^d_{C\setminus f}\right)\right\}, & i \neq f, D+1 \\ \left\{-\beta^d_{C\setminus f}, \text{Var}\left(\beta^d_{C\setminus f}\right)\right\}, & i = f \\ \left\{\beta^{D+1}_{C\setminus f} + \beta^d_{C\setminus f}, \text{Var}\left(\beta^{D+1}_{C\setminus f} + \beta^d_{C\setminus f}\right)\right\}, & i = D+1 \end{cases} \tag{7}$$

Neither the estimated coefficients $\boldsymbol{\beta}_{C\setminus d}$ or the variances $\text{Var}\left(\boldsymbol{\beta}_{C\setminus d}\right)$ corresponding to the different variables in $\mathbf{C}$ are permutation invariant. As the different variables in $\mathbf{C}$ are not free to vary independently, the best way to interpret each element in the vector of coefficients $\boldsymbol{\beta}_{C\setminus d}$ in (4) is that each measures the effect of an increase in the value of the corresponding variable in $\mathbf{C}_{\setminus d}$ in relation to the excluded d^{th} variable.

In many applications, there may indeed be a particular component in $\mathbf{C}$ that one would naturally choose to exclude. For example, when analysing the effect of education on growth using data that measures the share of the population with different levels of educational attainment, it seems natural to exclude the share of people with no education and treat that as the baseline category as in Petrakis and Stamatakis (2002).

In other cases the choice of a baseline category to omit is not obvious. Lindh (1999) investigates the relationship between growth and the shares of the population that belong to different age cohorts, with children under 15 as the omitted category. Lindh and Malmberg (2009) also omit the share of children, but use logarithms of the remaining shares to reduce collinearity. Devarajan et al. (1996) and Bose et al. (2007) examine the relationship between growth and shares of disaggregated public expenditure. The omitted category changes depending on which subset of shares is included in the different specifications. Voigt et al. (2015) include two variables in their regressions, investment (public and private) and government consumption, as shares of GDP. Hence, private consumption expenditure as a share of GDP is the omitted category. Hall and Jones (1999) and Rodrik et al. (2004) both include shares of the population that are native English speakers or, separately, any of four other major European languages (French, German, Portuguese, or Spanish), leaving the share of the population speaking a language other than these five as the omitted category. Putterman and Weil (2010) regress log GDP per capita on eleven ancestral regions, with sub-Saharan Africa as the omitted category.

Barro (1996) uses a seven-religion breakdown of world religions to examine the impact of religion on democracy. After choosing the fraction of Catholics as the omitted variable, he finds that only the fraction of Hindus has a significant (positive) effect on democracy. Sala-i-Martin (1997) and Sala-i-Martin et al. (2004) test the robustness of a large set of regressors as explanatory variables in a growth regression, including Barro's breakdown of religions. The fraction of the population practicing Confucianism, Buddhism, and Islam are all significant

and positive, while Protestantism and Catholicism are significant and negative. Hall and Jones (1999) estimate the impact of four religions, Catholicism, Hinduism, Islam and Protestantism, measured again as population shares on output per worker. Only the variables associated with Catholicism and Islam are statistically significant.

Noland (2005) considers the effect of seven categories of religious affiliation (Catholicism, Protestantism, Orthodox Christianity, Islam, Judaism, Hinduism, and Buddhism) on total factor productivity and economic growth, and finds a statistically significant effect for Catholicism, Protestantism and Judaism on the latter. As he states, other religions and the category of nonreligious "are omitted from the regression (i.e., are absorbed in the constant) and are the standard against which the included major world religions are judged." Similarly, both Barro and McCleary (2003) and McCleary and Barro (2006) consider the effect of eight religious categories—Catholicism, Protestantism, Orthodox Christianity, Islam, Judaism, Hinduism, Eastern religions (including Buddhism), and other religions—on both religious practice and growth. Barro and McCleary (2003) find statistically significant negative coefficients for economic growth associated with Hinduism, Islam, Orthodox Christianity, and Protestantism, and McCleary and Barro (2006) for shares of adherents to Islam and Protestantism. As in each case Catholicism is the excluded category, they state clearly that "each coefficient should be interpreted as the relationship with the indicated religion share, measured relative to the Catholic share" McCleary and Barro (2006)).

While these later papers are careful to state that the coefficients can only be interpreted in relation to the excluded category and any forecast generated by (3) is permutation invariant, this does not solve the problem of how to interpret their statistical significance. Theorem 4 shows that we can use the coefficient values associated with a regression that excludes category d, $\boldsymbol{\beta}_{C\setminus d}$, to derive the coefficient values associated with a regression that excludes category f, $\boldsymbol{\beta}_{C\setminus f}$. However the variances of the corresponding coefficients alone, the diagonal components of $\text{Var}\left(\boldsymbol{\beta}_{C\setminus d}\right)$—all that is usually reported in empirical research—is insufficient to derive the diagonal components of $\text{Var}\left(\boldsymbol{\beta}_{C\setminus f}\right)$.[3]

Theorem 3 states that regardless of which category is omitted, the overall statistical significance of the remaining categories as a group is unaltered. Nonetheless, while two permutation may include the same category i, and from (7) $\frac{\beta^i_{C\setminus d}}{\text{Var}\left(\beta^i_{C\setminus d}\right)} = \frac{\beta^i_{C\setminus f}-\beta^d_{C\setminus f}}{\text{Var}\left(\beta^i_{C\setminus f}-\beta^d_{C\setminus f}\right)}$, these do not equal $\frac{\beta^i_{C\setminus f}}{\text{Var}\left(\beta^i_{C\setminus f}\right)}$. Hence the t-statistics and p-values for the same category i can differ greatly across the different permutations. This leaves a researcher the freedom to choose the permutation that appears most convincing, one that perhaps yields the most p-values that cross a desired threshold of significance, or maximises the number of "stars", in a manner analogous to someone engaged in p-hacking as described in Brodeur et al. (2016).

To demonstrate just how different the model may appear depending on which category is excluded, I combine data on output, savings rates and population from the IMF with data on

3. Since $\text{Var}\left(\beta^i_{C\setminus f}-\beta^d_{C\setminus f}\right)$ does not generally equal $\text{Var}\left(\beta^i_{C\setminus f}\right)$-$\text{Var}\left(\beta^d_{C\setminus f}\right)$, to derive the associated variances of these coefficients requires the full variance covariance matrix of the coefficients.

religious affiliation from the World Religion Project (WRP) database (Maoz and Henderson (2013)). The dependent variable is the log difference in per-capita purchasing power parity gross domestic product between the years 2001 and 2020. The explanatory variables associated with the Solow growth model, the logarithm of per-capita GDP (PPP) in the year 2001, and both the average rate of savings and rate of population growth between 2001 and 2020 correspond to the matrix $\mathbf{N}$.[4] I consolidate the different religious affiliations in the WRP database into D=10 religious categories in the following order: Catholicism, Protestantism, Orthodox Christianity, Other Christian Denominations, Islam, Judiasm, Buddhism, Hinduism, Eastern Religions (Confucianism, Shintoism and Taoism), and a last category of Other which includes those not in the previous nine categories (including Sikhism, Zoroastrianism, Bahaism, Jainism, Animism and the non-religious), each measured as a share of the population in each country. These shares, along with a column of ones, correspond to matrix $\mathbf{C}$. The ten columns in Table 1 each present one of the $D = 10$ different permutations of the regression, as each religious category is successively excluded.

The results in the first three rows of Table 1 are consonant with the standard predictions of conditional convergence of the Solow growth model; the coefficient on log GDP in levels is negative—implying conditional convergence—as are the coefficients on savings and population growth, all at a statistically significant p-level below 0.01. Moreover, as Theorems 1 and 2 imply, as each successive religious category is excluded, the coefficient estimates and standard errors associated with the noncompositional data in each of the estimated equations, along with the R^2, remain identical.

Column 1 of Table 1 excludes the Catholic share of the population. Column 2 reinstates the share of Catholics but excludes Protestants, and so forth. In accordance with Theorem 4, the coefficients for each religious category in any particular column d in Table 1 are identical to the values of the coefficients in the row corresponding to the d^{th} religious category, with the sign inverted, along with the same standard deviation. So though in Table 1 there are $D \times (D-1)$=90 estimated coefficients associated with the $D = 10$ different religious categories, along with their corresponding standard errors, the half on one side of the diagonal of excluded categories is the mirror, inverted-sign image of the other half. Though the coefficients, and particulary the pattern of statistical significance, may appear very different, each permutation represents a different representation of the same underlying statistical model.

4. Following Mankiw et al. (1992), we assume the annualised value is $g + \delta = 0.05$. As the dependent variable is not annualised, we multiply this by twenty—we add 1 to the change in population between 2001 and 2020, corresponding to the term in (1).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP (2001)	−0.167***	−0.167***	−0.167***	−0.167***	−0.167***	−0.167***	−0.167***	−0.167***	−0.167***	−0.167***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Pop. Growth	−0.957***	−0.957***	−0.957***	−0.957***	−0.957***	−0.957***	−0.957***	−0.957***	−0.957***	−0.957***
	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)	(0.150)
Savings	0.098**	0.098**	0.098**	0.098**	0.098**	0.098**	0.098**	0.098**	0.098**	0.098**
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
Catholic		0.175	−0.204	0.158	0.012	−0.378	−0.555***	0.657	−0.230	−0.139
		(0.139)	(0.135)	(0.359)	(0.091)	(0.410)	(0.163)	(0.553)	(0.229)	(0.180)
Protestant	−0.175		−0.380**	−0.017	−0.164	−0.554	−0.730***	0.482	−0.405	−0.315
	(0.139)		(0.163)	(0.388)	(0.126)	(0.419)	(0.186)	(0.559)	(0.249)	(0.212)
Orthodox	0.204	0.380**		0.363	0.216	−0.174	−0.350*	0.861	−0.025	0.065
	(0.135)	(0.163)		(0.375)	(0.152)	(0.425)	(0.190)	(0.560)	(0.250)	(0.198)
Other Christ.	−0.158	0.017	−0.363		−0.147	−0.537	−0.713*	0.499	−0.388	−0.298
	(0.359)	(0.388)	(0.375)		(0.353)	(0.534)	(0.376)	(0.656)	(0.421)	(0.414)
Muslim	−0.012	0.164	−0.216	0.147		−0.390	−0.566***	0.645	−0.241	−0.151
	(0.091)	(0.126)	(0.152)	(0.353)		(0.411)	(0.165)	(0.558)	(0.233)	(0.175)
Jewish	0.378	0.554	0.174	0.537	0.390		−0.176	1.035	0.148	0.239
	(0.410)	(0.419)	(0.425)	(0.534)	(0.411)		(0.434)	(0.680)	(0.460)	(0.437)
Buddhist	0.555***	0.730***	0.350*	0.713*	0.566***	0.176		1.212**	0.325	0.415*
	(0.163)	(0.186)	(0.190)	(0.376)	(0.165)	(0.434)		(0.607)	(0.275)	(0.231)
Other Eastern	−0.657	−0.482	−0.861	−0.499	−0.645	−1.035	−1.212**		−0.887	−0.796
	(0.553)	(0.559)	(0.560)	(0.656)	(0.558)	(0.680)	(0.607)		(0.590)	(0.580)
Hindu	0.230	0.405	0.025	0.388	0.241	−0.148	−0.325	0.887		0.090
	(0.229)	(0.249)	(0.250)	(0.421)	(0.233)	(0.460)	(0.275)	(0.590)		(0.263)
Other	0.139	0.315	−0.065	0.298	0.151	−0.239	−0.415*	0.796	−0.090	
	(0.180)	(0.212)	(0.198)	(0.414)	(0.175)	(0.437)	(0.231)	(0.580)	(0.263)	
Constant	2.053***	1.877***	2.257***	1.894***	2.041***	2.431***	2.607***	1.396**	2.282***	2.192***
	(0.204)	(0.226)	(0.208)	(0.398)	(0.204)	(0.465)	(0.233)	(0.606)	(0.296)	(0.240)
Observations	164	164	164	164	164	164	164	164	164	164
R^2	0.509	0.509	0.509	0.509	0.509	0.509	0.509	0.509	0.509	0.509
Adjusted R^2	0.469	0.469	0.469	0.469	0.469	0.469	0.469	0.469	0.469	0.469
Mean VIF	1.34	1.82	1.91	8.29	1.31	10.54	2.36	18.73	3.86	2.64

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Omitted category regressions (3). Dependent variable is log difference of per-capita GDP (PPP) between 2001 to 2020.

Yet given that only one permutation is usually chosen to be reported, that choice can have profoundly different implications on how the results are likely to be interpreted. In column 1, the coefficient for the share of Orthodox Christians, relative to the excluded category of Catholics, in not statistically significant, but in column 2 it is significant, at the 5% level, relative to the excluded category of Protestants. The share of Other Christians is only statistically significant (at the 10% level) in column 7, when the Buddhist category is omitted.

Moreover, a comparison of columns 6 or 9 with 7 produces a startling contrast. The Jewish and Hindu shares are not statistically significant in any of the regressions. This helps explain why in columns 6 and 9, when, respectively, Jews or Hindus are excluded, the other religious categories have no statistically significant relationship with economic growth, when compared to the omitted Jewish category. By Theorem 3, the joint hypothesis test on all the different religious categories is permutation invariant and each yields the same value of $F(9, 151)$=2.43, corresponding to a p-value of 0.013.[5] Yet someone reading an empirical study that reports the permutations in columns 6 or 9 alone might well conclude that including the shares of religious categories is of marginal value in explaining cross-country differences in growth rates, and that the particular coefficient estimates for the share of adherents of each religion, can be safely ignored.

By contrast, the share of Buddhists in seven of the ten columns in Table 1 are positive and statistically significant. That is why in column 7, when the Buddhist share is the omitted category, the coefficients for all the remaining categories, aside from the share of Jews and Hindus, are negative and significant at the 10% level or less. The coefficient for Other Eastern Religions is -1.212 and is statistically significant at the 5% level (p-value of 0.048), so a change in the share of these adherents relative to the share of Buddhists could potentially account for a sizable difference in growth rates. Similarly, the coefficients for the shares of Catholics, Protestants and Muslims vary in size between -0.555 and -0.730 but are statistically significant at p-levels of 0.001 or lower.

Looking at Table 1 as a whole, one might conclude that between the years 2001 and 2020, *ceteris paribus*, countries with higher shares of Buddhists enjoyed higher economic growth. While overall, growth was lower in countries with a higher share of Christians, this was less so for the Orthodox relative to other Christian denominations. Furthermore the share of Jews in the population seems completely unrelated to growth—even when Buddhists are the excluded category—and nothing conclusive can be said about the remaining categories. To avoid these problems we might adopt an alternative strategy and experiment with including some number less than $D-1$ of the categories. But which, and how many? The total number of possible regressions expands from merely D to equal 2^D-2 different possibilities, which for D=10 is 1022 possible regressions to choose from.

The larger point remains. It is hard to point to one particular category that can serve as a natural choice for a baseline case, and the column with the most "stars" will inevitably look

5. Including the religious categories raises the adjusted R^2 from 0.429 to 0.469.

the most convincing. Excluding the eight "stars" associated with the Solow model, and the intercept term, the number of "stars" associated with compositional data varies widely from zero in the case of columns 6 and 9, where Jews or Hindus are the omitted category, to fourteen across the nine categories that remain, when Buddhists are excluded. Finally, the intercept term changes with every permutation and sometimes, as in the case of the growth regression in (3), the constant term does convey useful information—for example if we wish to isolate the value of the level of technology, A.

In some contexts, one might choose to include the compositional dataset to serve merely as a control. Permutation invariance matters less if knowing the values of the different coefficients themselves is not perceived to be important. If that is the case, and the intercept term does not convey useful information, the standard approach might seem appropriate. In the next section we consider more reasons for adopting an alternative based on logarithmic ratios.

4 Vector Spaces, Distorted Distances and Angles

The common practice of excluding one share in a composition to circumvent the singularity problem, and then including the remaining variables as raw shares, creates more problems than the lack of permutation invariance described in Section 3. Doing so means we treat compositional data as if they are real coordinates relative to a canonical basis to which we can apply the usual Euclidean geometry. Unfortunately, applying standard Euclidean measures of distance and angle to points in a simplex generates a distorted relationship between the coordinates. Furthermore, these are not points in a vector space and so, particularly if they fall near the edges of the simplex, can be easily misused to generate incorrect or misleading counterfactual experiments.

Suppose for argument's sake we thought that the link between economic growth and the share of adherents of the various religious categories as represented by the vector of coefficients $\boldsymbol{\beta}_{C\setminus d}$ from (3) in Table 1 implies not merely a statistical correlation, but captures some causal relationship as well. Given the strong evidence of a positive relationship between the share of Buddhists in our sample and economic growth, we might then ask how much economic growth might change overall if the number of Buddhists in every country increased or decreased by one unit, perhaps offset by changes in the share of Muslims. In 103 of the 164 countries in our sample, the share of Buddhists is zero.[6] In sixty countries the share of Buddhists falls between zero and

6. Albania, Algeria, Andorra, Antigua and Barbuda, Armenia, Austria, Azerbaijan, Bahamas, Bahrain, Belarus, Benin, Bosnia and Herzegovina, Bulgaria, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Colombia, Comoros, Congo, Croatia, Cuba, Cyprus, Czech Republic, Djibouti, Dominica, Dominican Republic, Egypt, El Salvador, Equatorial Guinea, Eritrea, Ethiopia, Fiji, Gabon, Gambia, Georgia, Ghana, Greece, Grenada, Guatemala, Guinea-Bissau, Guyana, Haiti, Honduras, Hungary, Iran, Iraq, Israel, Jordan, Kazakstan, Kenya, Kosovo, Kuwait, Kyrgyzstan, Lesotho, Liberia, Libya, Lithuania, Luxembourg, Macedonia, Malawi, Mali, Malta, Marshall Islands, Mauritania, Mexico, Moldova, Monaco, Morocco, Namibia, Nauru, Nicaragua, Niger, Poland, Romania, Russia, Rwanda, Saint Kitts and Nevis, Saint Vincent and the Grenadines, Saint Lucia, San Marino, Sao Tome and Principe, Saudi Arabia, Serbia and Montenegro, Sierra Leone, Slovakia,

half of one percent.[7] In Afghanistan and Somalia, Muslims constitute more than 99.5% of the population, and in 45 of our sample of 164 countries, less than half of one percent.[8] What is the meaning of a counterfactual that posits the impact associated with raising the share of Muslims and lowering the share of Buddhists by the same amount, if the resulting share of Muslims in Afghanistan and Somalia rises above 100% and the share of Buddhists below zero? How would we interpret the effect of raising the share of Muslims, Buddhists or Catholics and decreasing the shares of Hindus, Jews or Orthodox Christians if the database records that there are no adherents of the latter three religions in 42 countries? And note that in this application, we are not in fact asserting that the relationship between the dependent variable and the explanatory compositional variables is necessarily a causal one. For applications in which the relationship is explicitly causal, this problem greatly limits the usefulness of a regression formulated this way. Confidence intervals, associated with compositional data can also fall outside the simplex.

The underlying issue is that the familiar Euclidean geometry applied to a simplex does not generate a vector space and ordinary operations such as adding two vectors in the simplex, as if they were Cartesian coordinates, or multiplying one vector by a scalar, can yield coordinates that fall outside the simplex. To overcome these limitations, Aitchison (1986) developed the concepts of perturbation and powering (as defined in Appendix A.1), which respectively take the place of vector addition and scalar multiplication. Billheimer et al. (2001) demonstrates that these two concepts are sufficient to make the simplex a vector space (the point where all shares are equal is the barycentre). By adding the Aitchison inner product and norm in Appendix A.1 the simplex is also turned into a Hilbert space (a complete, inner product space). Aitchison distances and angles follow directly from these.

To visualise the implications of using Aitchison geometry, we amalgamate our ten religious categories into three broader categories by combining Catholics, Protestants, Orthodox Christians and Other Christians into a single Christian category, and then combining all the remaining religions, save Islam, together into the expanded category of Other Religions, so that it now encompasses everyone who is not Muslim or Christian. The composition of amalgamated religious shares for each of the 164 countries in our dataset are plotted on the two dimensional ternary diagram in Figure 2.

Slovenia, Somalia, Sudan, Suriname, Swaziland, Sweden, Syria, Togo, Trinidad and Tobago, Tunisia, Turkey, Tuvalu, Uruguay, Uzbekistan, Yemen, Zimbabwe.

7. Afghanistan, Angola, Argentina, Barbados, Belgium, Belize, Bolivia, Botswana, Brazil, Chad, Chile, Costa Rica, Cote d'Ivoire, Democratic Republic of Congo, Denmark, East Timor, Ecuador, Estonia, Finland, Germany, Guinea, Iceland, Ireland, Italy, Jamaica, Kiribati, Latvia, Liechtenstein, Madagascar, Maldives, Mauritius, Montenegro, Mozambique, Nigeria, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Samoa, Senegal, Seychelles, Solomon Islands, South Africa, Spain, Switzerland, Tajikistan, Tanzania, Tonga, Turkmenistan, Uganda, Ukraine, United Arab Emirates, United Kingdom, Vanuatu, Venezuela, Zambia.

8. Antigua and Barbuda, Armenia, Belarus, Belize, Bolivia, Brazil ,Chile, Colombia, Costa Rica, Czech Republic, Dominica, Dominican Republic, Ecuador, El Salvador, Estonia, Finland, Grenada, Guatemala, Haiti, Honduras, Hungary, Jamaica, Japan, Korea, Latvia, Lesotho, Lithuania, Mexico, Namibia, Nicaragua, Panama, Paraguay, Peru, Poland, Romania, Slovak Republic, Solomon Islands, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Taiwan, The Bahamas, Uruguay, Venezuela, Vietnam.

The orthogonal basis of the ternary diagram is represented by the straight dashed line that connects the vertex for the expanded category of Other Religions, with a point on the edge midway between Christian and Muslim, and the dashed curve that connects the vertices for Christians and Muslims. The latter is in fact a straight line when generated using Aitchison geometry or plotted on the surface of Figure 1b, it only appears curved on the two dimensional projection of that surface in Figure 2.

We choose three sets of four coordinates in Figure 2 and connect them with straight lines to generate three polygons. As with the orthogonal basis, the straight lines connecting the coordinates appear curved on the two dimensional surface. These edges between the vertices are nearly perpendicular, and so the polygons are nearly rectangular.[9] Estimating (3) with just the shares of Christians and Muslims is akin to treating the compositional data as the coordinates we see in Figure 3. Translating the orthogonal basis that connects the two vertices in Figure 2 to this Cartesian graph results in a curve, not a straight line. The lines connecting the polygons in Figure 2 are curved as well. Even if we draw conventional straight lines between the vertices, the resulting quadrilateral shapes are not rectangular or even parallelograms—this simple "naive" transformation from $\mathbb{S}^D$ to $\mathbb{R}^{D-1}$ distorts both angles and distances.

One final issue relates to how much comparing coordinates in terms of their Euclidean distance offers an appropriate insight into how they truly differ. For example, according to the WRP database, the Muslim share of Italy's population in 2010 was just over 1%, whereas in neighboring France it was 8%, having risen from only half of one percent in 1960. That difference of seven percent is identical to the difference between the Muslim share of the population in Oman, 90%, and Jordan, 97%. Yet, it is hard to see these as equivalent. The eight fold difference between France and Italy reflects the former's much longer and wider colonial presence throughout the Middle East and North Africa. By comparison, when comparing two countries with an overwhelming Muslim majority, the salience of the additional 7% Muslim share in Jordan is less meaningful. Similarly, in the last few decades, the share of Protestants has grown rapidly in many (historically Catholic) Latin American countries, most notably in the largest, Brazil, where they now constitute 27% of the population. The experience is similar in Colombia, though the growth has not been nearly as rapid; Protestants there constitute 15% as of 2010. By contrast, this has not been the experience in neighboring Ecuador, where Protestants make up only 2% of the population. For centuries, Lutheran Protestantism was the state religion in all five Scandinavian countries. In 2010, Protestants comprised 81% of the population in both Denmark and Norway and 68% in Sweden, the same difference in shares as between Brazil and

9. The angles associated with the vertices in the polygon in green are Chad: 89.86° (0.499π), Nigeria: 90.19° (0.501π), Zambia: 90.96° (0.505π), Eswatini: 88.99° (0.494π). The ratio of the area of the polygon to the area of the minimum bounding rectangle is 0.9950. The angles associated with the vertices in the polygon in red are Portugal: 92.66° (0.515π), Ireland: 87.30° (0.485π), Bosnia: 89.86° (0.499π), Lebanon: 90.17° (0.501π). The ratio of the area of the polygon to the area of the minimum bounding rectangle is 0.9944. The angles associated with the vertices in the polygon in blue are Togo: 89.18° (0.495π), Korea: 89.40° (0.497π), Lithuania: 90.50° (0.503π), Rwanda: 90.92° (0.505π). The ratio of the area of the polygon to the area of the minimum bounding rectangle is 0.9924.

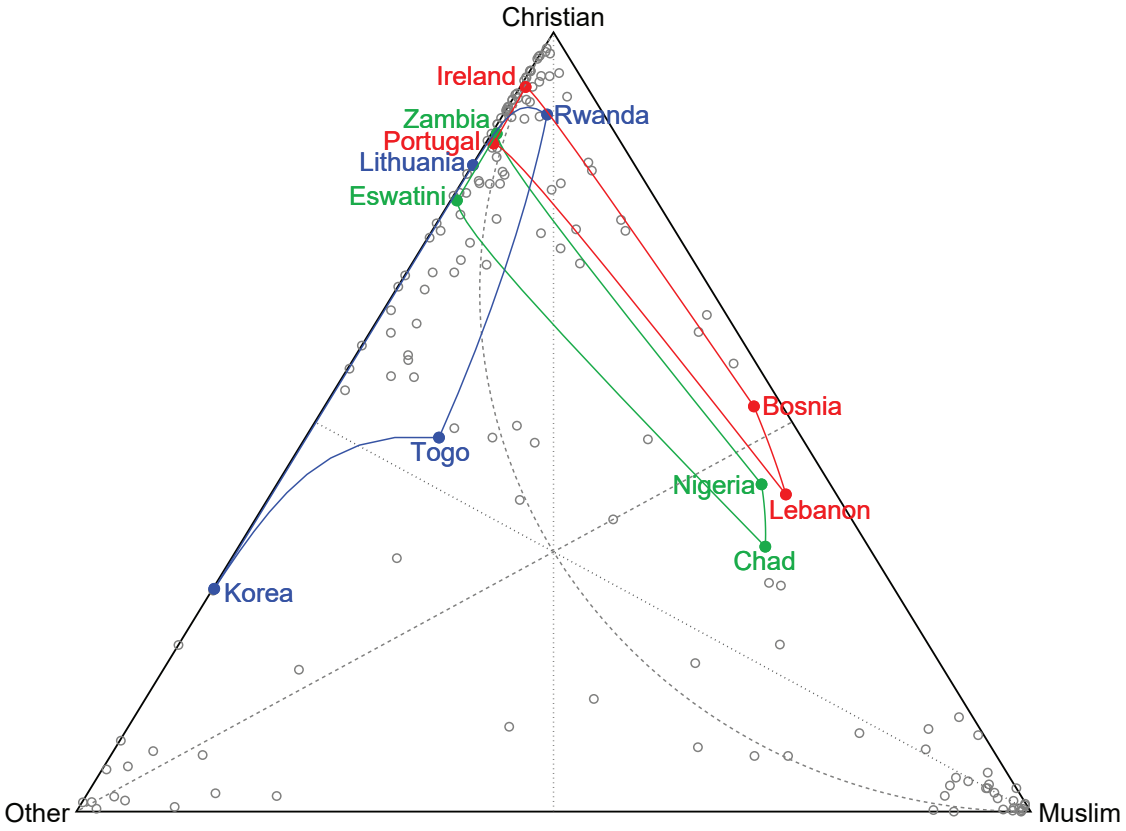


Figure 2: Ternary diagram for three part religious composition for the 164 countries in the dataset. Dashed curves represent the basis for ILR coordinates and dotted curves the basis for ALR coordinates. Curves in Blue, Green and Red represent the sides of rectangles in terms of Aitchison geometry.

Colombia, and nearly the same as between Colombia and Ecuador. Again, it is hard to see these differences as equivalent:the Protestant share of Brazil's population is more than thirteen times larger than Ecuador's. Yet the coefficients in Table 1 are estimated in a way that implicitly assumes they are.

Inner products and distances associated with Aitchison geometry rely on logarithmic differences, which mitigate many of these issues. In the next sections, I demonstrate how using logarithmic ratios can turn compositional data into coordinates in a vector space, which can then be incorporated into regressions that are permutation invariant.

5 Additive Logarithmic Ratios

One way forward is to use a logarithmic transformation to translate the points in the simplex $\mathbb{S}^D$ in (2) to Euclidean space $\mathbb{R}^D$. Before doing so we must first replace any zeros in $\mathbf{CQ}_{D+1}$

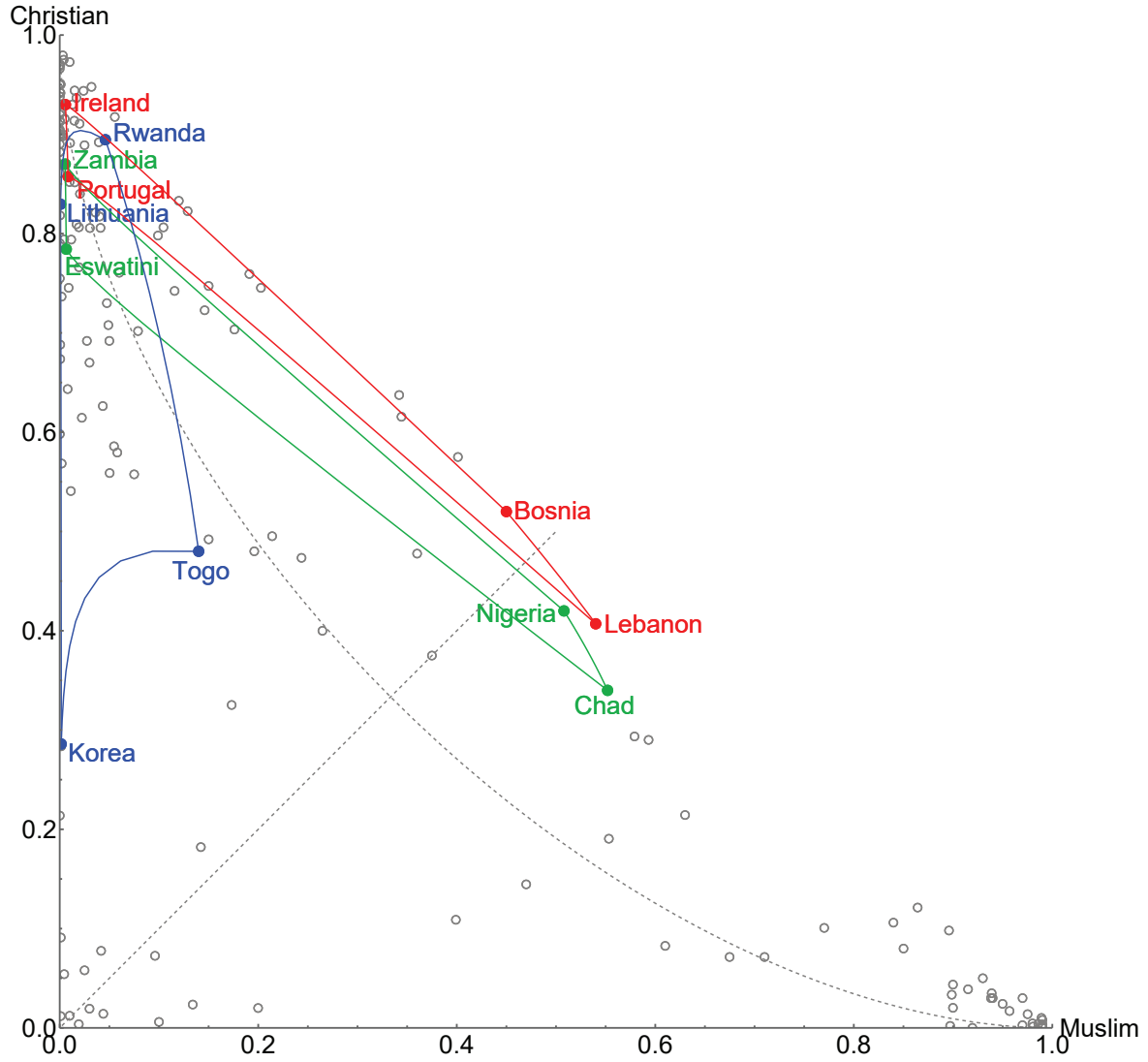


Figure 3: "Naive" transformation of the three part composition in Figure 2 to a canonical basis. The omitted category is Other Religions—the higher the share of adherents of religions other than Christianity and Islam in the population, the closer it is to the origin of the plot.

with small numbers and adjust the remaining data so that it sums to one.[10] Define the natural logarithm of all the elements in the matrix of compositional data $\mathbf{CQ}_{D+1}$, along with a column

10. Following Aitchison (1986), p.269, I set the Z zero components for each observation equal to $\delta(Z+1)(D-Z)/D^2$ and the positive elements are reduced by $\delta Z(Z+1)/D^2$. Setting δ equal to 1.0×10^{-5} ensures that the fraction of the population measured as zero are replaced by values of between 1.0×10^{-6} and 3.0×10^{-6}, in the latter case, no more than 3 per million people. Since the lowest nonzero fraction in the dataset is 100 per million and $D = 10$, this procedure reduces those observations by 2 per million if $Z = 1$, and 9 per million if $Z = 9$. This is to avoid logarithms of zero, but is also consistent with Cromwell's Rule as stated by the Bayesian statistician Dennis Lindley (1985), p. 104: "So leave a little probability for the moon being made of green cheese; it can be as small as 1 in a million, but have it there since otherwise an army of astronauts returning with samples of the said cheese will leave you unmoved. . . So never believe in anything absolutely, leave some room for doubt: as Oliver Cromwell told the Church of Scotland, 'I beseech you, in the bowels of Christ, think it possible you may be mistaken'."

of ones, as $\check{\mathbf{C}} = \left[\ln\left(c_{i,j}\right)_{i=1,...,D;j=1,...,n} : \mathbf{i}_n\right]$. Though all the shares are represented in $\check{\mathbf{C}}$ we can regress $\boldsymbol{y}$ on $\mathbf{N}$ and $\check{\mathbf{C}}$:

$$\boldsymbol{y} = \mathbf{N}\check{\boldsymbol{\beta}}_N + \check{\mathbf{C}}\check{\boldsymbol{\beta}}_C + \check{\boldsymbol{\varepsilon}}. \tag{8}$$

The values of $\check{\boldsymbol{\beta}}_{\boldsymbol{C}}$ are elasticities, and since all the categories are included, permutation invariance is no longer an issue. However, unless we restrict the parameter estimates (see below), there is a conceptual problem in interpreting how the D coefficients of $\check{\boldsymbol{\beta}}_{\boldsymbol{C}}$ associated with the shares affect the dependent variable. While changing the units from fractions to percentages does not change the corresponding coefficient estimates, the value of the constant term does change. More importantly, there is nothing to distinguish $\check{\mathbf{C}}$ from a set of noncompositional values and no corresponding restriction on the values of $\check{\boldsymbol{\beta}}_{\boldsymbol{C}}$.[11]

To overcome this problem Aitchison (1982) suggests replacing $\check{\mathbf{C}}$ with an $n \times D-1$ matrix of additive log-ratios (ALRs), the logarithmic differences between any $D-1$ of the first D components in $\mathbf{C}$ and the missing d^{th} component $\mathbf{c}_d$:

$$\begin{aligned}&\left[\check{\mathbf{c}}_1 - \check{\mathbf{c}}_d, \check{\mathbf{c}}_2 - \check{\mathbf{c}}_d, ..., \check{\mathbf{c}}_{d-1} - \check{\mathbf{c}}_d, \check{\mathbf{c}}_{d+1} - \check{\mathbf{c}}_d, ..., \check{\mathbf{c}}_D - \check{\mathbf{c}}_d\right] \\ &\quad = \left[\log\left(\mathbf{c}_1 \oslash \mathbf{c}_d\right), \log\left(\mathbf{c}_2 \oslash \mathbf{c}_d\right), ..., \log\left(\mathbf{c}_{d-1} \oslash \mathbf{c}_d\right), \log\left(\mathbf{c}_{d+1} \oslash \mathbf{c}_d\right), ..., \log\left(\mathbf{c}_D \oslash \mathbf{c}_d\right)\right] \end{aligned} \tag{9}$$

where $\oslash$ represents Hadamard division and the logarithm is taken for each element in the vector, translating the points in the simplex $\mathbb{S}^D$ in (2) to the Euclidean space $\mathbb{R}^{D-1}$.

Taking the exponentials of each component in the canonical basis and then normalising so that each vector sums to one (applying the closure operator $\mathcal{C}$ as in Appendix A.1), we generate the $D \times D$ matrix $\mathbf{W}$ whose D columns $\mathbf{w}_1, ..., \mathbf{w}_D$ are the elements:

$$w_{i,j} = \begin{cases} \frac{e}{e+D-1} & i = j \\ \\ \frac{1}{e+D-1} & i \neq j \end{cases} \tag{10}$$

Then, in terms of Aitchison geometry, the matrix of compositional data can be expressed as:

$$\begin{aligned}\mathbf{CQ}_{D+1} &= \left[\left(\check{\mathbf{c}}_{1,i} \odot \mathbf{w}_1\right) \oplus \left(\check{\mathbf{c}}_{2,i} \odot \mathbf{w}_2\right) \oplus ... \oplus \left(\check{\mathbf{c}}_{D,i} \odot \mathbf{w}_D\right)\right] \\ &= \left[\left(\left(\check{\mathbf{c}}_{1,i} - \check{\mathbf{c}}_{d,i}\right) \odot \mathbf{w}_1\right) \oplus \left(\left(\check{\mathbf{c}}_{2,i} - \check{\mathbf{c}}_{d,i}\right) \odot \mathbf{w}_2\right) \oplus ... \oplus \left(\left(\check{\mathbf{c}}_{d-1,i} - \check{\mathbf{c}}_{d,i}\right) \odot \mathbf{w}_{d-1}\right) \oplus \left(\left(\check{\mathbf{c}}_{d+1,i} - \check{\mathbf{c}}_{d,i}\right) \odot \mathbf{w}_{d+1}\right)\right. \\ &\qquad \left. \oplus ... \oplus \left(\left(\check{\mathbf{c}}_{D,i} - \check{\mathbf{c}}_{d,i}\right) \odot \mathbf{w}_D\right)\right], \qquad i = 1, ..., n \end{aligned} \tag{11}$$

where $\oplus$ represents powering (as defined in Appendix A.1)—the equivalent in Aitchison geometry of scalar multiplication. The second equality means that the set of $D-1$ vectors $\left[\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{d-1},\right.$
$\left.\mathbf{w}_{d+1}, ..., \mathbf{w}_D\right]$ are a basis for the additive log-ratios (9). For the case of D=3, the two basis vectors are represented by the dotted lines in the ternary diagram in Figure 2.

11. Estimating (8), the sum of the coefficients associated with the ten religious categories is 0.0331.

The dependent variable in the regression is not necessarily compositional, and we are concerned with models that include both compositional and noncompostional explanatory variables. To make this more applicable to econometric applications and compatible with Section 3, we can transform the compositional data into ALRs using matrices so that the regression is expressed in a way that is compatible with conventional Euclidean geometry.

Define the $D \times D-1$ and $D+1 \times D$ matrices:

$$\mathbf{F}_d = \begin{bmatrix} \mathbf{I}_{d-1,d-1} & \mathbf{0}_{d-1,D-d-1} \\ -\mathbf{i}'_{D-1} & \\ \mathbf{0}_{D-d,d-1} & \mathbf{I}_{D-d,D-d} \end{bmatrix}, \qquad \tilde{\mathbf{F}}_d = \begin{bmatrix} \mathbf{F}_d & \mathbf{0}_{D,1} \\ \mathbf{0}_{1,D} & 1 \end{bmatrix}. \tag{12}$$

The transpose $\mathbf{F}'_d$ is analogous to the matrix $\mathbf{F}$ in Aitchison (1986) where it generates additive log-ratios by premultiplying the log compositional data alone, i.e.: $(\check{\mathbf{C}}\mathbf{Q}_{D+1})$. The added row and column in $\tilde{\mathbf{F}}'_d$ accommodate the all-ones vector in $\check{\mathbf{C}}$ that generates the intercept term. Postmultiplying $\check{\mathbf{C}}$ by $\bar{\mathbf{F}}_d$ generates an $n \times D$ matrix $\tilde{\mathbf{C}}_{/d} = \check{\mathbf{C}}\tilde{\mathbf{F}}_d$, where each column is the log deviation of the remaining $i \neq d$ columns from the d^{th} variable, and the last column is a $D+1$ all-ones vector.[12] Augmenting the regression with ALR coordinates first suggested by Aitchison and Bacon-Shone (1984) to include noncompositional covariates, we can now regress:

$$\boldsymbol{y} = \mathbf{N}\tilde{\boldsymbol{\beta}}_{N/d} + \tilde{\mathbf{C}}_{/d}\tilde{\boldsymbol{\beta}}_{C/d} + \tilde{\boldsymbol{\varepsilon}}_{/d}, \qquad d \in \{1, ..., D\}. \tag{13}$$

How then do the D possible permutations of (13) differ? Define the $D \times D$ matrix:

$$\mathbf{M}_{d,f} = \begin{cases} \mathbf{K}_d\mathbf{S}_{d+1}\mathbf{S}_d\mathbf{S}_{d-1}...\mathbf{S}_{f-1} & \text{for all } d < f \\ \mathbf{K}_{d-1}\mathbf{S}_{d-1}\mathbf{S}_{d-2}...\mathbf{S}_{f+1} & \text{for all } d > f \end{cases}$$

where:

$$\mathbf{K}_j = \begin{bmatrix} \mathbf{I}_{j-1,j-1} & \mathbf{0}_{j-1,D-j-1} \\ -\mathbf{i}'_{D-1} & 0 \\ \mathbf{0}_{D-j-1,j-1} & \mathbf{I}_{D-j-1,D-j-1} \end{bmatrix}.$$

Postmultiplying any $\tilde{\mathbf{C}}_{/d}$ by $\mathbf{K}_j$ transforms the log deviations with respect to variable d, into log deviations with respect to variable j, and postmultipling the result by the permutation matrix $\mathbf{S}_j$, exchanges the j and j-1 columns in the new matrix. Hence for all $d, f \in \{1, ..., D\}$ where $d \neq f$: $\tilde{\mathbf{C}}_{/d} = \tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}$, $\tilde{\mathbf{C}}_{/f} = \tilde{\mathbf{C}}_{/d}\mathbf{M}_{f,d}$ and $\mathbf{M}_{d,f}=\left(\mathbf{M}_{f,d}\right)^{-1}$. Define two projection matrices $\tilde{\mathbf{P}}_{C/d} = \tilde{\mathbf{C}}_{/d}\left(\tilde{\mathbf{C}}'_{/d}\tilde{\mathbf{C}}_{/d}\right)^{-1}\tilde{\mathbf{C}}'_{/d}$ and $\tilde{\mathbf{P}}_{X/d} = \tilde{\mathbf{X}}_{/d}\left(\tilde{\mathbf{X}}'_{/d}\tilde{\mathbf{X}}_{/d}\right)^{-1}\tilde{\mathbf{X}}'_{/d}$ where $\tilde{\mathbf{X}}_{/d} = \left[\mathbf{N} \vdots \tilde{\mathbf{C}}_{/d}\right]$.

Lemma 2. *The projection matrices $\tilde{\mathbf{P}}_{C/d}$ and $\tilde{\mathbf{P}}_{X/d}$, as well as the error terms $\tilde{\boldsymbol{\varepsilon}}_{/d}$ in (13) are all permutation invariant. Hence $\tilde{\mathbf{P}}_{C/d}=\tilde{\mathbf{P}}_{C/f}=\tilde{\mathbf{P}}_C$, $\tilde{\mathbf{P}}_{X/d}=\tilde{\mathbf{P}}_{X/f}=\tilde{\mathbf{P}}_X$ and $\tilde{\boldsymbol{\varepsilon}}_{/d}=\tilde{\boldsymbol{\varepsilon}}_{/f}=\tilde{\boldsymbol{\varepsilon}}$ for all $d, f \in \{1, ..., D\}$.*

12. In Section 3 we use the subscript $\setminus d$ to designate the d^{th} variable or column omitted from the data matrix and its associated coefficients, and in Sections 5-8 we use the subscript $/d$ to designate log deviations from the d^{th} variable.

Proof.

$$\begin{aligned}
\tilde{\mathbf{P}}_{C/d} &= \tilde{\mathbf{C}}_{/d}\left(\tilde{\mathbf{C}}'_{/d}\tilde{\mathbf{C}}_{/d}\right)^{-1}\tilde{\mathbf{C}}'_{/d} \\
&= \tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\left(\left(\tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\right)'\tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\right)^{-1}\left(\tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\right)' \\
&= \tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\mathbf{M}_{d,f}^{-1}\left(\tilde{\mathbf{C}}'_{/f}\tilde{\mathbf{C}}_{/f}\right)^{-1}\left(\mathbf{M}'_{d,f}\right)^{-1}\mathbf{M}'_{d,f}\tilde{\mathbf{C}}'_{/f} \\
&= \tilde{\mathbf{C}}_{/f}\left(\tilde{\mathbf{C}}'_{/f}\tilde{\mathbf{C}}_{/f}\right)^{-1}\tilde{\mathbf{C}}'_{/f} = \tilde{\mathbf{P}}_{C/f} = \tilde{\mathbf{P}}_{C}
\end{aligned}$$

$$\begin{aligned}
\tilde{\mathbf{P}}_{X/d} &= \tilde{\mathbf{X}}_{/d}\left(\tilde{\mathbf{X}}'_{/d}\tilde{\mathbf{X}}_{/d}\right)^{-1}\tilde{\mathbf{X}}'_{/d} \\
&= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right]'(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right\}^{-1}\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right]' \\
&= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\right]'(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\right\}^{-1}\mathbf{M}'_{d,f}\tilde{\mathbf{C}}'_{/f}(\mathbf{I}_n - \mathbf{P}_N)' \\
&= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\mathbf{M}_{d,f}\left(\mathbf{M}_{d,f}\right)^{-1}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right]'(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\right\}^{-1}\left(\mathbf{M}'_{d,f}\right)^{-1}\mathbf{M}'_{d,f}\tilde{\mathbf{C}}'_{/f}\,(\mathbf{I}_n - \mathbf{P}_N)' \\
&= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\right]'(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\right\}^{-1}\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/f}\right]' \\
&= \tilde{\mathbf{X}}_{/f}\left(\tilde{\mathbf{X}}'_{/f}\tilde{\mathbf{X}}_{/f}\right)^{-1}\tilde{\mathbf{X}}'_{/f} = \tilde{\mathbf{P}}_{X/f} = \tilde{\mathbf{P}}_{X}
\end{aligned}$$

Finally the error terms are:

$$\tilde{\boldsymbol{\varepsilon}} = \left(\mathbf{I}_n - \tilde{\mathbf{P}}_X\right)\boldsymbol{y}.$$

□

From Lemma 2 we can generate the log-ratio analogue to Theorems 1–3.

Theorem 5. *The estimated coefficients and variances associated with the non-compositional data* $\mathbf{N}$ *in (13) are permutation invariant:* $\tilde{\boldsymbol{\beta}}_{N/d} = \tilde{\boldsymbol{\beta}}_{N/f} = \tilde{\boldsymbol{\beta}}_N$, $Var\left(\tilde{\boldsymbol{\beta}}_{N/d}\right) = Var\left(\tilde{\boldsymbol{\beta}}_{N/f}\right) = Var\left(\tilde{\boldsymbol{\beta}}_N\right)$ *for all* $d, f \in \{1, ..., D\}$. *Furthermore, the* R^2 *and F-test for the regression (13) are invariant to which compositional variable is omitted from the regression, and the F-test for the joint hypothesis that for any choice* $d \in \{1, ..., D\}$, *the vector* $\tilde{\boldsymbol{\beta}}_{C/d} = 0$ *is also permutation invariant.*

Proof. Follows directly from Lemma 2 and the arguments in the proofs of Theorems 1-3. □

The coefficients in (13) can be written as:

$$\begin{bmatrix} \tilde{\boldsymbol{\beta}}_N \\ \tilde{\boldsymbol{\beta}}_{C/d} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{N}'\left(\mathbf{I}_n - \tilde{\mathbf{P}}_C\right)\mathbf{N}\right)^{-1}\mathbf{N}'\left(\mathbf{I}_n - \tilde{\mathbf{P}}_C\right)\boldsymbol{y} \\ \left(\tilde{\mathbf{C}}'_{/d}\left(\mathbf{I}_n - \mathbf{P}_N\right)\tilde{\mathbf{C}}_{/d}\right)^{-1}\tilde{\mathbf{C}}'_{/d}\left(\mathbf{I}_n - \mathbf{P}_N\right)\boldsymbol{y} \end{bmatrix} \tag{14}$$

and their variances are:

$$\begin{bmatrix} \mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_N\right) \\ \mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_{C/d}\right) \end{bmatrix} = \frac{1}{n-K}\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}\left(\mathbf{N}'\left(\mathbf{I}_n - \tilde{\mathbf{P}}_C\right)\mathbf{N}\right)^{-1} \\ \tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}\left(\tilde{\mathbf{C}}'_{/d}\left(\mathbf{I}_n - \mathbf{P}_N\right)\tilde{\mathbf{C}}_{/d}\right)^{-1} \end{bmatrix} \tag{15}$$

Theorem 6. *The estimated coefficients and variances associated with the compositional data that are common to both* $\tilde{\mathbf{C}}_{/d}$ *and* $\tilde{\mathbf{C}}_{/f}$ *in (13) are identical.*

Proof. Follows the same logic as the proof in Theorem 4 and $\mathbf{M}_{f,d}^{-1} = \mathbf{M}_{d,f}$ since:

$$\tilde{\boldsymbol{\beta}}_{C/d} = \mathbf{M}_{d,f}\tilde{\boldsymbol{\beta}}_{C/f},$$

and:

$$\text{Var}\left(\tilde{\boldsymbol{\beta}}_{C/d}\right) \;=\; \mathbf{M}_{d,f}\text{Var}\left(\tilde{\boldsymbol{\beta}}_{C/f}\right)\mathbf{M}'_{d,f}.$$

The salient factor in $\mathbf{M}_{d,f}$ is the matrix $\mathbf{K}_d$ (or $\mathbf{K}_{d-1}$), analogous to $\mathbf{A}_d$ in Section 3. Unlike $\mathbf{A}_d$, $\mathbf{K}_d$ is an identity matrix for all but row d (or $d-1$). □

So in contrast to Theorem 4 in Section 3, not only are $\tilde{\boldsymbol{\beta}}_N$ and its associated variance, permutation invariant, but so are all the coefficients that correspond to the noncompositional variables, along with the intercept.

The values of the coefficients across the different permutations can be summarised by:

$$\left\{\tilde{\beta}^i_{C/d}, \text{Var}\left(\tilde{\beta}^i_{C/d}\right)\right\} = \begin{cases} \left\{\tilde{\beta}^i_{C/f}, \text{Var}\left(\tilde{\beta}^i_{C/f}\right)\right\}, & i \neq f, i = D+1 \\ \left\{-\sum\limits_{\forall j\neq i, j\neq D+1}\tilde{\beta}^j_{C/f}, \sum\limits_{\forall j\neq i, j\neq D+1}\sum\limits_{\forall k\neq i, k\neq D+1}\text{Cov}\left(\tilde{\beta}^j_{C/f}, \tilde{\beta}^k_{C/f}\right)\right\}, & i = f, i \neq D+1 \end{cases} \tag{16}$$

Furthermore not only are the coefficient values for all the common elements between $\tilde{\boldsymbol{\beta}}_{C/d}$ and $\tilde{\boldsymbol{\beta}}_{C/f}$, $f \neq d$, identical, but the f^{th} element of $\tilde{\boldsymbol{\beta}}_{C/d}$ equals the sum of all the values of $\tilde{\boldsymbol{\beta}}_{C/f}$ multiplied by -1. Therefore a hypothetical sum of the elasticities associated with all D possible elements of $\check{\mathbf{C}}$ is by construction, equal to zero. Furthermore, all the common diagonal terms between $\text{Var}\left(\tilde{\boldsymbol{\beta}}_{C/f}\right)$ and $\text{Var}\left(\tilde{\boldsymbol{\beta}}_{C/d}\right)$ are identical. The missing variance for the coefficient of the f variable, chosen as a basis in $\tilde{\boldsymbol{\beta}}_{C/f}$, is equal to the sum of all the elements of $\text{Var}\left(\tilde{\boldsymbol{\beta}}_{C/d}\right)$. Hence all of the coefficient values and corresponding variances can be found by estimating a restricted version of (8):

$$\min_{\check{\boldsymbol{\beta}}_N, \check{\boldsymbol{\beta}}_C}\left(\boldsymbol{y} - \mathbf{N}\check{\boldsymbol{\beta}}_N - \check{\mathbf{C}}\check{\boldsymbol{\beta}}_C\right)'\left(\boldsymbol{y} - \mathbf{N}\check{\boldsymbol{\beta}}_N - \check{\mathbf{C}}\check{\boldsymbol{\beta}}_C\right) \tag{17}$$

s.t.:

$$r'\begin{bmatrix}\check{\boldsymbol{\beta}}_N \\ \check{\boldsymbol{\beta}}_C\end{bmatrix} = 0$$

where $r = \begin{bmatrix}\mathbf{0}_{K,1} \\ \mathbf{i}_D \\ 0\end{bmatrix}$. It then follows that $\check{\boldsymbol{\beta}}_N = \tilde{\boldsymbol{\beta}}_N$ and $\mathbf{Q}_d\check{\boldsymbol{\beta}}_C = \tilde{\boldsymbol{\beta}}_{C/d}$ for all $d \in \{1, ..., D\}$.

A note of caution. The vectors of coefficients $\boldsymbol{\beta}_{C\setminus d}$, $\tilde{\boldsymbol{\beta}}_{C/d}$ and $\check{\boldsymbol{\beta}}_C$ can be interpreted as the marginal impact of each element in the vector of compositional data, either relative to an omitted category in (3), the log difference with respect to a baseline category in (13), or the log of the elements themselves in (17). As explained in Section 4, using (3) to predict how changing the composition for a particular observation in the sample, or indeed all of them, by a non-infinitesimal amount alters the value of the corresponding dependent variable may be invalid, if elements in the resulting composition fall outside the simplex. Given the nonlinearity of a logarithmic transformation, this problem is potentially even more acute if we attempt to perform a counterfactual analysis using the constrained regression (17), and change all D variables simultaneously, but fail to ensure that all the new coordinates remain in the simplex. By contrast, there are no restrictions on performing counterfactual experiments

involving changes to all $D-1$ log differences in (13). The resultant compositions implied by these changes will always correspond to coordinates within the simplex.

As discussed in Section 3, the variances of the coefficients associated with any permutation d in Table 1 cannot be inferred without either estimating each permutation of (3), or having access to all the elements of the matrix $\text{Var}\left(\boldsymbol{\beta}_{C\setminus d}\right)$. When using additive log-ratios, two permutations suffice to generate all the information we typically need. To see this, we present in Table 2 the different permutations of the Barro growth equation, using log-ratios of the compositional data as in (13). Each column represents a different permutation, where the missing religious category is no longer an excluded variable, but one that is employed as a basis by which the others are divided. Hence, in column (1), where the category not listed is Catholic, the coefficient associated with Protestants is no longer the raw population share, but rather the elasticity that relates to the logarithm of the ratio of Protestants to Catholics, and the subsequent coefficient associated with Orthodox is the elasticity that relates to the logarithm of the ratio of Orthodox Christians to Catholics.

As in Table 1, the estimates of (13) in Table 2, in accordance with Theorem 5, demonstrate the permutation invariance of the coefficients and variances associated with the noncompositional variables that relate to the Solow growth model. Unlike Table 1, in Table 2, as Theorem 6 indicates, the coefficients for the compositional data—i.e, the shares of religious adherence in each country—are the same as well, regardless of which religious category (the missing term on the diagonal) is chosen as the baseline against which the other categories are log differenced. That invariance applies to the intercept terms as well. Given that the intercept relates to underlying parameters associated with the Solow growth model in (1), this is another useful attribute of this methodology. Only the measure of multicollinearity, mean VIF, differs across the different columns (more on this in Section 8). By estimating the model using log-ratios, we eliminate any ambiguity about the association between the individual shares and the dependent variable.

The results in Table 2 reveal a much less ambiguous pattern than in Table 1—the shares of the population that adhere to two religions, Orthodox Christianity and Buddhism, are associated at a statistically significant 1% level with higher rates of growth in the years between 2001 to 2020. The category of Other Religions is associated with lower rates of growth, but only at the 10% significance level. The coefficients associated with the compositional covariates are not statistically significant. There are hints of this pattern in Table 1—in the permutations in columns (1), (2) and (5) the coefficient for the share of Buddhists in the population is statistically significant at the 1% level, as of course are the corresponding categories of Catholics, Protestants and Muslims in column (7), with the Buddhist category omitted. For Orthodox Christianity, the pattern in Table 1 is weaker—only significant relative to Protestants (at the 5% level) and relative to Buddhists (at the 10% level).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP (2001)	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Pop. Growth	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***
	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)
Savings	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Catholic		−0.003	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Protestant	−0.002		−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002
	(0.010)		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Orthodox	0.020***	0.020***		0.020***	0.020***	0.020***	0.020***	0.020***	0.020***	0.020***
	(0.006)	(0.006)		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Other Christ.	0.001	0.001	0.001		0.001	0.001	0.001	0.001	0.001	0.001
	(0.006)	(0.006)	(0.006)		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Muslim	−0.009	−0.009	−0.009	−0.009		−0.009	−0.009	−0.009	−0.009	−0.009
	(0.008)	(0.008)	(0.008)	(0.008)		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Jewish	0.002	0.002	0.002	0.002	0.002		0.002	0.002	0.002	0.002
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)		(0.010)	(0.010)	(0.010)	(0.010)
Buddhist	0.027***	0.027***	0.027***	0.027***	0.027***	0.027***		0.027***	0.027***	0.027***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)		(0.007)	(0.007)	(0.007)
Other Eastern	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002		−0.002	−0.002
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)		(0.011)	(0.011)
Hindu	−0.007	−0.007	−0.007	−0.007	−0.007	−0.007	−0.007	−0.007		−0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)		(0.007)
Other	−0.027*	−0.027*	−0.027*	−0.027*	−0.027*	−0.027*	−0.027*	−0.027*	−0.027*	
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	
Constant	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***
	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)
Observations	164	164	164	164	164	164	164	164	164	164
R^2	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526
Adjusted R^2	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488
Mean VIF	2.73	2.65	4.45	3.84	3.44	2.53	3.30	1.98	3.39	1.61

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Additive log ratio regressions (8). Dependent variable is log difference of GDP (PPP) between 2001 to 2020.

Working in log differences of shares rather than changes in shares might seem cumbersome. However, the benefits of permutation invariance can more than compensate for any awkwardness. Take as an example Singapore, one of the more religiously diverse countries in our sample, where the seven main religious categories are: Buddhism, 0.33; Other, 0.186; Islam, 0.142; Eastern Religions, 0.108; Catholicism, 0.07; Protestantism, 0.065; Hinduism, 0.05; and Other Christian, 0.046. The log differences between the largest group, adherents of Buddhism and the number of people in the other categories are as follows: Other Religions, 0.573; Islam, 0.843; Eastern Religions, 1.117; Catholicism 1.551; Protestantism, 1.625; Hinduism, 1.887; and Other Christian, 1.97. The log difference in per-capita GDP between 2001 and 2020 implies Singapore's population enjoyed an annualised economic growth rate that averaged 4.05%. The coefficient corresponding to the Buddhism category of 0.027 means that in a hypothetical country, identical to Singapore in every other way, but where the number of Buddhists relative to any of the other categories is greater by the same log difference of 0.1, we would expect to observe an annualised growth rate during that period of 4.32 instead. By contrast, working in shares as in Section 3 means that a similar question regarding a change in the number of Buddhists in the population must be addressed relative to each religious category independently, according to each specific coefficient whose statistical significance varies widely.

It is impossible to assert a priori whether log-ratios yield regressions that better fit the data. There is indeed a marginal improvement in our baseline example—in Table 2, the value of R^2 is 0.526, which is marginally higher than the value of 0.509 in Table 1. The F-test on the significance of the religious covariates is $F(9, 151)$=3.13 (p-value 0.002) for the former and $F(9, 151)$=2.43 (p-value 0.013) in the latter. However, suppose we return to the World Religion Project (Maoz and Henderson (2013)) and slightly change the way we aggregate the different sects into ten different religious categories by amalgamating Orthodox Christians and Other Christians into one category, but separating Muslims into two categories, Sunni Muslims and non Sunni Muslims. The value of R^2 is 0.500 in Table 9 (Appendix B) where additive log-ratios are used, but 0.509 in Table 8 (Appendix B), where the regression uses the compositional data and omits one of the variables. Similarly, the F-test on the significance of the religious covariates is $F(9, 151)$=2.09 (p-value 0.034) in the former and $F(9, 151)$=2.44 (p-value 0.013) in the latter.

At the same time, in the Shapley-Owen decomposition in Table 3, the Owen value for the religious category is only 0.122 when the model is estimated using additive log-ratios for our baseline religious categories, but 0.168 when raw shares are used. Instead, the inclusion of log-ratios for the religious categories raises the Owen values for the noncompositional variables associated with the Solow model, particularly for the level of GDP growth in 2001. This pattern is almost identical when we use our alternative aggregation of religious categories.

To illustrate this for the baseline aggregates, Figure 4a plots the orthogonal components of growth and the log level of GDP against each other. The slope of the regression line between the two corresponds to the coefficient for the log level of GDP, associated with both the test for convergence and estimates of its speed, in Tables 1 and 2 (in Appendix B, Figures 9a and

	Original Composition				Alternative Composition			
	Excluded Variable		Additive Log-Ratio		Excluded Variable		Additive Log-Ratio	
Variables	Owen Values	Percent	Owen Values	Percent	Owen Values	Percent	Owen Values	Percent
GDP (2001)	0.168	33.10%	0.214	40.75%	0.160	31.52%	0.197	39.36%
Pop. Growth	0.161	31.61%	0.175	33.31%	0.169	33.29%	0.172	34.36%
Savings	0.012	2.27%	0.015	2.83%	0.011	2.20%	0.012	2.37%
Religion	0.168	33.02%	0.122	23.12%	0.168	33.00%	0.119	23.92%
Total	0.509	100%	0.526	100%	0.509	100%	0.500	100%

Table 3: Shapley-Owen decomposition.

10a correspond to the other two variables, population growth and savings). Using additive log-ratios for the religious categories generates a slightly steeper slope and a narrower confidence interval. Figures 4b and 4c show how the orthogonal components shift and the kernel density estimates of their distributions change when we switch from the estimates using the raw data to additive log-ratios. The differences are subtle, but there appear to be fewer outliers when log-ratios are employed, which is consistent with the results in Table 3.

Can we detect any systematic differences between the squared error terms for the $h \in \{1, ..., N\}$ observations $\tilde{\varepsilon}_h^2$ and ε_h^2, where $\varepsilon_h \in \boldsymbol{\varepsilon}$, from (3) and where $\tilde{\varepsilon}_h \in \tilde{\boldsymbol{\varepsilon}}$ from (13)? Using the baseline composition, it is hard to see differences that pertain to particular regions in the map in Figure 5. In Table 4, we regress the difference between the two squared error terms on three indices that pertain specifically to the distribution of religious categores: (1) the maximum share of a religion—in effect how close a particular observation is to an edge of the simplex; (2) similarly, a Herfindahl index of religious concentration; and (3) the Aitchison distance of a country's religious composition to the geometric mean of religious compositions across all the different countries:

$$\tilde{\varepsilon}_h^2 - \varepsilon_h^2 = \phi_0 + \phi_1 \text{Maximum Share}_h + \phi_2 \text{Hefindahl}_h + \phi_3 \text{Aitchison}_h + u_h. \tag{18}$$

There is no significant relationship between the three different indices and the differences in the squared error terms for either composition. Whether goodness of fit improve, or deteriorates does not appear, in our example, to be influenced by the distribution of the coordinates in the simplex.

6 From Additive Log-Ratios to Isometric Log-Ratios

The two key benefits to estimating regressions, that include compositional covariates using ALRs are permutation invariance and the fact that the estimated coefficients are associated with coordinates in a vector space. There is another benefit: namely when we compare angles and

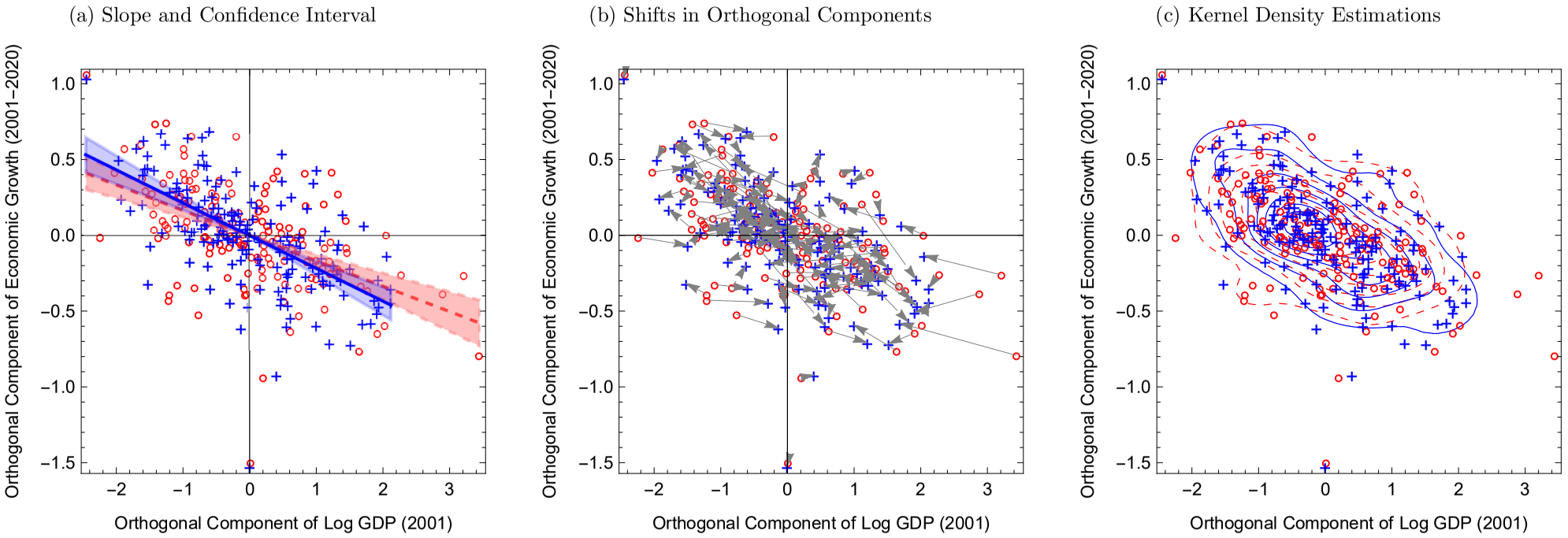


Figure 4: Components of per-capita economic growth between 2001 and 2020 orthogonal to average savings rates and population growth rates between 2001 and 2020 as well as religious composition, against the component of log per-capita GDP in 2001 orthogonal to the same variables. The red circles ○ represent the regression with an excluded religious category, and the blue crosses + the regressions with the religious composition included as additive log-ratios.

	Original Composition				Alternative Composition			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Maximum Share	−0.019 (0.020)			0.128 (0.101)	−0.013 (0.020)			−0.073 (0.097)
Herfindahl Index		−0.023 (0.019)		−0.141 (0.093)		−0.010 (0.019)		0.059 (0.094)
Aitchison Distance			−0.000 (0.002)	0.000 (0.002)			−0.000 (0.002)	−0.000 (0.002)
Constant	0.010 (0.014)	0.010 (0.011)	−0.002 (0.018)	−0.014 (0.024)	0.010 (0.014)	0.007 (0.011)	0.006 (0.017)	0.019 (0.023)
Observations	164	164	164	164	164	164	164	164
R^2	0.005	0.009	0.000	0.020	0.003	0.002	0.000	0.005
Adjusted R^2	−0.001	0.003	−0.006	0.002	−0.003	−0.004	−0.006	−0.013

Table 4: Estimation results for (18).

distances using Aitchison geometry, the ALR transforms the rectangles in Figure 2 to the parallelograms in Figure 6—the coordinates with respect to the basis $[\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{d-1}, \mathbf{w}_{d+1}, ..., \mathbf{w}_D]$ in (5) are not as distorted as they are when the composition, with one component omitted, is treated as if they were coordinates in Euclidean space, as in Figure 3. Yet these basis vectors are neither normal or orthogonal. The Aitchison norm for any of the $i = 1, ..., D-1$ is $\|\mathbf{w}_i\|_a = \sqrt{\frac{D-1}{D}}$, and the Aitchison inner product between any two i, j vectors is $\langle \mathbf{w}_i, \mathbf{w}_j \rangle_a = -1/D$. Hence the angle between the vectors (in radians) is $\arccos(\frac{1}{1-D})$. Since the ALR transformation from the simplex $\mathbb{S}^D$ to Euclidean space $\mathbb{R}^{D-1}$ generates coordinates with respect to an oblique basis, some distortions of the distances and angles between the coordinates in the regression remain—distortions that are inversely related to the value of D. In Figure 2 where D=3, the angles between the basis vectors are equal to $\frac{2\pi}{3}$ or 120°.[13]

In some applications a better solution is available. One can transform the data to Euclidean space $\mathbb{R}^D$ as log deviations from a geometric mean or centred log-ratios (CLR) by postmultiplying $\check{\mathbf{X}}_C \mathbf{Q}_{D+1}$ by the matrix $\frac{1}{D}\mathbf{I}_D - \mathbf{H}_D$ (where $\mathbf{H}_D$ is a $D \times D$ unit matrix of ones), which generates for each point in the simplex:

$$\left(\ln \frac{c_1}{\sqrt[D]{\prod_{i=1}^{D} c_i}}, \ln \frac{c_2}{\sqrt[D]{\prod_{i=1}^{D} c_i}}, ..., \ln \frac{c_D}{\sqrt[D]{\prod_{i=1}^{D} c_i}} \right), c_1, ..., c_D \in \mathbf{c}. \tag{19}$$

However as each row represented by (19) sums to zero, this transformation does not resolve the

13. In Figure 6 the angles associated with the vertices in the polygon in green are Chad: 115.64° (0.642π), Nigeria: 64.41° (0.358π), Zambia: 116.36° (0.646π), Eswatini: 63.58° (0.353π). The angles associated with the vertices in the polygon in red are Portugal: 64.57° (0.359π), Ireland: 115.40° (0.641π), Bosnia: 62.56° (0.348π), Lebanon: 117.48° (0.653π). The angles associated with the vertices in the polygon in blue are Togo: 119.20° (0.662π), Korea: 59.63° (0.331π), Lithuania: 120.28° (0.668π), Rwanda: 60.89° (0.338π).

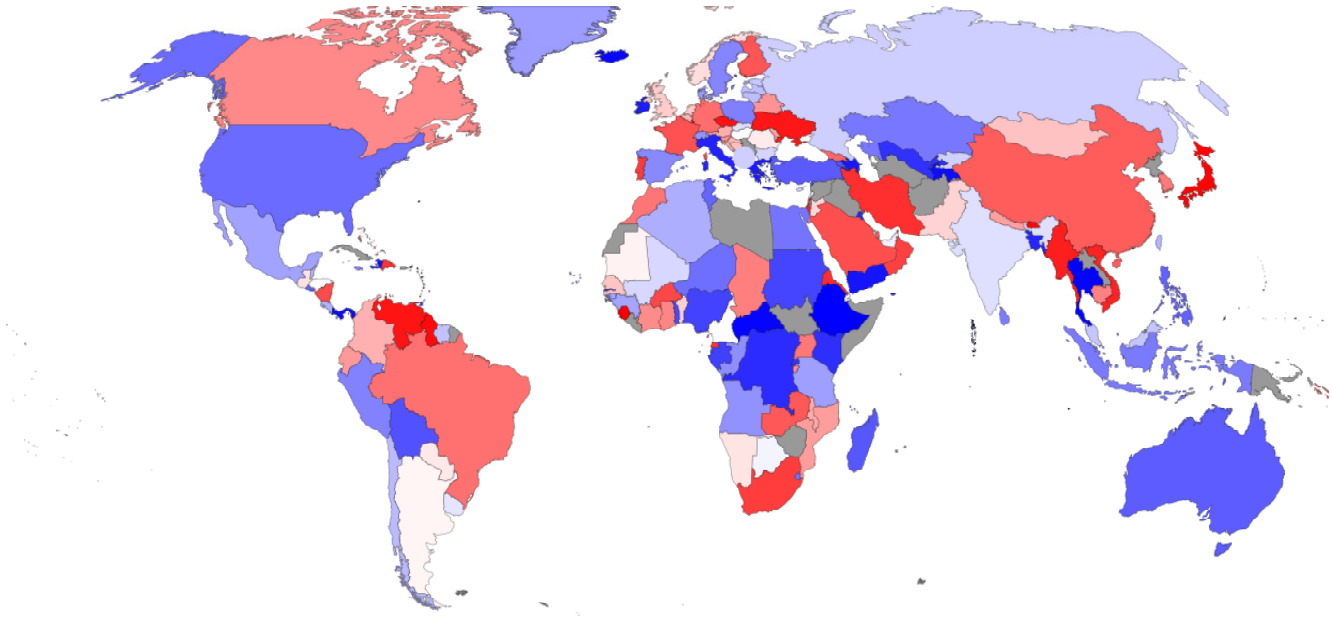


Figure 5: The deeper a country is shaded red, the more the squared errors ε^2 from the regression with an omitted category (3) exceeds the squared errors $\tilde{\varepsilon}^2$ from the additive log-ratio (13). Conversely the deeper a country is shaded blue, the more the squared errors $\tilde{\varepsilon}^2$ from the additive log-ratio (13) exceeds the squared errors ε^2 from the regression with an omitted category (3). Countries shaded grey were omitted for lack of data during the sample period.

problem of linear dependence and is therefore unsuitable for linear regression.

What is needed for estimating regressions is a transformation that generates coordinates with respect to an orthonormal basis that are linearly independent. Egozcue et al. (2003) suggest taking the Aitchison inner product between the points on the simplex and any orthonormal basis $\mathbf{e}$ on the simplex, which transforms the data into a length $D-1$ vector of isometric log-ratios (ILR). This is the equivalent to postmultiplying $\check{\mathbf{C}}\mathbf{Q}_{D+1}$ by the transpose of a $(D-1)\times D$ contrast matrix $\mathbf{U}$, where the $D-1$ rows are the centered log-ratios (19) of the chosen orthonormal basis $\mathbf{e}$. Transposing $\mathbf{U}$, the $D-1$ columns in $\mathbf{U}'$ form a basis in a $D-1$ subspace where each of the D elements, the "balances," sum to zero. Egozcue et al. (2003) show that post multiplying an ALR composition by $\mathbf{F}^{+}$ (the Moore Penrose generalised inverse of $\mathbf{F}$ from (5)) generates a CLR, and postmultiplying that by $\mathbf{U}'$ generates an ILR composition.[14]

Extending this to accommodate the vector of ones in $\tilde{\mathbf{C}}$ that generate the intercept term, we can convert additive log-ratios to a generic set of isometric log-ratios by postmultiplying $\tilde{\mathbf{C}}_{/d}$ by the $D\times D$ matrix $\bar{\mathbf{F}}_d\bar{\mathbf{U}}'$ where:

$$\bar{\mathbf{F}}_d = \begin{bmatrix} & \mathbf{F}_d^{+} & \\ \hdashline & 1/D \quad 1/D \quad \ldots 1/D & \end{bmatrix}, \tag{20}$$

14. See Proposition 4 in Egozcue et al. (2003).

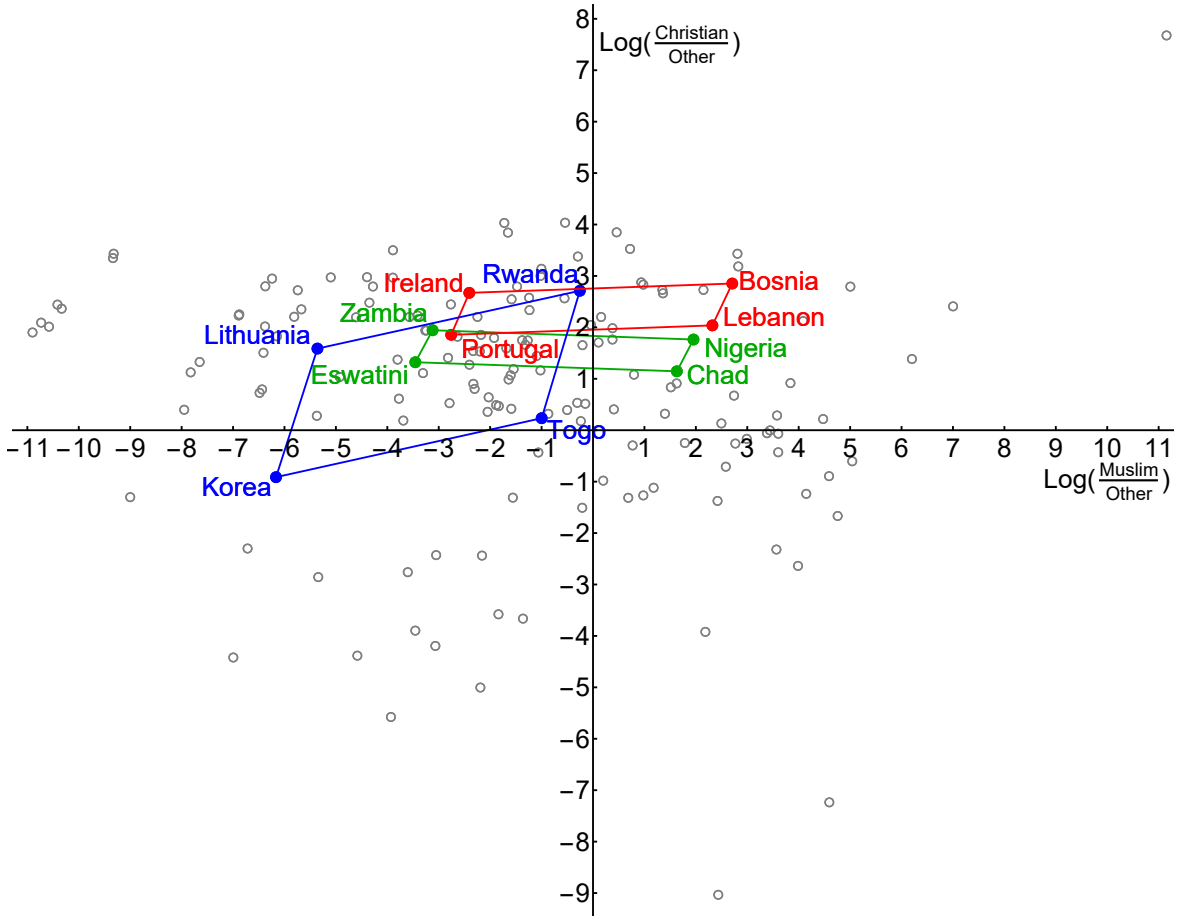


Figure 6: ALR transformation for the three-part composition in Figure 2.

and $\bar{\mathbf{U}}$ is a generic contrast matrix $\mathbf{U}$, appended with a row vector of ones, $\mathbf{i}_D$. Note that each coordinate in $\tilde{\mathbf{C}}_{/d}\bar{\mathbf{F}}_d$ is the sum of the CLR and the constant term $1/D$.

What orthonormal basis is appropriate for multiple regression? Adopting a version of the sequential binary partition method in Pawlowsky-Glahn et al. (2015):

$$\mathbf{U}_1 = \begin{bmatrix} -\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} & 0 & \dots & \dots & \dots & 0 \\ -\frac{1}{2}\sqrt{\frac{2}{3}} & -\frac{1}{2}\sqrt{\frac{2}{3}} & \sqrt{\frac{2}{3}} & 0 & \dots & \dots & 0 \\ -\frac{1}{3}\sqrt{\frac{3}{4}} & -\frac{1}{3}\sqrt{\frac{3}{4}} & -\frac{1}{3}\sqrt{\frac{3}{4}} & \sqrt{\frac{3}{4}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \ddots & 0 \\ -\frac{1}{D-1}\sqrt{\frac{D-1}{D}} & -\frac{1}{D-1}\sqrt{\frac{D-1}{D}} & -\frac{1}{D-1}\sqrt{\frac{D-1}{D}} & -\frac{1}{D-1}\sqrt{\frac{D-1}{D}} & \dots & \dots & \sqrt{\frac{D-1}{D}} \end{bmatrix} \tag{21}$$

which generates a set of balances with these coordinates:

$$\tilde{c}^l_{i/1} = \sqrt{\frac{i-1}{i}} \ln\left[c^l_i \left(\prod_{j=1}^{i-1} c^l_j \right)^{\frac{1}{i-1}} \right], i = 2, ..., D;\ l = 1, ..., n, \tag{22}$$

also known as pivot log-ratios.[15] For each observation l, the first term in (22), represented by the vector $\tilde{c}^l_{2/1}$ (where the term /1 in the subscript refers to the orthonormal basis $\mathbf{U}_1$), represents

15. See Greenacre (2018) for an alternative matrix transformation that generates pivot log-ratios.

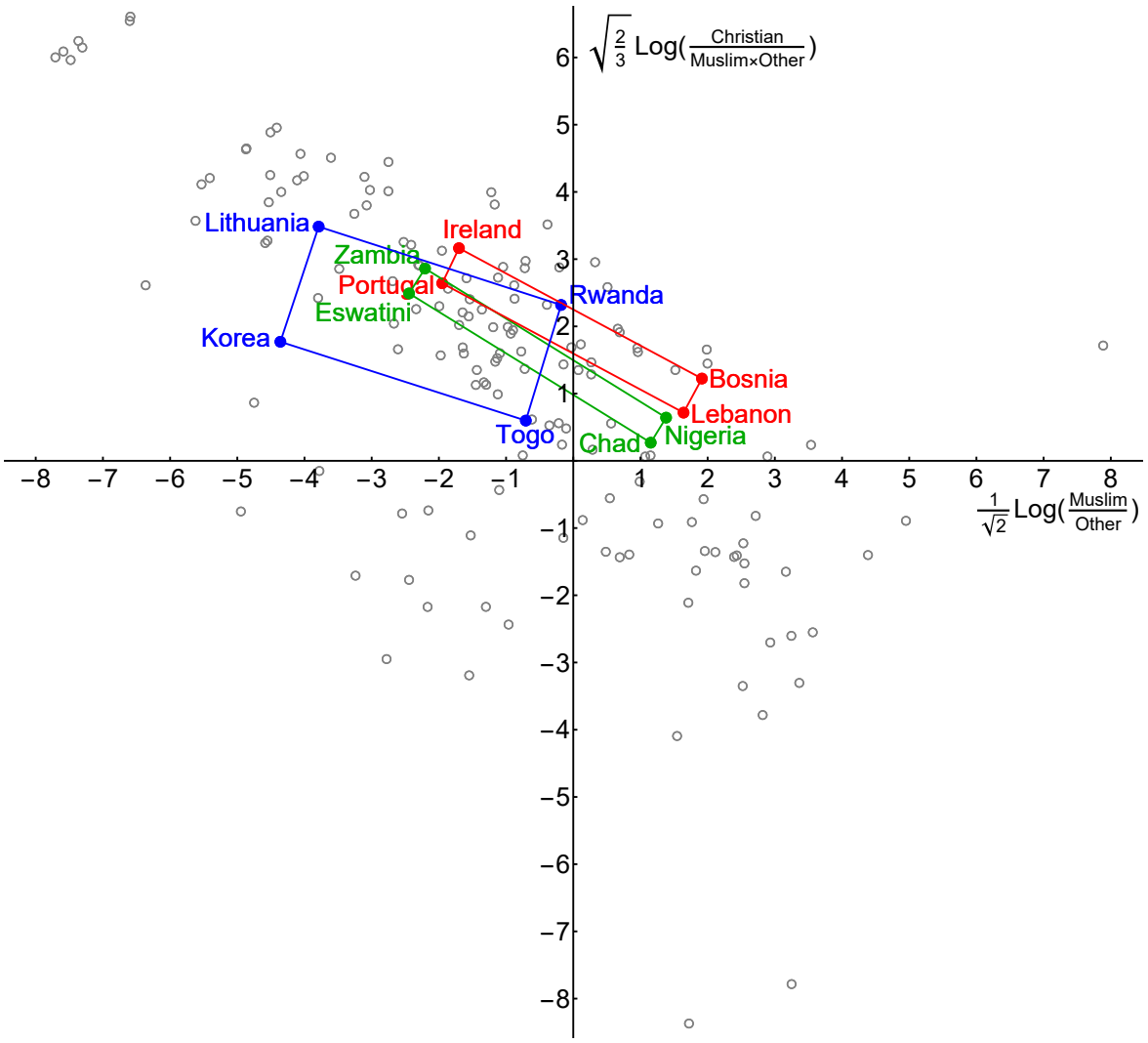


Figure 7: ILR transformation for the three-part composition in Figure 2.

the log difference between c_2^l and c_1^l, the second term, $\bar{c}_{1,3}^l$, the log difference between c_3^l and the geometric mean of c_2^l and c_1^l, and each subsequent term the log deviation of c_i^l from the geometric mean of c_1^l through c_{i-1}^l. The last remaining term is potentially the most interesting and useful: $\bar{c}_{1,D}^l$, which represents the log deviation of c_D^l from the geometric mean of all the preceding $D-1$ variables. Contrasting the three sets of points that form rectangles in Figure 6 with Figure 7 illustrates how the shift from ALR to ILR means the transformed compositional data are not only coordinates in a vector space, but are now coordinates with respect to an orthogonal basis.

The regression we estimate has the compositional data included as isometric log-ratios:

$$\boldsymbol{y} = \mathbf{N}\bar{\boldsymbol{\beta}}_{N/1} + \bar{\mathbf{C}}_{/1}\bar{\boldsymbol{\beta}}_{C/1} + \bar{\boldsymbol{\varepsilon}}_{/1}, \tag{23}$$

where $\bar{\mathbf{C}}_{/1} = \bar{\mathbf{F}}_d\bar{\mathbf{U}}_1'$. This can now be generalized to generate additional permutations.

Define the $D \times D$ exchange matrix $\mathbf{J}_D$, where each element is defined as:

$$j_{i,k} = \begin{cases} 1, & k = D - i + 1 \\ 0, & k \neq D - i + 1 \end{cases},$$

which is the mirror image of the identity matrix $\mathbf{I}_D$. We can now extend (21) to incorporate D different permutations of the contrast matrix $\mathbf{U}_1$, where $\mathbf{U}_h$ refers to (21) with column h exchanged with column 1 and column D exchanged with column $D-h+1$. If D is an even number:

$$\mathbf{U}_h = \begin{cases} \mathbf{U}_1\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_h\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_h, & 1 \le h \le D/2 \\ \mathbf{U}_1\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_{h-1}\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_h, & D/2+1 \le h \le D; \end{cases} \tag{24}$$

and if D is odd:

$$\mathbf{U}_h = \begin{cases} \mathbf{U}_1\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_h\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_h, & 1 \le h < (D+1)/2 \\ \mathbf{U}_1\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_h\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_{h-1}, & h = (D+1)/2 \\ \mathbf{U}_1\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_{h-1}\mathbf{J}_D\mathbf{S}_1...\mathbf{S}_h, & (D+1)/2 < h \le D. \end{cases} \tag{25}$$

As before, we define the augmented contrast matrix $\bar{\mathbf{U}}_h$ as the matrix $\mathbf{U}_h$ with the addition of a last row vector of ones.

Define the matrix $\mathbf{G}_{d,h} \equiv \bar{\mathbf{F}}_d\bar{\mathbf{U}}_h'$, $h = 1,...,D$. The symmetry of the CLR means that $\tilde{\mathbf{C}}_{/d}\bar{\mathbf{F}}_d = \tilde{\mathbf{C}}_{/f}\bar{\mathbf{F}}_f$ for all $d, f \in \{1,...,D\}$. Hence, we can generate a matrix of isometric log-4ratios $\bar{\mathbf{C}}_{/h}$, associated with an augmented contrast matrix $\bar{\mathbf{U}}_h$, from any set of additive log-ratios so that $\bar{\mathbf{C}}_{/h} = \tilde{\mathbf{C}}_{/d}\mathbf{G}_{d,h} = \tilde{\mathbf{C}}_{/f}\mathbf{G}_{f,h}$.[16] We can estimate $h = 1,...,D$ different versions of the regression:

$$\boldsymbol{y} = \mathbf{N}\bar{\boldsymbol{\beta}}_{N/h} + \bar{\mathbf{C}}_{/h}\bar{\boldsymbol{\beta}}_{C/h} + \bar{\boldsymbol{\varepsilon}}_{/h}. \tag{26}$$

Defining the matrix $\bar{\mathbf{X}}_{/h} = \left[\mathbf{N} \vdots \bar{\mathbf{C}}_{/h}\right]$, consider the properties of the projection matrices associated with (26) for the different permutations $h \in \{1,...,D\}$.

Lemma 3. *The projection matrices associated with the ILR coordinates are permuation invariant and equal to the projection matrices associated with the ALR coordinates so that* $\bar{\mathbf{P}}_{C/h} = \tilde{\mathbf{P}}_C$ *and* $\bar{\mathbf{P}}_{X/h} = \tilde{\mathbf{P}}_X$ *for all* $\in \{1,...,D\}$.

Proof.

$$\begin{aligned} \bar{\mathbf{P}}_{C/h} &= \bar{\mathbf{C}}_{/h}\left(\bar{\mathbf{C}}_{/h}'\bar{\mathbf{C}}_{/h}\right)^{-1}\bar{\mathbf{C}}_{/h}' \\ &= \tilde{\mathbf{C}}_{/d}\mathbf{G}_{d,h}\left(\mathbf{G}_{d,h}'\tilde{\mathbf{C}}_{/d}'\tilde{\mathbf{C}}_{/d}\mathbf{G}_{d,h}\right)^{-1}\mathbf{G}_{d,h}'\tilde{\mathbf{C}}_{/d}' \\ &= \tilde{\mathbf{C}}_{/d}\mathbf{G}_d\mathbf{G}_{d,h}^{-1}\left(\tilde{\mathbf{C}}_{/d}'\tilde{\mathbf{C}}_{/d}\right)^{-1}\left(\mathbf{G}_{d,h}'\right)^{-1}\mathbf{G}_{d,h}'\tilde{\mathbf{C}}_{/d}' \\ &= \tilde{\mathbf{C}}_{/d}\left(\tilde{\mathbf{C}}_{/d}'\tilde{\mathbf{C}}_{/d}\right)^{-1}\tilde{\mathbf{C}}_{/d}' = \tilde{\mathbf{P}}_C \end{aligned}$$

$$\begin{aligned} \bar{\mathbf{P}}_{X/h} &= \bar{\mathbf{X}}_{/h}\left(\bar{\mathbf{X}}_{/h}'\bar{\mathbf{X}}_{/h}\right)^{-1}\bar{\mathbf{X}}_{/h}' \\ &= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\bar{\mathbf{C}}_{/h}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\bar{\mathbf{C}}_{/h}\right]'(\mathbf{I}_n - \mathbf{P}_N)\,\bar{\mathbf{C}}_{/h}\right\}^{-1}\left[(\mathbf{I}_n - \mathbf{P}_N)\,\bar{\mathbf{C}}_{/h}\right]' \\ &= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\mathbf{G}_{d,h}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\mathbf{G}_{d,h}\right]'(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\mathbf{G}_{d,h}\right\}^{-1}\mathbf{G}_{d,h}'\tilde{\mathbf{C}}_{/d}'(\mathbf{I}_n - \mathbf{P}_N)' \\ &= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\mathbf{G}_{d,h}\,(\mathbf{G}_{d,h})^{-1}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right]'(\mathbf{I}_n - \tilde{\mathbf{P}}_1)\,\tilde{\mathbf{C}}_{/d}\right\}^{-1}\left(\mathbf{G}_{d,h}'\right)^{-1}\mathbf{G}_{d,h}'\tilde{\mathbf{C}}_{/d}'(\mathbf{I}_n - \mathbf{P}_N)' \\ &= \mathbf{P}_N + (\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\left\{\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right]'(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right\}^{-1}\left[(\mathbf{I}_n - \mathbf{P}_N)\,\tilde{\mathbf{C}}_{/d}\right]' \\ &= \tilde{\mathbf{X}}_{/d}\left(\tilde{\mathbf{X}}_{/d}'\tilde{\mathbf{X}}_{/d}\right)^{-1}\tilde{\mathbf{X}}_{/d}' = \tilde{\mathbf{P}}_{X/d} = \tilde{\mathbf{P}}_X \end{aligned}$$

16. Note that each row of $\tilde{\mathbf{C}}_{/d}$ represents the $D-1$ log deviations of the $d \notin \{1,...,D\}$ shares with respect to d^{th} share and $\bar{\mathbf{C}}_{/h}$ the log deviations of the $h \notin \{1,...,D\}$ shares with respect to the cascading geometric means that all have one element, the h^{th} share, in common. If $h = d$, the first column in $\bar{\mathbf{C}}_{/d}$ always equals the first column in $\tilde{\mathbf{C}}_{/d}$, divided by $\sqrt{2}$.

□

From Lemma 3 we can derive the values of the estimated coefficients and their associated variances:

$$\begin{bmatrix} \bar{\boldsymbol{\beta}}_N \\ \bar{\boldsymbol{\beta}}_{C/h} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{N}'\left(\mathbf{I}_n - \tilde{\mathbf{P}}_C\right)\mathbf{N}\right)^{-1}\mathbf{N}'\left(\mathbf{I}_n - \tilde{\mathbf{P}}_C\right)\boldsymbol{y} \\ \mathbf{G}_{d,h}^{-1}\left(\tilde{\mathbf{C}}'_{/d}\left(\mathbf{I}_n - \mathbf{P}_N\right)\tilde{\mathbf{C}}_{/d}\right)^{-1}\tilde{\mathbf{C}}'_{/d}\left(\mathbf{I}_n - \mathbf{P}_N\right)\boldsymbol{y} \end{bmatrix} \tag{27}$$

$$\begin{aligned} \mathrm{Var}\left(\bar{\boldsymbol{\beta}}_{C/h}\right) &= \frac{1}{n-K}\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}\left(\bar{\mathbf{C}}'_{\diagdown d}\left(\mathbf{I}_n - \mathbf{P}_N\right)\bar{\mathbf{C}}_{\diagdown d}\right)^{-1} \\ &= \frac{1}{n-K}\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}\left(\mathbf{G}'_{d,h}\bar{\mathbf{C}}'_{\diagdown f}\left(\mathbf{I}_n - \mathbf{P}_N\right)\bar{\mathbf{C}}_{\diagdown f}\mathbf{G}_{d,h}\right)^{-1} \\ &= \mathbf{G}_{d,h}^{-1}\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_{C/d}\right)\left(\mathbf{G}'_{d,h}\right)^{-1} \end{aligned} \tag{28}$$

Theorem 7. *The estimated coefficients, variances, and residuals associated with the non-compositional data* $\mathbf{N}$ *in the regression using ILR coordinates (23), are all permutation invariant and identical to those in the regression using ALR coordinates (13):* $\bar{\boldsymbol{\beta}}_N = \tilde{\boldsymbol{\beta}}_N$, $Var\left(\bar{\boldsymbol{\beta}}_N\right) = Var\left(\tilde{\boldsymbol{\beta}}_N\right)$ *and* $\bar{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{\varepsilon}}$. *Hence, the* R^2 *and* F*-test for the regression (26) and the* F*-test for the joint hypothesis for the compositional variables* $\bar{\boldsymbol{\beta}}_N = 0$ *are also equal to those in (13).*

Proof. Follows directly from Lemma 3, (27) and (28). □

There are several options for using ILR coordinates in a regression. One option is to simply choose one permutation of (26). However, interpreting the regression is cumbersome, and, aside from the coefficient associated with the log difference between one variable and all the others, not very intuitive. Given that one implication of Theorem 7 is that the coefficients associated with the noncompositional data $\bar{\boldsymbol{\varepsilon}}$ and the error terms $\bar{\boldsymbol{\varepsilon}}$ do not change in the transition from ALR to ILR coordinates, the only value in using the latter is when the compositional data are included in the regression not merely as controls or instruments.

One way to generate a model which can provide coefficients on the compositional data that are more intuitive is to follow Hron et al. (2012) and estimate D different permutations of (26), where in each regression the log deviation of a different variable against all the remaining variables in the dataset is changed. We can then build a statistical model using only those last coefficients.

As I demonstrate below, there is a simpler way to incorporate ILR transformations into a regression and generate the estimates of these D coefficients and associated variances, but by implementing the method in Hron et al. (2012), we produce the results in Table 5 that create a pattern most easily compared with Tables 1 and 2. In accordance with Theorem 7, the estimates of the coefficients that correspond to the noncompositional data in Table 5 (the log level of GDP in 2001, the savings rate and population growth), along with the constant in the last row, are not only invariant across the different columns but identical to the estimates using the ALR transformation in Table 2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP (2001)	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***	−0.217***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Pop. Growth	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***	−0.971***
	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)
Savings	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***	0.126***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Catholic		−0.001	−0.016**	−0.003	0.004	−0.003	−0.021***	−0.001	0.003	**−0.003**
		(0.012)	(0.007)	(0.008)	(0.008)	(0.010)	(0.008)	(0.011)	(0.008)	**(0.010)**
Protestant	0.001		−0.008	−0.000	0.004	−0.001	−0.011	0.001	**−0.002**	0.018
	(0.012)		(0.011)	(0.011)	(0.010)	(0.011)	(0.011)	(0.012)	**(0.010)**	(0.015)
Orthodox	0.018***	0.018***		0.018***	0.021***	0.018**	0.011*	**0.021*** **	0.020***	0.028***
	(0.006)	(0.006)		(0.006)	(0.007)	(0.007)	(0.006)	**(0.006)**	(0.006)	(0.009)
Other Christ.	−0.004	−0.004	−0.004		−0.001	−0.003	**0.001**	0.003	−0.002	0.003
	(0.007)	(0.007)	(0.007)		(0.007)	(0.006)	**(0.006)**	(0.007)	(0.007)	(0.007)
Muslim	−0.012	−0.012	−0.012	−0.012		**−0.009**	−0.017**	−0.007	−0.010	−0.006
	(0.008)	(0.008)	(0.008)	(0.008)		**(0.009)**	(0.008)	(0.008)	(0.009)	(0.009)
Jewish	0.000	0.000	0.000	0.000	**0.002**		−0.005	0.004	0.001	0.005
	(0.010)	(0.010)	(0.010)	(0.010)	**(0.010)**		(0.010)	(0.010)	(0.010)	(0.011)
Buddhist	0.024***	0.024***	0.024***	**0.029*** **	0.023***	0.021***		0.027***	0.024***	0.027***
	(0.007)	(0.007)	(0.007)	**(0.008)**	(0.007)	(0.007)		(0.008)	(0.007)	(0.007)
Other Eastern	−0.007	−0.007	**−0.002**	−0.003	−0.007	−0.009	−0.007		−0.006	−0.003
	(0.012)	(0.012)	**(0.012)**	(0.011)	(0.012)	(0.011)	(0.012)		(0.011)	(0.012)
Hindu	−0.010	**−0.007**	−0.011	−0.007	−0.011	−0.012*	−0.011	−0.008		−0.008
	(0.007)	**(0.007)**	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)		(0.007)
Other	**−0.029***	−0.030*	−0.029*	−0.026	−0.029*	0.030*	−0.029*	−0.027	−0.029*	
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	
Constant	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***	2.668***
	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)	(0.272)
Observations	164	164	164	164	164	164	164	164	164	164
R^2	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526
Adjusted R^2	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488
Mean VIF	1.47	1.46	1.59	1.55	1.62	1.56	1.48	1.58	1.62	1.74

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Isometric log ratio regressions. Values in bold are coefficients for the log deviation from the geometric mean of all nine other variables. Dependent variable is log difference of per-capita GDP (PPP) between 2001 to 2020.

In column (1) the coefficient on the second category, Protestants, represents the elasticity with respect to the log deviation between the share of Protestants and the first category, the share of Catholics. The coefficient on the third category, the Orthodox Christians, is the elasticity of the log deviation between the share of Orthodox Christians and the geometric mean of Protestants and Catholics. The coefficient on the fourth, Other Christians is the elasticity of the log deviation between that category and the geometric mean of Protestants, Catholics and Orthodox Christians. This pattern continues to the last religious category: Other Religions. That coefficient, highlighted in bold, is the elasticity with respect to the remaining categories and the one we retain for our statistical model. Note that in each entry in column (1) there is a log deviation of the named category with respect to a geometric mean that always includes the first category, Catholics.

In column (2) the coefficient for Hindus, highlighted in bold, corresponds to the log deviation between that category and the geometric mean of all the others. The category Other Religions, is the log deviation between that category and the geometric mean and the eight remaining categories, and the other coefficients follow the same pattern as in column (1), except that it is now the Protestant category, rather than the Catholic, that always appears in every geometric mean. In each subsequent column the coefficients that correspond to the log deviation between that category and all of the nine remaining categories are in ascending order, culminating in column (10) where the coefficient for Catholics, highlighted in bold, represents that category relative to all nine remaining categories. The missing categories along the diagonal in each column are the ones that feature in every geometric mean. There are other ways to generate the $D = 10$ regressions we want, but as mentioned above, this is the pattern that most closely matches the permutations in Tables 1 and 2.

As in Table 2, in Table 5 only the Buddhist and Orthodox Christian categories, in columns (4) and (8) respectively, are statistically significant. Indeed, closer inspection reveals that along the diagonal in bold in Table 5 the coefficients and standard errors are slightly larger than the corresponding estimates in Table 2 by the same fixed factor. To see why, consider the different

elements that constitute the inverse of the matrix $\mathbf{G}_{1,h}$ the components of which take the form:

$$\left\langle \mathbf{G}_{1,h}^{-1} \right\rangle_{i,j} = \begin{cases} j < h-1 & \begin{cases} i \leq j & -\frac{1}{\sqrt{(j+1)j}} \\ i = j+1 & \sqrt{\frac{j}{j+1}} \\ i > j+1 & 0 \end{cases} \\ j = h-1 & \begin{cases} i < h & -\sqrt{\frac{j+1}{j}} \\ i \geq h & -\sqrt{\frac{j}{j+1}} \end{cases} \\ h \leq j < D & \begin{cases} i < j & 0 \\ i = j & \sqrt{\frac{j+1}{j}} \\ j < i < D & \frac{1}{\sqrt{(j+1)j}} \\ i = D & 0 \end{cases} \\ j = D & \begin{cases} i < D & 0 \\ i = D & 1 \end{cases} \end{cases} \tag{29}$$

In Appendix A.2 there are three examples of what (29) looks like, for the cases $\mathbf{G}_{1,1}^{-1}$ in (33), $\mathbf{G}_{1,D-1}^{-1}$ in (34) and $\mathbf{G}_{1,D}^{-1}$ in (35). In each, the salient part of is the $D-1$ row. If $h < D$ as in (33) and (34), all the entries are zero except for the one that corresponds to $\left\langle \mathbf{G}_{1,h}^{-1} \right\rangle_{D-1,D-1}$, which equals $\sqrt{\frac{D+1}{D}}$. Multiplying this row by the coefficients for the ALR estimation $\tilde{\boldsymbol{\beta}}_{C/d}$ yields $\sqrt{\frac{D+1}{D}}\tilde{\beta}_{C/d}^{D}$. If $h = D$ as in (35), then the last entry is still zero, but all the other entries in that row equal $-\sqrt{\frac{D+1}{D}}$, which when multiplied by $\tilde{\boldsymbol{\beta}}_{C/D}$ yields $-\sqrt{\frac{D+1}{D}}\sum_{d=1}^{D-1}\tilde{\beta}_{C/D}^{d}$. Since the entire set of ALR coefficients sums to zero, this too equals $\sqrt{\frac{D+1}{D}}\tilde{\beta}_{C/d}^{D}$. Hence, having estimated the model using additive log-ratios as in (13) for any possible value of d, the last ILR coefficient, $\bar{\beta}_{C/D}$, which captures the impact of the log deviation of $\bar{x}_{C/D}$ from the geometric mean of all the other $D-1$ variables, can be easily derived as $\bar{\beta}_{C/D}=\sqrt{\frac{D+1}{D}}\tilde{\beta}_{C/D}$. Furthermore, this result holds when the compositional variables are reordered, or can be generalised from $\mathbf{G}_{1,d}^{-1}$ to $\mathbf{G}_{h,d}^{-1}$ to cover all the other variables. If $h \neq D-d+1$, all the elements in the $D-1$ row are equal to zero except the one in the $D-h+1$ column, which is equal to $\sqrt{\frac{D+1}{D}}$. Where $h = D-d+1$ the first $D-1$ entries are equal to $-\sqrt{\frac{D+1}{D}}$ and the last to zero. By the same logic, the variance, $\mathrm{Var}\left(\bar{\boldsymbol{\beta}}_{C/h}\right)$, in (28) equals $\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_{C/d}\right)$ in (15), multiplied by $\frac{D+1}{D}$. Hence the t-statistics associated with all the coefficients in the model using ILR coordinates are identical to those using ALR coordinates.

What does all this mean? Aitchison (1988) advocated using additive log-ratios over isometric log-ratios on the grounds of simplicity of interpretation and calculation, an argument reiterated by Greenacre et al. (2021) and Greenacre et al. (2022). While this concern is possibly valid for other applications, this is not the case for the case of multiple regression examined here, even if only a subset of the variables are compositional. Simply put, it is not necessary to transform the D composite variables into D sets of pivot log-ratios and estimate each permutation of the model separately, as in Table 5 or Hron et al. (2012). Estimating two permutations of equation (13)—any two columns in Table 2 will suffice—or the constrained regression (17) once, as described in Section 5, and then inflating both the coefficients and standard errors

associated with the composition variables by the simple factor $\sqrt{\frac{D+1}{D}}$, is sufficient to convert the regressions with ALR coordinates in Aitchison and Bacon-Shone (1984), into those that would result from converting the data to D sets of pivot log-ratios and estimating each separately to obtain the coefficients in bold in Table 5. As in Section 5, counterfactual experiments are best conducted using log differentials as in (26) and not in terms of logs themselves (as in (17)), to ensure that the compositions remain within the simplex. Either way, the resulting statistical model is permutation invariant and the estimated coefficients are associated with data that are coordinates with respect to an orthogonal basis—the remaining distortions that ALR does not eliminate are removed.

7 Expressing ILR Coefficients as Coordinates in the Simplex

Using additive log-ratios in an estimation requires a change in how we use and interpret regressions with compositional data. The regression coefficients are elasticities and so shifts in the composition are in terms of percentages of fractions (or percentages), rather than the fractions (percentages) themselves, relative to the share chosen as a base. Any one of the permutations of the ILR regression (26) is bound to be less intuitive than ALR, which is why building a model where each coefficient is the elasticity of a share relative to the geometric mean of all the other shares is probably more useful. In Section 6 we demonstrate that it is not necessary to follow Hron et al. (2012) and estimate D different regressions; it is sufficient to estimate (17) and multiply the coefficients and standard errors corresponding to the compositional variables by $\sqrt{\frac{D+1}{D}}$.

Nonetheless, calculating one permutation of (26) is still potentially useful. Extending Van den Boogaart and Tolosana-Delgado (2013), we can restate the coefficients associated with the compositional data as coordinates in the simplex. This yields an alternative expression analogous to (26):

$$\boldsymbol{y} = \mathbf{N}\tilde{\boldsymbol{\beta}}_N + \bar{\beta}_{D+1} + \langle \mathbf{C}_{\setminus D+1}, \bar{\boldsymbol{b}}_C \rangle_a + \bar{\boldsymbol{\varepsilon}} \tag{30}$$

where $\bar{\beta}_{D+1}$ is the same constant term and $\bar{\boldsymbol{b}}_C = \exp[\bar{\boldsymbol{\beta}}_{C/h}\bar{\mathbf{U}}_h]/(\mathbf{i}'\exp[\bar{\boldsymbol{\beta}}_{C/h}\bar{\mathbf{U}}_h])$ is a point in the simplex whose Aitchison product with each point in the dataset $\mathbf{C}_{\setminus D+1}$ yields the same values as the Cartesian product $\bar{\mathbf{C}}_{/h}\bar{\boldsymbol{\beta}}_{C/h}$. The isometry between $\mathbf{C}_{\setminus D+1}$ and $\bar{\mathbf{C}}_{/h}$ means that this is the same for all values of h. For the case of $h = 1$, the D coordinates for regression associated with $\bar{\mathbf{U}}_1'$ can be calculated using the formula:

$$\bar{b}_C^j = \frac{\sqrt{j-1}}{\sqrt{j}}\bar{\beta}_{C/1}^j - \sum_{i=j+1}^{D} \frac{\bar{\beta}_{C/1}^i}{\sqrt{i-1}\sqrt{i}} \tag{31}$$

.

The results for our example yield a point $\bar{\boldsymbol{b}}_C$ in the simplex whose coordinates are: Catholic: 0.0997; Protestant: 0.0998; Orthodox: 0.102; Other Christian: 0.1001; Muslim: 0.0991; Jewish: 0.1001; Buddhist: 0.1027; Hindu: 0.0998; Other Eastern Religions: 0.0993; Other Religions: 0.0973. Note that these are best understood in relation to the barycentre in the simplex

$\{0.1, 0.1, ..., 0.1\}$. Expressing the coefficients that measure the impact of the compositional data on the dependent variable as a vector inside the simplex itself can provide additional intuition as long as one remembers that the Aitchison product between that vector and the data behaves differently. For example, where the coefficients in Table 2 are negative the coordinates here are below 0.1 and where positive above, but this need not always be the case even for what is effectively the same model. Suppose the dependent variable, the rate of growth, is multiplied by 100 and expressed as percentages; the values of all the coefficients in (30) scale up by 100 as well. However, the coordinates of $\bar{\boldsymbol{b}}_C$ (power) scale further away from 0.1 within the simplex according to Aitchison, not Euclidean geometry:

$$\exp[\lambda \times \bar{\boldsymbol{\beta}}_{C/h} \bar{\mathbf{U}}_h] / (\mathbf{i}' \exp[\lambda \times \bar{\boldsymbol{\beta}}_{C/h} \bar{\mathbf{U}}_h]) = ((\bar{b}_C^1)^\lambda, (\bar{b}_C^2)^\lambda, ..., (\bar{b}_C^D)^\lambda). \tag{32}$$

Setting λ=100, the new coordinates are Catholic: 0.0266; Protestant: 0.0307; Orthodox: 0.2602; Other Christian: 0.0395; Muslim: 0.0147; Jewish: 0.0425; Buddhist: 0.5363; Hindu: 0.0181; Other Eastern Religions: 0.0292; Other Religions: 0.0023.

In Table 5, the coefficients for Other Christian and Jewish are positive, but when λ=100, their associated coordinates in the nine dimensional unit simplex now both fall below 0.1. Note that owing to the nonlinearity and nonmonotonicity of (31), we cannot use it to translate the confidence intervals for each coefficient in (23) into confidence intervals in the simplex. Pawlowsky-Glahn et al. (2015) demonstrate, for the case of $D = 3$, how to draw an ellipse that corresponds to a confidence region for the data inside the ternary diagram. Using the variance-covariance matrix from (23), one could do the same for the vector of the coefficients expressed as a vector in the simplex, and even extend this to a three dimensional ellipsoid when $D = 4$. Higher-dimension hyperellipsoids that correspond to confidence regions for any $D > 4$ also exist, but cannot be visualised.

8 Multicollinearity

Incorporating compositional data in regressions as log-ratios, rather than as raw shares, means we introduce extra information into the model—the data are not unrelated points in Euclidean space, but rather coordinates in a simplex. In this section we consider whether the additional information introduced by these procedures, risks making the estimates of the coefficients related to the compositional data less precise, by exacerbating the problem of multicollinearity that results from the intrinsic correlation between the shares in any set of compositional variables.

Ordinary sets of explanatory variables in a regression, of the type that can be described as coordinates in Euclidean space, can be correlated to a degree that complicates estimation. With data that are points in a simplex, this problem is obviously more acute. Omitting one variable from $\mathbf{C}_{\setminus D+1}$ or transforming the data into logs or log-ratios makes it possible to overcome the perfect linear relationship between the variables and incorporate the compositional data into a regression. However, none of these procedures can suppress the inherent correlation between the variables.

To see this, consider a generic set of D compositional variables, each distributed according to a symmetric Dirichlet function $\{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_D\} \sim \mathcal{D}(\alpha)$. The probability density function is $\frac{\Gamma(D\alpha)}{\Gamma(\alpha)^D} \prod_{i=1}^{D} \mathbf{z}_i^{\alpha-1}$, where $\sum_{i=1}^{D} \mathbf{z}_i = \mathbf{i}_D$ and the special case of $\alpha = 1$ represents the uniform distribution. The expected value and variance of each variable are $\mathrm{E}(\mathbf{z}_i) = \frac{1}{D}$ and $\mathrm{Var}(\mathbf{z}_i) = \frac{(D-1)}{D(1+D\alpha)}$; the covariance between two variables $i \neq j$ is $\mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j) = -\frac{1}{D^2(1+D\alpha)}$. This means that whether or not there is a strong correlation between the different categories in a given composition, the compositional nature itself imposes an underlying pairwise correlation between any two variables: $\mathrm{Corr}(\mathbf{z}_i, \mathbf{z}_j) = -\frac{1}{D(D-1)}$.

By contrast, the expected value and variance of any set of $D-1$ log-ratios applied to the same set of compositional variables is $\mathrm{E}(\log(\mathbf{z}_i \oslash \mathbf{z}_d)) = 0$ and $\mathrm{Var}(\log(\mathbf{z}_i \oslash \mathbf{z}_d)) = 2\varphi(\alpha)$, where the logarithm is again with respect to each element in the vector and $\varphi(\alpha) = \frac{\partial^2 \ln \Gamma(\alpha)}{\partial^2 \alpha}$ is the trigamma function. The covariance between two log-ratios $i \neq j \neq d$ is $\mathrm{Cov}(\log(\mathbf{z}_i \oslash \mathbf{z}_d), \log(\mathbf{z}_i \oslash \mathbf{z}_d)) = \varphi(\alpha)$. Whereas the pairwise correlation between any two variables in the compositional dataset is negative and its absolute magnitude diminishes rapidly as the overall number of variables grows larger, the correlation between any pair of log-ratios of these variables is positive and fixed at $\mathrm{Corr}(\log(\mathbf{z}_i \oslash \mathbf{z}_d), \log(\mathbf{z}_i \oslash \mathbf{z}_d)) = \frac{1}{2}$. Does this mean that the cost of switching to log-ratios is to introduce a higher degree of multicollinearity? The answer is no.

To see why, it is best to consider the main measure used to quantify multicollinearity in regression, the variance inflation factor (VIF).[17] Assume each vector $\mathbf{z}_i$ is length n and define the $n \times (D+1)$ matrix $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_D, \mathbf{i}_n\}$ and the $n \times D$ matrix $\mathbf{Z}_{\setminus d} = \mathbf{Z}\mathbf{Q}_{D,d}$ which is the matrix $\mathbf{Z}$ with the $d \in \{1, ..., D+1\}$ column removed. Further define $\mathbf{Z}_{\setminus d,i} = \mathbf{Z}_{\setminus d}\mathbf{Q}_{D-1\setminus i}$, $i \neq d$ as the matrix $\mathbf{Z}$ with both columns d and i removed. Solving the normal equations

$$\mathbf{Z}'_{\setminus d\setminus i}\mathbf{Z}_{\setminus d\setminus i}\beta^z_{\setminus d\setminus i} = \mathbf{Z}'_{\setminus d\setminus i}\mathbf{z}_i$$

where

$$\mathbf{Z}'_{\setminus d\setminus i}\mathbf{Z}_{\setminus d\setminus i} = \frac{n}{D(1+D\alpha)} \begin{bmatrix} 1+\alpha & \alpha & \alpha & \cdots & \alpha & 1+D\alpha \\ \alpha & 1+\alpha & \alpha & \cdots & \alpha & 1+D\alpha \\ \alpha & \alpha & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \alpha & \vdots \\ \alpha & \alpha & \cdots & \alpha & 1+\alpha & 1+D\alpha \\ 1+D\alpha & \cdots & \cdots & \cdots & 1+D\alpha & D(1+D\alpha) \end{bmatrix}$$

and

$$\mathbf{Z}'_{\setminus d\setminus i}\mathbf{z}_i = \frac{n}{D(1+D\alpha)} \begin{bmatrix} (1+D\alpha) \\ \alpha \\ \alpha \\ \vdots \\ \vdots \\ \alpha \end{bmatrix}$$

17. What is notable in the estimates that use ALR and ILR in Tables 2 and 5 is that they are not permutation invariant, though particularly the latter varies a great deal less than the measures of VIF in Table 1. Still, though in some permutations in Table 1 the value of the VIF is very high, some are lower than for the corresponding log-ratios.

yields the vector of $D-1$ coefficients $\beta^z_{\setminus d \setminus i} = \left[\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, ..., -\frac{1}{2}\right]$, and the coefficient of determination of the regression of a each vector $\mathbf{z}_i$ on the matrix $\mathbf{Z}_{\setminus d,i}$ is:

$$R^2_{d,i} = \frac{\beta^{z\prime}_{\setminus d \setminus i}\mathbf{Z}'_{\setminus d \setminus i}\mathbf{Z}_{\setminus d \setminus i}\beta^z_{\setminus d \setminus i} - n\mathrm{E}\left(\mathbf{z}_i\right)^2}{\mathbf{z}'_i\mathbf{z}_i - n\mathrm{E}\left(\mathbf{z}_i\right)^2} = \frac{D-2}{2\left(D-1\right)}.$$

Hence the VIF$=1/\left(1-R^2_{d,i}\right)$, for a hypothetical regression with the matrix of compositional data $\mathbf{Z}_{\setminus d}$ as the only explanatory variables, is equal to $2\frac{D-1}{D}$.

Similarly the matrix $\tilde{\mathbf{Z}}_{/d \setminus i}$ is the matrix $\tilde{\mathbf{Z}}_{/d}$ of additive log-ratios with respect to $\mathbf{z}_d$, and a last column of ones with the $i \neq d$ column removed. Solving the normal equations:

$$\tilde{\mathbf{Z}}'_{/d \setminus i}\tilde{\mathbf{Z}}_{/d \setminus i}\tilde{\beta}^z_{/d \setminus i} = \tilde{\mathbf{Z}}'_{/d \setminus i}\tilde{\mathbf{z}}_i$$

where

$$\tilde{\mathbf{Z}}'_{/d \setminus i}\tilde{\mathbf{Z}}_{/d \setminus i} = n\begin{bmatrix} 2\varphi\left(\alpha\right) & \varphi\left(\alpha\right) & \cdots & \cdots & \varphi\left(\alpha\right) & 0 \\ \varphi\left(\alpha\right) & 2\varphi\left(\alpha\right) & \cdots & \cdots & \varphi\left(\alpha\right) & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots & \vdots \\ \varphi\left(\alpha\right) & \varphi\left(\alpha\right) & \cdots & \cdots & 2\varphi\left(\alpha\right) & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

and

$$\tilde{\mathbf{Z}}'_{/d \setminus i}\tilde{\mathbf{z}}_i = n\begin{bmatrix} 0 \\ \varphi\left(\alpha\right) \\ \varphi\left(\alpha\right) \\ \vdots \\ \vdots \\ \varphi\left(\alpha\right) \end{bmatrix}$$

yields the vector of $D-1$ coefficients $\tilde{\beta}^{z\prime}_{/d \setminus i} = [0, 1, 1, ..., 1]$. Though the matrices appear very different, the corresponding coefficient of determination is the same:

$$\tilde{R}^2_{d,i} = \frac{\tilde{\beta}^{z\prime}_{/d \setminus i}\tilde{\mathbf{Z}}'_{/d \setminus i}\tilde{\mathbf{Z}}_{/d \setminus i}\tilde{\beta}^z_{/d \setminus i} - n\mathrm{E}\left(\tilde{\mathbf{z}}_i\right)^2}{\tilde{\mathbf{z}}'_i\tilde{\mathbf{z}}_i - n\mathrm{E}\left(\tilde{\mathbf{z}}_i\right)^2} = \frac{D-2}{2\left(D-1\right)}.$$

Though the pair-wise correlations between the additive log-ratio vectors are much greater, the VIF is once again equal to $2\frac{D-1}{D}$. This means that switching to log-ratios, whether ALR or ILR (for which the results above would be identical), does not in itself systematically introduce more multicollinearity into the estimation.

9 Conclusion

Two dimensional representations of a three-dimensional object can look very different, depending on the angle at which the object is painted or photographed. If we were to analogize Table 1 to a film set, column (7)—where all but two of the religious categories are at minimum statistically significant at the 10% level, with three at the 1% level—would be the view the director would wish to be captured by the camera. Columns (6) and (9), where none of the religious categories appear significant, would be the very same film set, but viewed from the direction of the backlot. The latitude to choose which share in a compositional data set to omit allows

researchers to "reach for the stars," in a way that can subtly exaggerate the significance of the findings. Moreover, treating compositional data as though they represent vectors in Euclidean space means ignoring the fact that they are actually points in a simplex. The resulting regression may be presented as a hyperplane in a vector space, but it is not.

In this paper I offer several ways to incorporate compositional data alongside noncompositional covariates, as they often appear together in applied work. All of these are based on using log-ratios as first introduced by John Aitchison in the 1980's (Aitchison (1982, 1986)) and further extend the subsequent work done by others, particularly Egozcue et al. (2003) and Hron et al. (2012). The simplest to implement is ALR and this can be extended to a contrained regression in logarithms of the shares. Using ILRs ensures that not only does the data enter the regression as coordinates in a vector space, but that these are coordinates with respect to an orthogonal basis. Finally, I demonstrate that it is a simple way to translate a regression that uses ALR's into one where all the coefficients that relate to compositions are log-ratios with respect to the geometric means of all the remaining variables. The properties of ALR apply equally to this symmetric version of a regression with ILR coordinates. The choice between the two is not necessarily dispositive, the former are easier to describe and explain, the latter benefit from being coordinates with respect to a more appropriate, orthogonal basis. For a composition with a sufficiently large numbers of shares, the choice is somewhat moot—the t-statistics are identical and the coefficients and standard errors themselves differ by less than ten per cent for a composition with six or more parts.

There are a small number of papers in the economics literature that incorporate some version of log-ratios—Fry et al. (1996) use log-ratios to examine household budgeting, Jackson and Khaled (2017) to analyse labour force statistics, and Kynčlová et al. (2015) integrates ILR in a vector autoregression. While the examples here all relate to econometric growth regressions, particularly those developed by Barro and McCleary (2003) and McCleary and Barro (2006), the emphasis throughout is not on explaining economic growth, or how it relates to population shares of religious adherents, but to demonstrate the broader usefulness of these methods within empirical economics and beyond so that they may be adopted more widely and perhaps further refined.

References

Aitchison, John. 1982. "The statistical analysis of compositional data." *Journal of the Royal Statistical Society: Series B (Methodological)* 44 (2): 139–160.

———. 1986. *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability (Reprinted in 2003).* Chapman / Hall London.

———. 1988. "The Single Principle of Compositional Data Analysis, Continuing Fallacies, Confusions and Misunderstandings and Some Suggested Remedies." *Keynote address, CODAWORK08.*

Aitchison, John. 2008. "The single principle of compositional data analysis, continuing fallacies, confusions and misunderstandings and some suggested remedies." *Universitat de Girona. Departament d'Informàtica i Matemàtica Aplicada.*

Aitchison, John, and John Bacon-Shone. 1984. "Log contrast models for experiments with mixtures." *Biometrika* 71 (2): 323–330.

Aitchison, John, and Sheng M Shen. 1980. "Logistic-normal distributions: Some properties and uses." *Biometrika* 67 (2): 261–272.

Barro, Robert J. 1991. "Economic growth in a cross section of countries." *The Quarterly Journal of Economics* 106 (2): 407–443.

———. 1996. "Democracy and growth." *Journal of Economic Growth* 1 (1): 1–27.

Barro, Robert J, and Rachel M McCleary. 2003. "Religion and Economic Growth." *American Sociological Review* 68:760–781.

Billheimer, Dean, Peter Guttorp, and William F Fagan. 2001. "Statistical interpretation of species composition." *Journal of the American Statistical Association* 96 (456): 1205–1214.

Bose, Niloy, M Emranul Haque, and Denise R Osborn. 2007. "Public expenditure and economic growth: A disaggregated analysis for developing countries." *The Manchester School* 75 (5): 533–556.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star wars: The empirics strike back." *American Economic Journal: Applied Economics* 8 (1): 1–32.

Cavallo, Eduardo, and Christian Daude. 2011. "Public investment in developing countries: A blessing or a curse?" *Journal of Comparative Economics* 39 (1): 65–81.

Chen, Jiajia, Xiaoqin Zhang, and Shengjia Li. 2017. "Multiple linear regression with compositional response and covariates." *Journal of Applied Statistics* 44 (12): 2270–2285.

Devarajan, Shantayanan, Vinaya Swaroop, and Heng-fu Zou. 1996. "The composition of public expenditure and economic growth." *Journal of Monetary Economics* 37 (2): 313–344.

Egozcue, Juan José, Vera Pawlowsky-Glahn, Glòria Mateu-Figueras, and Carles Barcelo-Vidal. 2003. "Isometric logratio transformations for compositional data analysis." *Mathematical Geology* 35 (3): 279–300.

Fry, Jane M, Tim RL Fry, and Keith R McLaren. 1996. "The stochastic specification of demand share equations: Restricting budget shares to the unit simplex." *Journal of Econometrics* 73 (2): 377–385.

Greenacre, Michael. 2018. *Compositional data analysis in practice.* CRC Press.

Greenacre, Michael, Eric Grunsky, and John Bacon-Shone. 2021. "A comparison of isometric and amalgamation logratio balances in compositional data analysis." *Computers & Geosciences* 148.

Greenacre, Michael, Eric Grunsky, John Bacon-Shone, Ionas Erb, and Thomas Quinn. 2022. "Aitchison's Compositional Data Analysis 40 Years On: A Reappraisal." *arXiv preprint arXiv:2201.05197.*

Hall, Robert E, and Charles I Jones. 1999. "Why do some countries produce so much more output per worker than others?" *The Quarterly Journal of Economics* 114 (1): 83–116.

Hron, Karel, Peter Filzmoser, and Katherine Thompson. 2012. "Linear regression with compositional explanatory variables." *Journal of Applied Statistics* 39 (5): 1115–1128.

Ioannidis, John PA, TD Stanley, and Hristos Doucouliagos. 2017. "The Power of Bias in Economics Research." *Economic Journal* 127 (605): 236–265.

Jackson, L Fraser, and Mohammed S Khaled. 2017. "Plotting labour force status shares: Interdependence and ternary plots."

Kynčlová, Petra, Peter Filzmoser, and Karel Hron. 2015. "Modeling compositional time series with vector autoregressive models." *Journal of Forecasting* 34 (4): 303–314.

Lindh, Thomas. 1999. "Age structure and economic policy: The case of saving and growth." *Population Research and Policy Review* 18 (3): 261–277.

Lindh, Thomas, and Bo Malmberg. 2009. "European Union economic growth and the age structure of the population." *Economic Change and Restructuring* 42 (3): 159–187.

Lindley, Dennis V. 1985. *Making Decisions.* 2nd ed. Wiley.

Mankiw, N Gregory, David Romer, and David N Weil. 1992. "A contribution to the empirics of economic growth." *The Quarterly Journal of Economics* 107 (2): 407–437.

Maoz, Zeev, and Errol A. Henderson. 2013. "The World Religion Dataset, 1945-2010: Logic, Estimates, and Trends." *International Interactions* 39 (3): 265–291.

McCleary, Rachel M, and Robert J Barro. 2006. "Religion and economy." *Journal of Economic Perspectives* 20 (2): 49–72.

Noland, Marcus. 2005. "Religion and economic performance." *World Development* 33 (8): 1215–1232.

Pawlowsky-Glahn, Vera, Juan José Egozcue, and Raimon Tolosana-Delgado. 2015. *Modeling and Analysis of Compositional Data.* John Wiley & Sons.

Petrakis, Panagiotis E, and Dimitrios Stamatakis. 2002. "Growth and educational levels: a comparative analysis." *Economics of Education Review* 21 (5): 513–521.

Putterman, Louis, and David N Weil. 2010. "Post-1500 population flows and the long-run determinants of economic growth and inequality." *The Quarterly Journal of Economics* 125 (4): 1627–1682.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions rule: the primacy of institutions over geography and integration in economic development." *Journal of Economic Growth* 9 (2): 131–165.

Sala-i-Martin, Xavier X. 1997. "I just ran two million regressions." *American Economic Review* 87 (2): 178–83.

Sala-i-Martin, Xavier X, Gernot Doppelhofer, and Ronald I Miller. 2004. "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach." *American Economic Review* 94 (4): 813–835.

Van den Boogaart, K Gerald, and Raimon Tolosana-Delgado. 2013. *Analyzing compositional data with R.* Vol. 122. Springer.

Voigt, Stefan, Jerg Gutmann, and Lars P Feld. 2015. "Economic growth and judicial independence, a dozen years on: Cross-country evidence using an updated set of indicators." *European Journal of Political Economy* 38:197–211.

A Appendix: Further Results

A.1 Aitchison Geometry

Barycentre:	$\left[\frac{1}{D}, \frac{1}{D}, ..., \frac{1}{D}\right]$
Closure:	$\mathcal{C}(\mathbf{z}) = \left[\frac{z_1}{\sum_{i=1}^{D} z_i}, \frac{z_2}{\sum_{i=1}^{D} z_i}, ..., \frac{z_D}{\sum_{i=1}^{D} z_i}\right]$
Perturbation (in place of addition):	$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, x_2 y_2, ..., x_D y_D)$
Powering (in place of multiplication):	$\alpha \odot \mathbf{x} = \mathcal{C}(x_1^{\alpha}, x_2^{\alpha}, ..., x_D^{\alpha})$
Aitchison Inner Product:	$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left[\ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}\right]$
Aitchison Norm:	$\|\mathbf{x}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left[\ln \frac{x_i}{x_j}\right]^2}$
Aitchison Distance:	$d_a(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left[\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j}\right]^2}$
Angle between $\mathbf{x}$ and $\mathbf{y}$:	$\arccos \frac{\langle \mathbf{x}, \mathbf{y} \rangle_a}{\|\mathbf{x}\|_a \|\mathbf{y}\|_a}$

A.2 ILR Inverse Matrices

$$\mathbf{G}_{1,1}^{-1} = \begin{bmatrix} \sqrt{\frac{2}{1}} & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} & \cdots & \cdots & \cdots & \sqrt{\frac{1}{2}} & 0 \\ 0 & \sqrt{\frac{3}{2}} & \sqrt{2 \times 3} & \sqrt{2 \times 3} & \sqrt{2 \times 3} & \cdots & \cdots & \cdots & \sqrt{2 \times 3} & 0 \\ 0 & 0 & \sqrt{\frac{4}{3}} & \sqrt{3 \times 4} & \sqrt{3 \times 4} & \cdots & \cdots & \cdots & \sqrt{3 \times 4} & 0 \\ \vdots & \vdots & \ddots & \ddots & & & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & & & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & 0 & \sqrt{\frac{D-1}{D-2}} & \sqrt{(D-1)(D-2)} & \sqrt{(D-1)(D-2)} & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & \sqrt{\frac{D}{D-1}} & \sqrt{D(D-1)} & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 & \sqrt{\frac{D+1}{D}} & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 \end{bmatrix} \tag{33}$$

$$\mathbf{G}_{1,D-1}^{-1} = \begin{bmatrix} -\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \frac{-1}{\sqrt{2 \times 3}} & \frac{-1}{\sqrt{2 \times 3}} & \sqrt{\frac{2}{3}} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \frac{-1}{\sqrt{3 \times 4}} & \frac{-1}{\sqrt{3 \times 4}} & \frac{-1}{\sqrt{3 \times 4}} & \sqrt{\frac{3}{4}} & & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \cdots & \cdots & \vdots \\ \frac{-1}{\sqrt{(D-1)(D-2)}} & \frac{-1}{\sqrt{(D-1)(D-2)}} & \frac{-1}{\sqrt{(D-1)(D-2)}} & \frac{-1}{\sqrt{(D-1)(D-2)}} & \cdots & \frac{-1}{\sqrt{(D-1)(D-2)}} & \sqrt{\frac{D-2}{D-1}} & 0 & 0 \\ -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & \cdots & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D}{D+1}} & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & \sqrt{\frac{D+1}{D}} & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 1 \end{bmatrix} \tag{34}$$

$$\mathbf{G}_{1,D}^{-1} = \begin{bmatrix} -\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} & 0 & 0 & \dots & \dots & \dots & \dots & 0 \\ \frac{-1}{\sqrt{2\times3}} & \frac{-1}{\sqrt{2\times3}} & \sqrt{\frac{2}{3}} & 0 & \dots & \dots & \dots & \dots & 0 \\ \frac{-1}{\sqrt{3\times4}} & \frac{-1}{\sqrt{3\times4}} & \frac{-1}{\sqrt{3\times4}} & \sqrt{\frac{3}{4}} & & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \dots & \dots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \dots & \dots & \dots & \dots & \vdots \\ \frac{-1}{\sqrt{(D-2)(D-3)}} & \frac{-1}{\sqrt{(D-2)(D-3)}} & \frac{-1}{\sqrt{(D-2)(D-3)}} & \frac{-1}{\sqrt{(D-2)(D-3)}} & \dots & \frac{-1}{\sqrt{(D-2)(D-3)}} & \sqrt{\frac{D-3}{D-2}} & 0 & 0 \\ \frac{-1}{\sqrt{(D-1)(D-2)}} & \frac{-1}{\sqrt{(D-1)(D-2)}} & \frac{-1}{\sqrt{(D-1)(D-2)}} & \frac{-1}{\sqrt{(D-1)(D-2)}} & \dots & \frac{-1}{\sqrt{(D-1)(D-2)}} & \frac{-1}{\sqrt{(D-1)(D-2)}} & \sqrt{\frac{D-2}{D-1}} & 0 \\ -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & \dots & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & -\sqrt{\frac{D+1}{D}} & 0 \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 & 1 \end{bmatrix} \tag{35}$$

B Appendix: Further Figures and Tables

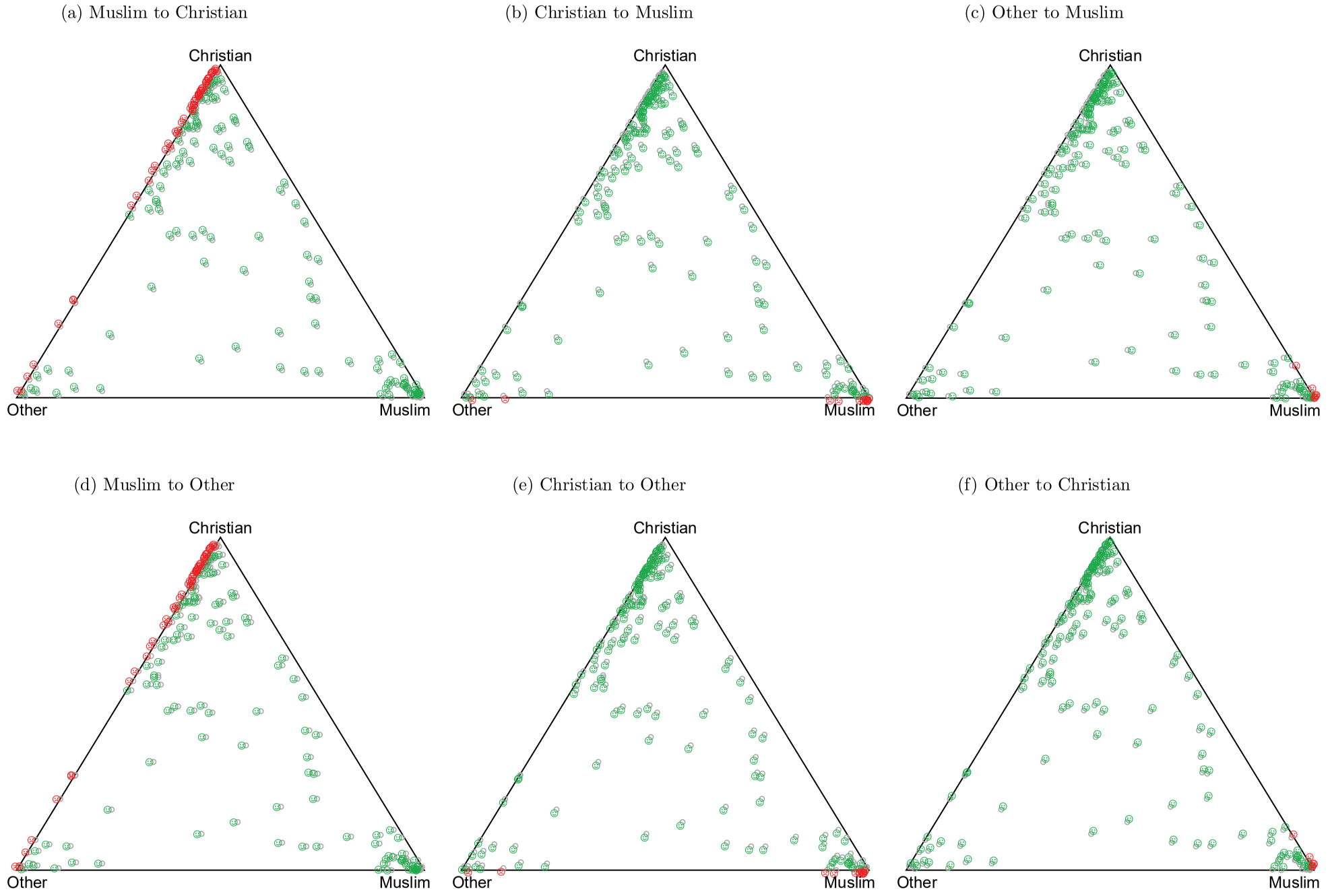


Figure 8: Shifting the religious shares in the three part composition in the ternary diagram, designated by ◦, by one unit leaves some coordinates within the unit simplex, designated by ☺, and other coordinates outside, designated by ☹.

	Zambia	Chad	Nigeria	Eswatini	Ireland	Lebanon	Bosnia	Portugal	S. Korea	Rwanda	Togo	Lithuania
Zambia	0	0.761	0.675	0.086	0.06	0.707	0.566	0.013	0.584	0.047	0.413	0.041
Chad	4.242	0	0.091	0.703	0.804	0.068	0.207	0.75	0.553	0.751	0.435	0.737
Nigeria	4.218	0.437	0	0.620	0.716	0.035	0.116	0.664	0.524	0.662	0.373	0.652
Eswatini	0.440	4.227	4.248	0	0.146	0.653	0.516	0.073	0.498	0.117	0.332	0.046
Ireland	0.588	4.064	3.982	1.002	0	0.748	0.604	0.073	0.644	0.054	0.469	0.1
Lebanon	4.404	0.662	0.27	4.454	4.142	0	0.144	0.697	0.552	0.694	0.407	0.685
Bosnia	4.434	1.221	0.789	4.539	4.103	0.576	0	0.556	0.506	0.551	0.313	0.546
Portugal	0.337	3.906	3.886	0.516	0.579	4.075	4.117	0	0.571	0.053	0.4	0.029
S. Korea	2.413	5.713	5.85	2.046	3.	6.092	6.298	2.561	0	0.610	0.238	0.544
Rwanda	2.096	2.445	2.292	2.269	1.739	2.426	2.365	1.799	4.214	0	0.425	0.079
Togo	2.715	1.888	2.088	2.569	2.751	2.35	2.694	2.392	3.837	1.799	0	0.376
Lithuania	1.702	5.895	5.9	1.673	2.113	6.095	6.136	2.023	1.805	3.792	4.223	0

Table 6: Aitichison distances between countries associated with rectangles in Figure 2 on lower left part of table and Euclidean distances between countries associated with rectangles in Figure 3 on upper right part of table.

	Zambia	Chad	Nigeria	Eswatini	Ireland	Lebanon	Bosnia	Portugal	S. Korea	Rwanda	Togo	Lithuania
Zambia	0	4.817	5.077	0.707	1.019	5.442	5.898	0.372	4.176	2.965	2.725	2.268
Chad	4.242	0	0.697	5.09	4.315	1.129	2.018	4.446	8.062	2.452	2.783	7.003
Nigeria	4.218	0.437	0	5.428	4.453	0.459	1.325	4.713	8.548	2.404	3.325	7.315
Eswatini	0.440	4.227	4.248	0	1.71	5.821	6.351	0.878	3.512	3.489	2.688	1.921
Ireland	0.588	4.064	3.982	1.002	0	4.769	5.117	0.889	5.194	2.15	2.815	3.146
Lebanon	4.404	0.662	0.27	4.454	4.142	0	0.901	5.083	8.984	2.664	3.779	7.693
Bosnia	4.434	1.221	0.789	4.539	4.103	0.576	0	5.557	9.639	2.967	4.538	8.166
Portugal	0.337	3.906	3.886	0.516	0.579	4.075	4.117	0	4.389	2.645	2.393	2.614
S. Korea	2.413	5.713	5.85	2.046	3.	6.092	6.298	2.561	0	6.932	5.292	2.627
Rwanda	2.096	2.445	2.292	2.269	1.739	2.426	2.365	1.799	4.214	0	2.587	5.226
Togo	2.715	1.888	2.088	2.569	2.751	2.35	2.694	2.392	3.837	1.799	0	4.566
Lithuania	1.702	5.895	5.9	1.673	2.113	6.095	6.136	2.023	1.805	3.792	4.223	0

Table 7: Aitichison distances between countries associated with rectangles in Figure 2 on lower left part of table and Euclidean distances between countries associated with the rectangles in Figure 6 associated with ALR coordinates on upper right part of table.

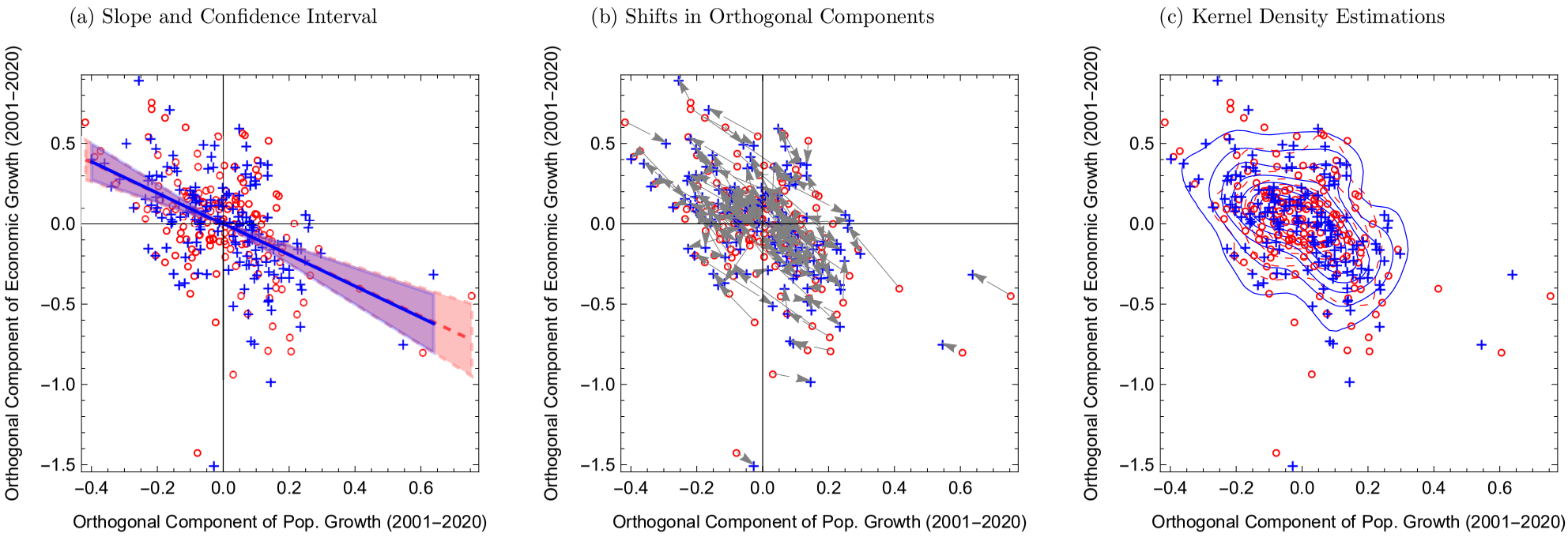


Figure 9: Components of per-capita economic growth between 2001 and 2020 orthogonal to log per-capita GDP in 2001 and orthogonal to average savings rates between 2001 and 2020 as well as religious composition, against the component of population growth rates between 2001 and 2020 orthogonal to the same variables. The red circles ○ represent the regression with an excluded religious category and the blue crosses +, the regressions with the religious composition included as additive log-ratios.

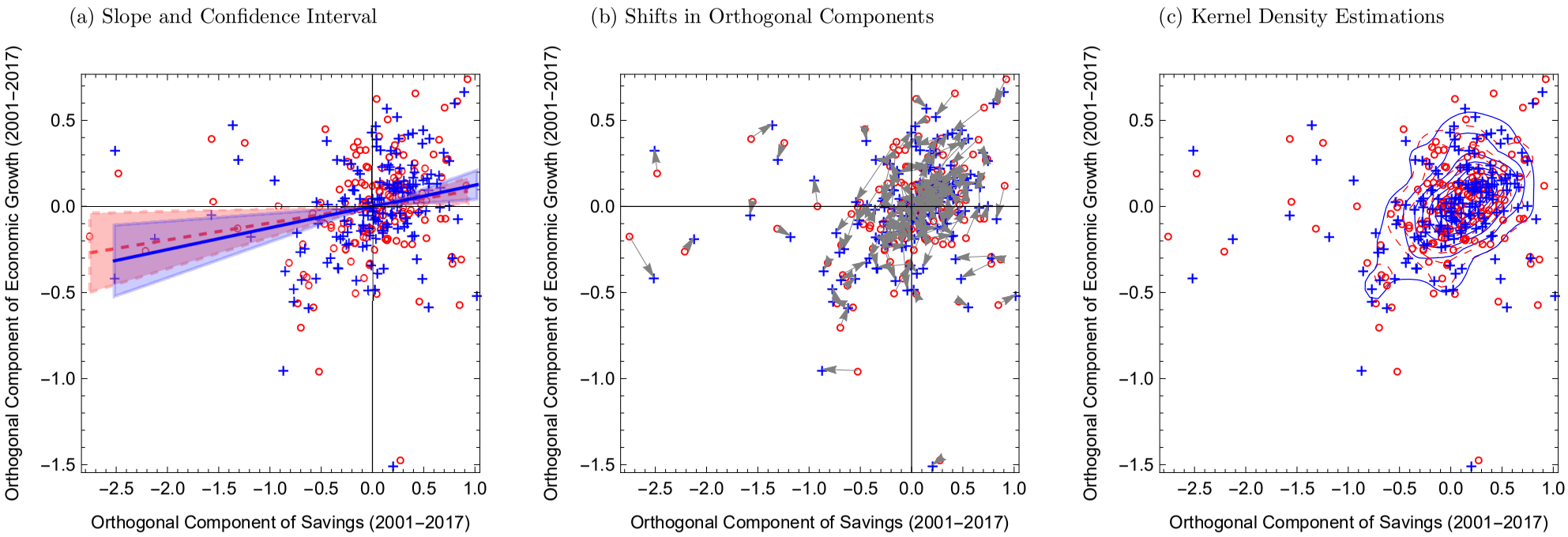


Figure 10: Components of per-capita economic growth between 2001 and 2020 orthogonal to log per-capita GDP in 2001 and population growth rates between 2001 and 2020 as well as religious composition, against the component of average savings rates between 2001 and 2020 orthogonal to the same variables. The red circles ○ represent the regression with an excluded religious category and the blue crosses +, the regressions with the religious composition included as additive log-ratios.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP (2001)	−0.163***	−0.163***	−0.163***	−0.163***	−0.163***	−0.163***	−0.163***	−0.163***	−0.163***	−0.163***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Pop. Growth	−0.989***	−0.989***	−0.989***	−0.989***	−0.989***	−0.989***	−0.989***	−0.989***	−0.989***	−0.989***
	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)	(0.145)
Savings	0.097**	0.097**	0.097**	0.097**	0.097**	0.097**	0.097**	0.097**	0.097**	0.097**
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
chrstcatpct		0.193	−0.166	−0.033	0.151	−0.375	−0.566***	0.673	−0.232	−0.107
		(0.138)	(0.129)	(0.096)	(0.174)	(0.410)	(0.163)	(0.553)	(0.229)	(0.178)
Protestant	−0.193		−0.359**	−0.226*	−0.042	−0.567	−0.759***	0.480	−0.425*	−0.300
	(0.138)		(0.162)	(0.127)	(0.193)	(0.418)	(0.184)	(0.559)	(0.249)	(0.212)
Other Christ.	0.166	0.359**		0.133	0.317	−0.208	−0.400**	0.840	−0.066	0.059
	(0.129)	(0.162)		(0.147)	(0.207)	(0.422)	(0.184)	(0.559)	(0.248)	(0.198)
Sunni Muslim	0.033	0.226*	−0.133		0.184	−0.342	−0.533***	0.706	−0.199	−0.074
	(0.096)	(0.127)	(0.147)		(0.182)	(0.414)	(0.167)	(0.559)	(0.233)	(0.175)
Other Muslim	−0.151	0.042	−0.317	−0.184		−0.525	−0.717***	0.523	−0.383	−0.258
	(0.174)	(0.193)	(0.207)	(0.182)		(0.434)	(0.226)	(0.575)	(0.280)	(0.224)
Jewish	0.375	0.567	0.208	0.342	0.525		−0.192	1.048	0.142	0.267
	(0.410)	(0.418)	(0.422)	(0.414)	(0.434)		(0.434)	(0.680)	(0.460)	(0.435)
Buddhist	0.566***	0.759***	0.400**	0.533***	0.717***	0.192		1.240**	0.334	0.459**
	(0.163)	(0.184)	(0.184)	(0.167)	(0.226)	(0.434)		(0.607)	(0.275)	(0.229)
Other Eastern	−0.673	−0.480	−0.840	−0.706	−0.523	−1.048	−1.240**		−0.905	−0.781
	(0.553)	(0.559)	(0.559)	(0.559)	(0.575)	(0.680)	(0.607)		(0.590)	(0.580)
Hindu	0.232	0.425*	0.066	0.199	0.383	−0.142	−0.334	0.905		0.125
	(0.229)	(0.249)	(0.248)	(0.233)	(0.280)	(0.460)	(0.275)	(0.590)		(0.262)
Other and Non	0.107	0.300	−0.059	0.074	0.258	−0.267	−0.459**	0.781	−0.125	
	(0.178)	(0.212)	(0.198)	(0.175)	(0.224)	(0.435)	(0.229)	(0.580)	(0.262)	
Constant	2.023***	1.830***	2.190***	2.056***	1.873***	2.398***	2.590***	1.350**	2.255***	2.131***
	(0.207)	(0.229)	(0.208)	(0.203)	(0.275)	(0.467)	(0.234)	(0.607)	(0.297)	(0.243)
Observations	164	164	164	164	164	164	164	164	164	164
R^2	0.509	0.509	0.509	0.509	0.509	0.509	0.509	0.509	0.509	0.509
Adjusted R^2	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470
Mean VIF	1.31	1.76	1.80	1.33	2.56	10.37	2.31	18.43	3.80	2.55

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Omitted category regressions (3) with alternative amalgamation. Dependent variable is the growth of per-capita GDP (PPP) from 2001 to 2020.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP (2001)	−0.202***	−0.202***	−0.202***	−0.201***	−0.204***	−0.202***	−0.202***	−0.202***	−0.202***	−0.202***
	(0.024)	(0.024)	(0.024)	(0.036)	(0.028)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Pop. Growth	−1.010***	−1.010***	−1.010***	−0.917***	−1.022***	−1.010***	−1.010***	−1.010***	−1.010***	−1.010***
	(0.153)	(0.153)	(0.153)	(0.227)	(0.175)	(0.153)	(0.153)	(0.153)	(0.153)	(0.153)
Savings	0.106**	0.106**	0.106**	0.115**	0.100**	0.106**	0.106**	0.106**	0.106**	0.106**
	(0.042)	(0.042)	(0.042)	(0.056)	(0.050)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
Catholic		−0.005	−0.005	−0.009	−0.005	−0.005	−0.005	−0.005	−0.005	−0.005
		(0.009)	(0.009)	(0.014)	(0.012)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Protestant	−0.003		−0.003	0.003	0.004	−0.003	−0.003	−0.003	−0.003	−0.003
	(0.010)		(0.010)	(0.016)	(0.013)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Other Christ.	0.010	0.010		0.008	0.001	0.010	0.010	0.010	0.010	0.010
	(0.006)	(0.006)		(0.012)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Other Muslim	−0.003	−0.003	−0.003	−0.013		−0.003	−0.003	−0.003	−0.003	−0.003
	(0.007)	(0.007)	(0.007)	(0.023)		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Sunni Muslim	−0.003	−0.003	−0.003		−0.002	−0.003	−0.003	−0.003	−0.003	−0.003
	(0.005)	(0.005)	(0.005)		(0.014)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Jewish	0.009	0.009	0.009	0.013	0.019		0.009	0.009	0.009	0.009
	(0.010)	(0.010)	(0.010)	(0.015)	(0.012)		(0.010)	(0.010)	(0.010)	(0.010)
Buddhist	0.025***	0.025***	0.025***	0.033***	0.028***	0.025***		0.025***	0.025***	0.025***
	(0.007)	(0.007)	(0.007)	(0.012)	(0.010)	(0.007)		(0.007)	(0.007)	(0.007)
Other Eastern	−0.002	−0.002	−0.002	−0.016	−0.010	−0.002	−0.002		−0.002	−0.002
	(0.012)	(0.012)	(0.012)	(0.020)	(0.017)	(0.012)	(0.012)		(0.012)	(0.012)
Other and Non	−0.022	−0.022	−0.022	−0.021	−0.025	−0.022	−0.022	−0.022	−0.022	
	(0.017)	(0.017)	(0.017)	(0.023)	(0.020)	(0.017)	(0.017)	(0.017)	(0.017)	
Hindu	−0.007	−0.007	−0.007	−0.010	−0.008	−0.007	−0.007	−0.007		−0.007
	(0.007)	(0.007)	(0.007)	(0.011)	(0.009)	(0.007)	(0.007)	(0.007)		(0.007)
Constant	2.541***	2.541***	2.541***	2.385***	2.599***	2.541***	2.541***	2.541***	2.541***	2.541***
	(0.274)	(0.274)	(0.274)	(0.420)	(0.315)	(0.274)	(0.274)	(0.274)	(0.274)	(0.274)
Observations	164	164	164	164	164	164	164	164	164	164
R^2	0.500	0.500	0.500	0.596	0.628	0.500	0.500	0.500	0.500	0.500
Adjusted R^2	0.460	0.460	0.460	0.514	0.576	0.460	0.460	0.460	0.460	0.460
Mean VIF	2.63	2.55	3.61	2.51	4.74	2.52	3.19	1.94	3.29	1.58

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Additive log ratio regressions (13). Dependent variable is log difference of per-capita GDP (PPP) between 2001 to 2020.