ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Weights for Democratic Regional Consumer Price Indices

Andrew Aitken and Martin Weale

ESCoE Discussion Paper 2023-15

August 2023

ISSN 2515-4664

DISCUSSION PAPER

Weights for Democratic Regional Consumer Price Indices
Andrew Aitken and Martin Weale
ESCoE Discussion Paper No. 2023-15
August 2023

Abstract

This paper examines the reliability of regional democratic consumer price indices. The focus on the production of regional weights and the standard error of regional measures of price inflation arising from uncertainty in the regional weights. Issues in the use of modelling techniques applied to expenditure shares are discussed and it is shown that fixed effects methods may be preferable to random effects techniques. The use of small area techniques with mean log household income used as a covariate is compared with purely data-based measures, and it is found that the use of a covariate does not generally result in an increased level of accuracy.

Keywords: Democratic price index, Regional price indices, Small area modelling

JEL classification: C21, C43, R19

Martin Weale, King's College London
Martin.weale@outlook.com

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

Weights for Democratic Regional Consumer Price Indices

Andrew Aitken*and Martin Weale†
ESCoE

August 2023

Abstract

This paper examines the reliability of regional democratic consumer price indices. The focus is on the production of regional weights and the standard error of regional measures of price inflation arising from uncertainty in the regional weights. Issues in the use of modelling techniques applied to expenditure shares are discussed and it is shown that fixed effects methods may be preferable to ran-dom effects techniques. The use of small area techniques with mean log household income used as a covariate is compared with purely data-based measures, and it is found that the use of a covariate does not generally result in an increased level of accuracy.

Keywords; Democratic price index, Regional price indices, Small area modelling.

JEL Codes: C21, C43, R19.

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statis-tical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

This research has been funded by the Office for National Statistics as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE).

*ESCoE and Organisation of Economic Co-operation and Development
†ESCoE and King's College, London. Corresponding author. martin.weale@outlook.com

1 Introduction

There is considerable interest in the production of consumer price indices for regions or smaller areas of the country. Both consumption patterns and price movements may vary geographically. In particular both expenditure on housing and movements in housing costs can be very different in different parts of the country, with the implication that movements in the cost of living may be regionally disparate. Separately, of course, with local data on prices and consumption patterns, it becomes possible to compare living standards in different parts of the country taking into account not only differences in nominal incomes but also differences in the cost of living. The science of making purchasing power parity comparisons between different countries is well developed, while that of comparing different parts of the United Kingdom has not made much progress. On the other hand, with the increasing spread of on-line shopping, there is probably less variability in prices and price movements than would be the case if all shopping took place in person, and it may well be the case that regional differences in prices are most marked for goods and services such as housing which cannot be ordered on line.

This paper explores the problem of estimating regional consumption weights for the construction of a democratic price index (Prais 1958, Aitken & Weale 2020), focusing on the issue of uncertainty in the weights. It follows on from work by Dawber, Würz, Smith, Flower, Thomas, Schmid & Tzavidis (2022). They discussed the problem of producing a consumer index on a regional basis, making use of small area modelling methods to enhance the expenditure data from the Living Costs and Food Survey. They found considerable volatility in regional price indices, but it was not clear how far the source of this was uncertainty in the weights and how far it arose from the fact that the price collection process resulted in material volatility in the elementary aggregates. This study thus complements that piece of work. It considers the special requirements imposed on small area estimation techniques when estimating the weights needed for democratic price indices and examines how far use of modelling techniques can enhance regional weights constructed purely from the data. It identifies the standard errors in aggregate regional price indices which result from uncertainties in the regional average expenditure shares suggesting in fact that weights based solely on data are likely to result in reasonably accurate regional price indices. While the focus is on democratic weights, we do also show how to estimate the standard errors of plutocratic price indices.

Standard small area estimation relies on the production of statistical models which relate the variable of interest to a covariate. If the covariate is available for locations for which the variable of interest is unavailable, such a model can be used directly to produce an estimate for the location in question. But if some observations are also available, then it may be possible to enhance the aggregate produced from these by combining data-based and model-based estimates. Knowing the variance of the aggregate generated by the observations, and also the variance of the estimate generated by fitting the model, it is possible to compute a combination of the two figures weighted by the inverses of the respective variances; this combined estimate will have lower variance than either of the two initial estimates. In essence the local estimate is produced in a way which "draws strength" from the relationship between the aggregate and the covariates as represented by the model.

There is, however, one difficulty to be overcome before applying this sort of approach to the production of regional or small area price indices. In standard small area problems there are few constraints on the type of model that can be used to represent the relationship between the covariates and the small area aggregates for the variables of interest. That is, however, no longer the case if we are concerned about expenditure shares. All of the expenditure shares have to be modelled in a way which guarantees that the model-generated shares add to one. This imposes a sharp constraint on the nature of the models which can be used. But before we discuss that we describe our data and explore the uncertainty in expenditure shares estimated solely from survey data.

2 Data

The expenditure data with which we work are provided by the ONS and derived from the Living Costs and Food Survey (LCFS). The basic data in the LCFS are not consistent with the aggregate expenditure figures shown in the national accounts and consumption data and are provided at COICOP87 level. The data we use are scaled to be consistent with the national accounts and the weights used in the production of CPIH. ONS was also able to make available to us three potential covariates, showing average gas and electricity consumption in each household and decile points of household income adjusted for equivalent household size. These covariates are available on a middle-layer super output area (MSOA) basis. The household income data, available only for 2015/16, are compiled from PAYE and benefit data and are incomplete in that they do not reflect either self-employment or investment income. Since they are available for decile points we use the mean of the logs of income at the decile points to approximate the mean of log income. The income data are available only for England and Wales.

The data for price growth are derived from the consumer price inflation tables (Office for National Statistics 2021). The results here use the version from April 2021.

3 Expenditure Patterns in Summary Data

There are substantial regional variations in expenditure patterns. Tables 1 and 2 show the average of household expenditure shares for the twelve broad COICOP categories in each of the ten regions for which we also have access to small area income data. Even at this level of aggregation some important regional differences stand out. The mean share of income spent on housing in London is 47% while the highest share anywhere else is 36%, in the East of England. A high expenditure share on housing implies less spending on other things. So Londoners spend only 8.3% of their total consumption on food while in the North-East the mean share is 11.4%. Only an average of 7.1% of Londoners' spending goes on recreation while in the North-West 12.3% of consumption is devoted to this purpose.

The standard errors of each of the mean shares are also shown [1]. It is clear from these data that, while there may be substantial variation across households in each of the regions, the regional means at COICOP12 level are determined reasonably accurately and many of the differences in the regional means are unlikely to arise from sampling error. The accuracy of the expenditure shares of course reflects the fact that the standard error of the mean is decreasing in the square root of the sample size.

The fact that the spending patterns are estimated with uncertainty is, of course, a source of uncertainty in estimates of regional inflation rates produced using these estimates of expenditure patterns. We need an estimate of the covariance matrix of the expenditure shares in order to estimate the standard deviations of the regional inflation rates constructed either directly from the regional data or after making use of modelling techniques. This is presented in the next section.

4 Uncertainty in the Weights

A democratic price index is computed from the weighted average of the expenditure patterns of each household, while a plutocratic price index is derived from the shares in total expenditure allocated to each expenditure category. In either case uncertainty in the weights is transmitted into uncertainty in the aggregate measure of inflation for each

[1] These do not take account of effects arising from the design of the LCFS

	North-East	North-West	Yorks	E Mid	W Mid
Food	0.114	0.101	0.1	0.102	0.107
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Drink	0.041	0.046	0.046	0.041	0.036
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
Clothing	0.063	0.055	0.049	0.052	0.053
	(0.005)	(0.003)	(0.003)	(0.003)	(0.004)
Housing	0.284	0.311	0.315	0.307	0.343
	(0.009)	(0.007)	(0.008)	(0.008)	(0.008)
Furnishings	0.058	0.045	0.047	0.052	0.044
	(0.006)	(0.003)	(0.003)	(0.004)	(0.003)
Health	0.014	0.014	0.016	0.023	0.019
	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)
Observations	224	559	449	359	443

	East	London	South-East	South-West	Wales
Food	0.098	0.083	0.085	0.097	0.106
	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)
Drink	0.03	0.024	0.033	0.037	0.037
	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)
Clothing	0.046	0.045	0.044	0.041	0.051
	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
Housing	0.362	0.47	0.353	0.333	0.323
	(0.008)	(0.009)	(0.007)	(0.007)	(0.011)
Furnishings	0.044	0.034	0.048	0.045	0.057
	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)
Health	0.017	0.016	0.024	0.028	0.024
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Observations	445	389	603	464	215

Standard errors in parentheses

Table 1: Expenditure Shares in 2016

	North-East	North-West	Yorks	E Mid	W Mid
Transport	0.114	0.114	0.112	0.125	0.105
	(0.007)	(0.005)	(0.005)	(0.006)	(0.005)
Comms	0.027	0.025	0.027	0.024	0.026
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Recreation	0.125	0.123	0.115	0.104	0.103
	(0.008)	(0.005)	(0.005)	(0.005)	(0.005)
Education	0.002	0.006	0.007	0.004	0.008
	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)
Hotels etc.	0.094	0.087	0.096	0.087	0.081
	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)
Misc	0.064	0.074	0.071	0.078	0.077
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Observations	224	559	449	359	443

	East	London	South-East	South-West	Wales
Transport	0.115	0.086	0.118	0.107	0.103
	(0.005)	(0.004)	(0.004)	(0.004)	(0.007)
Comms	0.024	0.023	0.021	0.023	0.025
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Recreation	0.106	0.071	0.106	0.12	0.105
	(0.005)	(0.004)	(0.004)	(0.006)	(0.006)
Education	0.008	0.008	0.012	0.01	0.004
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Hotels etc.	0.077	0.08	0.087	0.087	0.086
	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)
Misc	0.073	0.059	0.069	0.073	0.079
	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)
Observations	445	389	603	464	215

Standard errors in parentheses

Table 2: Expenditure Shares in 2016 (continued)

region. Before exploring how estimated models of the expenditure shares can help with estimation of weights, we discuss the variance in the inflation rate generated when the sole source of uncertainty concerns the weights themselves. In this section we examine both democratic and plutocratic weights although in the subsequent work we explore only democratic weights.

If we write our vector of weights for region k as $\mathbf{w}_k$ and a vector of percentage price increases for the different expenditure categories, assumed here to be the same in all regions, π, then the measure of the aggregate inflation rate, computed from the vector of component price indices, is measured as $\Pi_k = \pi'\mathbf{w}_k$. It follows that, to the extent that the vector of weights is uncertain, so too will be the resulting measure of aggregate inflation.

We can write the variance of the aggregate as $\mathrm{V}(\Pi_k) = \pi'\mathbf{V}(\mathbf{w}_k)\pi$ where $\mathbf{V(w}_k\mathbf{)}$ is the covariance matrix of the expenditure shares, assuming that the π are known precisely. $\mathrm{V}(\Pi_k)$ depends on the precise pattern of the category-specific price increases. In particular, since the weights sum to 1, $\mathbf{V(w}_j\mathbf{)}$ must be rank-deficient. If all of the price increases were equal, then there would be no uncertainty in the measure of inflation arising from uncertainty in the expenditure shares, whether those weights are democratic or plutocratic. Despite this, one might conjecture that the democratic weights would be determined better than the plutocratic weights, at least in a survey where the observations from each household counted equally. The reason for this is that the democratic weights are calculated by giving equal weight to each household observation, while the plutocratic weights give higher importance to the high-spending households, reducing the effective sample size.

To evaluate the variance matrices of the democratic weights we use the Stata command *mean*. If the weight given to observation i is q_i and $Q = \sum_i q_i$ then the variance of the mean is calculated as

$$V(\bar{y}) = \frac{\sum_i q_i(y_i - \bar{y})}{Q(Q-1)} \tag{1}$$

where $\bar{y}$ is the weighted mean of the variable in question. When the weights are interpreted as sampling weights, as is done here, the q_i are scaled automatically in the programme and the variance estimate is invariant to any rescaling of the weights.

For the plutocratic weights the primary data are the weighted sample means of expenditure in each of the expenditure categories and the plutocratic weights are derived from these, by dividing by the mean of the expenditure total. Stuart & Ord (1986) use a Taylor expansion to derive the covariance of functions of a vector of variables with known covariance matrix, and we use their expression to derive the covariance matrix of the plutocratic weights. We drop the subscript k to simplify the notation.

We use the STATA command *mean*[2] to evaluate the covariance matrix of the vector of the mean amounts of expenditure in each category, μ_i, and the mean value of total expenditure, μ_t. Using $\sigma_{i,j}$ to represent the covariance between mean expenditure in categories i and j, $\sigma_{i,t}$ to denote the covariance between the mean expenditure in category i and total expenditure, and σ_t^2 to denote the variance of total expenditure, we can write the covariance between the expenditure shares, w_i and w_j as

$$Cov(w_i, w_j) = \frac{\mu_i \mu_j \sigma_t^2}{\mu_t^4} + \frac{\sigma_{i,j}}{\mu_t^2} - \frac{\mu_j \sigma_{i,t}}{\mu_t^3} - \frac{\mu_i \sigma_{j,t}}{\mu_t^3}$$

while setting $i = j$ allows us to work out the terms on the diagonal of the covariance matrix.

5 Use of Small Area Techniques in Modelling Expenditure Shares

A standard framework for modelling expenditure shares is provided by the Almost Ideal Demand System (AIDS) developed by Deaton & Muellbauer (1976). Here expenditure shares are assumed to be linear in the logarithm of total expenditure. They also depend on relative prices, but we do not have any data from which to observe price differentials in different MSOAs. Working on a household basis, we replace the log of total expenditure by the mean log of our income covariate in the AIDS model. We can then specify an equation in which the mean expenditure share on each consumption category depends on a constant term and the mean of log income in the relevant MSOA. As noted earlier, the mean of log income can be approximated as the mean of the log of income at each decile.

5.1 Random Effects Models

We can model expenditure shares as

$$y_{jk} = \alpha_j + x_k \beta_j + \varepsilon_{jk}$$

Here y_{jk} is the mean share of expenditure (i.e. the mean in the MSOA of the expenditure shares observed for each household in the LCFS) on consumption category j in small

[2] As noted earlier this does not take account of design effects.

area k. x_k is the mean of logs of the income decile points in small area k and ε_{jk} is an error term. We now explore the requirement that fitted values of expenditure shares

$$\hat{y}_{jk} = \hat{\alpha}_j + x_k\hat{\beta}_j + \hat{\varepsilon}_{jk}$$

sum to 1, $\sum_j \hat{y}_{jk} = 1$. This is going to depend on the estimation technique and on the properties of the $\hat{\varepsilon}_{ij}$.

5.2 Estimation by Weighted Least Squares

Consider a consumer demand system which explains the shares of expenditure on different categories of goods. y_{jk} is consumption in category j by household k. $\mathbf{X}$ is a matrix of characteristics and $\mathbf{X}_k$ are the specific characteristics of household k. ε_{jk} is a random error term. While in the example we consider there is only on explanatory variable we set out here the algebra allowing for multiple explanatory variables. $\mathbf{i}$ is a vector of ones.

Consider a system in which variables are measured relative to their means. The equation for the share of total expenditure on good j is, with the vectors spanning the k households

$$\mathbf{y}_j = \alpha_j\mathbf{i} + \mathbf{X}\beta_j + \varepsilon_j$$

where the OLS coefficient vector is, with n the number of households, then given as

$$\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} = \begin{bmatrix} n & \mathbf{i}'\mathbf{X} \\ \mathbf{X}'\mathbf{i} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{i}' \\ \mathbf{X}' \end{bmatrix} \mathbf{y}_j.$$

The matrix $\begin{bmatrix} n & \mathbf{i}'\mathbf{X} \\ \mathbf{X}'\mathbf{i} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1}$ can be evaluated using the expression for the inverse of a partitioned matrix. We can write

$$\begin{bmatrix} n & \mathbf{i}'\mathbf{X} \\ \mathbf{X}'\mathbf{i} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{1}{n^2}\mathbf{i}'\mathbf{X}\mathbf{U}^{-1}\mathbf{X}'\mathbf{i} & -\frac{1}{n}\mathbf{i}'\mathbf{X}\mathbf{U}^{-1} \\ -\frac{1}{n}\mathbf{U}^{-1}\mathbf{X}'\mathbf{i} & \mathbf{U}^{-1} \end{bmatrix}$$

where $\mathbf{U} = \left[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{i}\mathbf{i}'\mathbf{X}/\mathbf{n}^2\right]$. Also adding over j we can see that $\sum_j \begin{bmatrix} \mathbf{i}' \\ \mathbf{X}' \end{bmatrix} \mathbf{y}_j = \begin{bmatrix} n \\ \mathbf{X}'\mathbf{i} \end{bmatrix}$. It now follows immediately that $\sum_j \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. So when estimated by OLS the constant terms sum to 1 and the other coefficients sum to zero, with the implication that the adding up constraints are satisfied. The out of sample fitted value of $\hat{\varepsilon}_{jk}$ is set to zero.

Now suppose that the model cannot be estimated by ordinary least squares, because cluster-specific random effects are present. In the OLS case we did not make the assumption that the variances of the error terms were the same across the different goods;

we simply assumed that the errors for each good were homoscedastic. Now we assume that, for each category j, the covariance matrix across households is known down to a scaling term, σ_j^2. Thus $\mathrm{E}\left(\varepsilon_j\varepsilon_j'\right) = \mathbf{V}_j = \sigma_j^2\mathbf{V}$. Also assume that $\mathbf{V}^{-1} = \mathbf{S}'\mathbf{S}$ where the factorisation is guaranteed because $\mathbf{V}$ is positive definite. The coefficient vector $[\alpha_j\ \beta_j]$ can be estimated by applying OLS to the following equation

$$\mathbf{S}\mathbf{y}_j = \alpha_j\mathbf{S}\mathbf{i} + \mathbf{S}\mathbf{X}\beta_j + \mathbf{S}\varepsilon_j$$

so that

$$\begin{bmatrix}\alpha_j\\ \beta_j\end{bmatrix} = \begin{bmatrix}\mathbf{i}'\mathbf{V}^{-1}\mathbf{i} & \mathbf{i}'\mathbf{V}^{-1}\mathbf{X}\\ \mathbf{X}'\mathbf{V}^{-1}\mathbf{i} & \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{i}'\mathbf{V}^{-1}\\ \mathbf{X}'\mathbf{V}^{-1}\end{bmatrix}\mathbf{y}_j.$$

Applying the partition once again

$$\begin{bmatrix}\alpha_j\\ \beta_j\end{bmatrix} = \begin{bmatrix}\frac{1}{\mathbf{i}'\mathbf{V}^{-1}\mathbf{i}} + \left(\frac{1}{\mathbf{i}'\mathbf{V}^{-1}\mathbf{i}}\right)^2\mathbf{i}'\mathbf{V}^{-1}\mathbf{X}\mathbf{U}^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{i} & -\frac{1}{\mathbf{i}'\mathbf{V}^{-1}\mathbf{i}}\mathbf{i}'\mathbf{V}^{-1}\mathbf{X}\mathbf{U}^{-1}\\ -\frac{1}{\mathbf{i}'\mathbf{V}^{-1}\mathbf{i}}\mathbf{U}^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{i} & \mathbf{U}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{i}'\mathbf{V}^{-1}\\ \mathbf{X}'\mathbf{V}^{-1}\end{bmatrix}\mathbf{y}_j \qquad (2)$$

so that, adding across the categories of consumption, we can see that, because $\sum_j \mathbf{y}_j = \mathbf{i}$ we have once again that $\sum_j \begin{bmatrix}\alpha_j\\ \beta_j\end{bmatrix} = \begin{bmatrix}1\\ \mathbf{0}\end{bmatrix}$ so that the adding up constraints on the model parameters are satisfied. This result obviously depends on the covariance matrix, subject to scaling, being the same for all the consumption categories, as well as having the same set of explanatory variables $\mathbf{X}$. The fact that the adding up constraints are automatically satisfied means that the parameters need be estimated for all but one of the categories of consumption. It does not matter which category is omitted (Barten 1969).

This proves adding up of the model coefficients but it does not consider specifically the cluster effects. With random cluster effects we can no longer set $\hat{\varepsilon}_{jk}$ to zero. When these take the form assumed by the multi-level model, we can write the model as

$$\mathbf{y}_j = \alpha_j\mathbf{i} + \mathbf{X}\beta_j + \mathbf{Q}\mathbf{v}_j + \mathbf{e}_j.$$

Here $\mathbf{e}_j$ is a vector of error terms with covariance matrix $\sigma_j^2\mathbf{I}$. We look at the case where there is only one random effect with h clusters, giving h different realisations of this random variable. The vector $\mathbf{v}_j$ of length h shows the cluster-specific random terms; their mean is zero. These are mapped into the household observations by the matrix $\mathbf{Q}$ of dimension $n \times h$. $\mathrm{Cov}(\mathbf{v}_j) = \sigma_j^2\sigma_v^2\mathbf{I}_h$ where $\mathbf{I}_h$ is an $h \times h$ unit matrix and the variance of $\mathbf{y}_t$ is given as

$$\mathbf{V}_j = E(\mathbf{y}_j\mathbf{y}_j') = \sigma_j^2\left[\mathbf{I} + \mathbf{Q}\mathbf{Q}'\sigma_v^2\right]$$

When σ_v^2 is known, we can write the estimate of the cluster random effects (Harville 1977, p.322) as

$$\mathbf{v}_j = \sigma_j^{-2}\mathbf{Q}'\mathbf{V}^{-1}(\mathbf{y}_j - \alpha_j\mathbf{i} - \mathbf{X}\beta_j). \tag{3}$$

For the shares to add to one across the different consumption categories, we require $\Sigma_j\alpha_j = 1, \Sigma_j\beta_j = 0$ and $\Sigma_j\mathbf{v}_j = \mathbf{0}$. We have already seen, working from equation (2) that $\Sigma_j\alpha_j = 1$, and $\Sigma_j\beta_j = 0$. For that to hold, we required the covariance structure of the residuals to be the same for all expenditure categories down to a scaling factor, σ_j^2. But we can see from equation (3) that this does not guarantee that $\Sigma_j\mathbf{v}_j = \mathbf{0}$. For this to be the case, we require that the values of σ_j^2 are the same for all consumption categories.

5.3 Testing the Adding Up Conditions

STATA allows us to estimate a multi-level model for each of the expenditure categories. It provides estimates of the logs of the idiosyncratic and cluster variances, σ_j^2 and σ_{jv}^2 and their covariance matrix. We face the limitation that it is not possible to estimate the covariances between the variance terms of two different expenditure categories. Subject to this limitation it is, however, possible to test the hypothesis that the relevant variance terms are equal to each other. We put $\mathbf{w}_j' = [\sigma_j^2\, \sigma_{jv}^2]$ and write its covariance matrix as $\mathbf{W}_j$.

Stacking the $\mathbf{w}_j$ and consolidating the $\mathbf{W}_j$ into a single array, we write

$$\mathbf{w}^* = \begin{bmatrix} \mathbf{w}_1 \\ \cdots \\ \mathbf{w}_j \\ \cdots \\ \mathbf{w}_N \end{bmatrix}, \quad \mathbf{W}^* = \begin{bmatrix} \mathbf{W}_1 \ldots & \mathbf{0} & \ldots & \mathbf{0} \\ \cdots \; \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \ldots \mathbf{W}_j & \ldots & \mathbf{0} \\ \cdots \; \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \ldots \; \mathbf{0} & \ldots & \mathbf{W}_N \end{bmatrix}$$

and put as $\mathbf{A}$ the constraint matrix; when the conditions are met, we will have $\mathbf{Aw^*=0}$. There are 2N-2 constraints to be met, so $\mathbf{A}$ is an array of dimension 2N-2×2N.

The weighted least squares estimate of $\mathbf{w}^{**}$, the estimate of $\mathbf{w}^*$ revised to meet the condition that $\mathbf{Aw}^{**} = \mathbf{0}$ is then given as

$$\mathbf{w}^{**} = \mathbf{w}^* - \mathbf{W}^*\mathbf{A}'(\mathbf{AW}^*\mathbf{A}')^{-1}\mathbf{Aw}^*.$$

It is clear that $\mathbf{Aw}^{**} = 0$. Further, subject to the estimates w^* being normally distributed.

$$(\mathbf{w}^{*\prime} - \mathbf{w}^{**\prime})\mathbf{W}^{*-1}(\mathbf{w}^* - \mathbf{w}^{**}) \tilde{}\, \chi^2_{2N-2}.$$

Thus it is possible to test whether the restrictions needed for the multi-level model to be applied to the system of expenditure share equations are in fact consistent with the data, at least while being aware that the covariance between $\mathbf{w}_j$ and $\mathbf{w}_k$ is neglected. A very large value of the χ^2 statistic would indicate that the data do not support the use of multi-level modelling as a means of representing the cluster effects.

5.4 Modelling with Fixed Effects

As we have seen, the use of multi-level models for estimating a coherent system of expenditure share equations requires very stringent conditions to be met. An obvious, less demanding, approach would be to use a fixed effects model, instead of the multi-level or random effects model, to represent regional effects. Fixed effects can be estimated by OLS and, provided the model coefficients are identified, there will always be a solution. The analysis in section 5.2 shows that, provided the same variables are included in the models for each consumption category, the estimated shares will add to one as required.

The distinction between the fixed and random effects models can be put as follows. With a fixed effects model, the cluster effects are designed to explain as much variance as possible, while with the multi-level model the cluster effects are intended only to explain the variance remaining after the other variables have accounted for as much as possible.

But the approach advocated by Fay & Herriot (1979) as a means of combining observed regional averages with the estimated averages generated by a model is unlikely to be useful in this case. If the survey data are available for all of the small areas for which the covariates are available, then, by design, the fixed effects model will exactly match the observed regional means. A fitted value may have lower variance than the observed regional average, because it also makes use of information from the covariate and this is the means by which the estimate “borrows strength”. Pooling model and data-based estimates will not add to this.

The framework can be set out as follows. We denote

$$\mathbf{Z} = [\mathbf{X} \;\; \mathbf{Q}]$$

where $\mathbf{X}$ is the matrix of exogenous variables as before and $\mathbf{Q}$ has h columns. For each observation the $\mathbf{Q}$ matrix contains a 1 to indicate the regional category from which the observation is drawn and 0 otherwise. Since $\mathbf{Q}$ includes columns for all the regions we do not also include a constant term. With one exogenous variable, small area mean log income, we can immediately write the parameter vector for consumption category j as

$$\beta_j = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_j$$

We stack the values of β_j and $\mathbf{y}_j$ side by side, so that

$$\mathbf{B} = [\beta_1....\beta_j....\beta_N]$$

and

$$\mathbf{Y} = [\mathbf{y}_1....\mathbf{y}_j....\mathbf{y}_N]$$

We can then write

$$\mathbf{B} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$$

and the array of model residuals can be set out as

$$\mathbf{E} = \mathbf{Y} - \mathbf{XB}$$

If we now stack the parameter vectors vertically

$$\mathbf{B}^* = [\beta_1'....\beta_j'....\beta_N']'$$

we can estimate, with $\otimes$ representing the Kronecker product:

$$V(\mathbf{B}^*) = \frac{(\mathbf{E}'\mathbf{E}) \otimes (\mathbf{Z}'\mathbf{Z})^{-1}}{n}.$$

For a cluster r ($r <= h$) we write $\mathbf{z}_r$ as the row vector which contains the category mean of the covariates- in this case the single covariate is the log of net income, and a 1 in the column appropriate to region r with zeros elsewhere. The fitted value of the vector of mean consumption shares in each cluster is the row vector

$$\hat{\mathbf{y}}_r = \mathbf{z}_r B \tag{4}$$

Since the expenditure shares are price index weights, the model generated value of inflation for category r is, with π the vector of percentage price increases for each expenditure component

$$\Pi_r = \hat{\mathbf{y}}_r \pi \tag{5}$$

It follows immediately that

$$V(\Pi_r) = \pi' V(\hat{\mathbf{y}}_r)\pi \tag{6}$$

allowing us to estimate the uncertainty in the inflation estimate due to uncertainty about the expenditure shares.

The easiest way of evaluating $V(\hat{\mathbf{y}}_r)$ is as follows. Reverting to the vector B^* we can write

$$\hat{\mathbf{y}}_r = \begin{bmatrix} \mathbf{z}_r \dots 0 \dots 0 \\ \dots \dots \dots \dots \dots \\ 0 \dots \mathbf{z}_t \dots 0 \\ \dots \dots \dots \dots \dots \\ 0 \dots 0 \dots \mathbf{z}_r \end{bmatrix} \begin{bmatrix} \beta_1 \\ \\ \beta_j \\ \\ \beta_n \end{bmatrix} = \mathbf{Z}_r^* \mathbf{B}^*$$

It then follows that, with

$$V(\hat{\mathbf{y}}_r) = \mathbf{Z}_r^* V(\mathbf{B}^*) \mathbf{Z}_r^{*\prime}$$

giving us the variance of the inflation rate for cluster r based on the expenditure patterns fitted by the model. Of course this can be calculated using also small areas for which there are covariates but not consumption observations. To the extent that the mean of the covariates for the small areas where consumption is observed differs from that where consumption is not observed, there will be a further gain in accuracy. The variance of the regional category inflation rate calculated only from the consumption data is making the assumption that the these data are representative of the regional category as a whole. That is unlikely to be the case if the means of the covariates are different in the two types of small areas.

One point should be mentioned. This model can be estimated with only the dummy variables present, and without any covariate. In that case the categorical fixed effects will equal the categorical means. The variances will, however, differ somewhat from the variances estimated using the covariance matrix for each regional category separately. This is because the fixed effects model assumes that the variance of the observations is the same across all regional categories. That is not the case if the variances for the categories are estimated independently.

Finally it should be noted that this method is perhaps easier to adapt to the presence of multiple categories than is the mixed model. It would be perfectly possible to introduce a second set of dummies which distinguish, for examples, cities, towns and non-urban areas so as to produce price index weights for these. In that case the means of the dummies for the regional categories will reflect the proportion of households in each agglomeration category in each of the regions. The principles of the calculations are nevertheless unaffected.

6 Results

6.1 The Random and Fixed Effects Models Compared

We first explore whether the conditions are met to allow us to ensure that the mixed model can be used consistent with fitted expenditure shares adding to 1. We fit the mixed model to the COICOP12 data since when applied to COICOP87 data the modelling routines do not converge. The results of this are presented in Appendix A using the mean of log household income as a covariate; we find that gas and electricity consumption are not helpful. We then test the hypothesis that the two variance parameters are equal across all twelve consumption categories. The resulting test statistic is $\chi^2_{22} = 1394$ suggesting that the mixed model is unlikely to be helpful or at least consistent with the data. We also note that it is not possible to estimate this model for all eighty-seven expenditure categories. Thus our conclusion is that use of this technique, which is widespread in small area estimation, is unlikely to be helpful for the production of regional expenditure weights.

In Appendix A we also present the coefficients on income in the fixed effects model, estimated on a comparable basis. The most striking feature to be seen on comparing the two is just how similar the two sets of income coefficients are. This of course carries with it the implication that, even though the statistical conditions for the adding up constraint are strongly rejected, the constraint cannot, in this case, be badly broken.

Rather than present the RMSEs of the fixed effects models, we show their logarithms, so as to facilitate comparison with the output of the random effects models. We can see that these idiosyncratic errors are scarcely distinguishable. All of these models are estimated with 4151 observations of household spending for 2016 and income data for 2015/16.

6.2 Data-based Estimates and their Variances

We next examine data-based estimates of regional inflation measured produced by using weights computed only for the households resident in the regions in question. These estimates are produced for both democratic and plutocratic price increases. The standard errors of the former are calculated using the estimated covariances of the mean expenditure shares as provided by STATA. The standard deviations of the latter are produced using the calculations of section 4. Table 3 shows the increase in the price indices for January 2017 over January 2016 calculated using weights derived from the

	Dem Infl	S.E.	Plut Infl	S.E.
North-East	1.601	0.068	1.571	0.077
North-West	1.602	0.040	1.642	0.041
Yorks & Humber	1.711	0.054	1.708	0.047
East Midlands	1.845	0.062	1.811	0.059
West Midlands	1.678	0.048	1.732	0.042
East	1.848	0.068	1.805	0.052
London	1.622	0.043	1.698	0.052
South-East	1.818	0.036	1.797	0.052
South-West	1.816	0.052	1.716	0.061
Wales	1.858	0.086	1.740	0.061

Table 3: Democratic and Plutocratic Measures of CPIH Price Growth and their Standard Errors: Jan 2017 compared to Jan 2016

scaled LCFS data for the calendar year 2016.

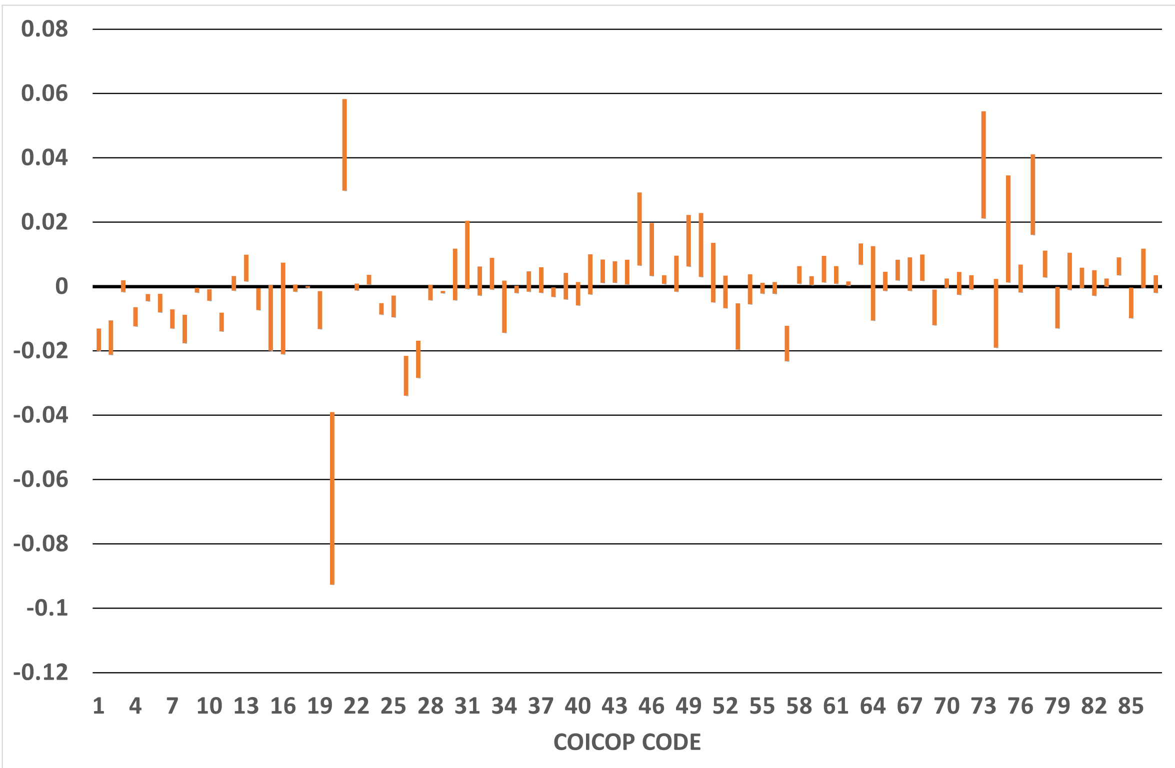
Both democratic and plutocratic measures have standard errors of about 0.04-0.09% p.a., which is probably an acceptable degree of accuracy.

6.3 Fixed Effects

Secondly we look at the results generated by the fixed effects model applied to eighty-seven COICOP expenditure categories. Focussing first on the income variable, we find it significant at a 5 per cent level for 49/87 of the consumption categories. The confidence ranges for the coefficients are shown in Figure 1. The first nine coefficients relate to food and are predominantly negative, confirming food's status as being more important in the budgets of households with low rather than high incomes. Category 20 shows the coefficient on rent payments and category 21 that on owner occupiers' housing costs. Not surprisingly the former are declining in mean log income while the latter are increasing in mean log income. The share of spending on item 73, package holidays, is also increasing sharply with income as are items 75 (restaurants and cafes) and 77 (accommodation away from home).

As we have noted, with regional dummies included, this will model the expenditure pattern of each region exactly and the regional inflation rates generated are therefore the same as those which would be produced simply by producing weights as the means of the expenditure shares in each region. But the fixed effects model using income as a covariate allows us to explore the impact of modelling the expenditure pattern for those MSOAs in each region for which there are no LCFS data. Any material difference

Figure 1: 95% Confidence Ranges for Expenditure Share Coefficients on log Mean Income



between the two might indicate that the LCFS is not representative, at least as far as expenditure shares go. Table 4 shows percentage growth in prices for 2016, measured relative to January 2016, calculated using the model while Table 5 shows the analogous results calculated using only the regional expenditure patterns of 2016. There is very little difference between the two suggesting that the expenditure patterns of the LCFS sample are representative. The tables also shows the standard errors of the estimates calculated using the model and then only the data. As with the figures themselves, these are generally close to each other. In particular, we can see that modelling does not result in a reduction in the standard errors of the estimates, and it follows that, if greater accuracy is required, it will be necessary to increase the sample size. It is also worth noting that the standard errors are in some cases higher when measuring the price

Increase in Price Index relative to January (%)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
North-East	0.00	0.23	0.48	0.66	0.84	1.03	0.96	1.31	1.48	1.54	1.77	2.22	1.63
North-West	0.00	0.19	0.39	0.62	0.81	0.97	0.90	1.17	1.42	1.50	1.72	2.07	1.62
Yorks & Humber	0.00	0.19	0.42	0.65	0.85	1.04	1.02	1.28	1.51	1.57	1.77	2.16	1.72
East Midlands	0.00	0.20	0.42	0.66	0.89	1.06	0.99	1.23	1.53	1.63	1.87	2.21	1.82
West Midlands	0.00	0.20	0.39	0.63	0.84	1.02	0.95	1.19	1.48	1.55	1.80	2.14	1.71
East	0.00	0.18	0.40	0.66	0.88	1.07	1.01	1.24	1.52	1.65	1.91	2.24	1.86
London	0.00	0.28	0.58	0.73	0.92	1.11	1.26	1.68	1.68	1.67	1.77	2.41	1.62
South-East	0.00	0.24	0.54	0.76	0.98	1.18	1.22	1.54	1.67	1.73	1.92	2.42	1.83
South-West	0.00	0.15	0.36	0.64	0.84	1.02	0.98	1.18	1.46	1.58	1.81	2.12	1.83
Wales	0.00	0.22	0.39	0.68	0.90	1.06	0.92	1.15	1.52	1.61	1.91	2.17	1.80

Standard Errors

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
North-East	0.00	0.02	0.05	0.04	0.05	0.06	0.10	0.12	0.07	0.07	0.07	0.11	0.07
North-West	0.00	0.01	0.03	0.02	0.03	0.04	0.06	0.08	0.05	0.05	0.04	0.07	0.05
Yorks & Humber	0.00	0.02	0.04	0.03	0.03	0.04	0.07	0.09	0.05	0.05	0.05	0.07	0.05
E Mid	0.00	0.02	0.04	0.03	0.04	0.05	0.08	0.10	0.06	0.06	0.05	0.08	0.06
W Mid	0.00	0.02	0.04	0.03	0.03	0.04	0.07	0.09	0.05	0.05	0.05	0.07	0.05
East	0.00	0.02	0.04	0.03	0.03	0.04	0.07	0.09	0.05	0.05	0.05	0.07	0.05
London	0.00	0.02	0.04	0.03	0.04	0.05	0.07	0.09	0.06	0.05	0.05	0.08	0.05
South-East	0.00	0.01	0.03	0.02	0.03	0.04	0.06	0.07	0.05	0.04	0.04	0.06	0.04
South-West	0.00	0.01	0.04	0.03	0.03	0.04	0.07	0.09	0.05	0.05	0.05	0.07	0.05
Wales	0.00	0.02	0.05	0.04	0.05	0.06	0.10	0.13	0.08	0.07	0.07	0.11	0.07

Table 4: Regional percentage price increases relative to Jan 2016- expenditure weights generated using a fixed effects model

increase over only a few months rather than over the full twelve months from January to January. Since the uncertainty comes from the expenditure shares, quarterly measures of price change are unlikely to be an improvement on monthly measures.

6.4 A Time-series of Regional Consumer Price Indices

For the production of regional weights, modelling seems to offer relatively little on top of what can be produced by a sub-sample of the data. Models may nevertheless be useful to meet any need for estimates of expenditure patterns for different cuts of the data. If, for example, it is desired to produce estimates of expenditure shares for a geography which is not well represented in the LCFS, then expenditure shares based on the model may be more precise than those based on an appropriate selection of respondents.

Increase in Price Index relative to January (%)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
North-East	0.00	0.22	0.45	0.65	0.82	1.00	0.91	1.24	1.46	1.52	1.75	2.16	1.60
North-West	0.00	0.18	0.38	0.61	0.80	0.96	0.90	1.17	1.41	1.48	1.69	2.05	1.60
Yorks & Humber	0.00	0.19	0.40	0.63	0.83	1.01	0.99	1.25	1.48	1.55	1.76	2.13	1.71
East Midlands	0.00	0.19	0.41	0.66	0.88	1.05	0.90	1.22	1.52	1.64	1.88	2.21	1.85
West Midlands	0.00	0.19	0.37	0.61	0.82	0.99	0.91	1.15	1.44	1.51	1.77	2.09	1.68
East	0.00	0.18	0.39	0.66	0.88	1.06	1.00	1.23	1.50	1.64	1.89	2.22	1.85
London	0.00	0.28	0.58	0.72	0.91	1.09	1.23	1.66	1.67	1.67	1.77	2.41	1.62
South-East	0.00	0.24	0.53	0.75	0.97	1.16	1.19	1.51	1.65	1.71	1.91	2.39	1.82
South-West	0.00	0.15	0.35	0.63	0.83	1.01	0.97	1.17	1.45	1.57	1.80	2.10	1.82
Wales	0.00	0.23	0.40	0.69	0.93	1.09	0.94	1.16	1.54	1.64	1.95	2.22	1.86

Standard Errors

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
North-East	0.00	0.02	0.06	0.04	0.05	0.07	0.10	0.13	0.08	0.07	0.07	0.11	0.07
North-West	0.00	0.02	0.04	0.03	0.03	0.04	0.07	0.08	0.05	0.05	0.04	0.07	0.04
Yorks & Humber	0.00	0.02	0.04	0.03	0.04	0.05	0.07	0.08	0.05	0.05	0.05	0.07	0.05
East Midlands	0.00	0.02	0.04	0.03	0.04	0.05	0.07	0.08	0.05	0.06	0.05	0.07	0.06
West Midlands	0.00	0.02	0.04	0.03	0.03	0.04	0.06	0.07	0.05	0.05	0.05	0.07	0.05
East	0.00	0.01	0.04	0.03	0.03	0.05	0.07	0.08	0.05	0.06	0.06	0.08	0.07
London	0.00	0.02	0.05	0.03	0.04	0.06	0.10	0.12	0.07	0.06	0.05	0.10	0.04
South-East	0.00	0.01	0.04	0.02	0.03	0.04	0.07	0.08	0.05	0.04	0.05	0.07	0.04
South-West	0.00	0.01	0.03	0.02	0.03	0.04	0.06	0.07	0.05	0.05	0.05	0.07	0.05
Wales	0.00	0.03	0.05	0.04	0.05	0.06	0.08	0.07	0.08	0.08	0.08	0.10	0.09

Table 5: Regional percentage price increases relative to Jan 2016- expenditure weights calculated from LCFS data

Tables 6 and 7 show regional price indices calculated using democratic weights. The weights are taken from the scaled LCFS two years before the year to which the data relate. The index is computed by single chain-linking carried out in January.

7 Relationship to Earlier Work for ONS

The results presented here contrast with those found by Dawber et al. (2022). Their study explored the problem of estimating regional weights for plutocratic rather than democratic price indices. As we have noted above, the weights for plutocratic indices are derived from estimates of total spending on each consumption category in each region as shares of total expenditure. They considered each consumption category in isolation and worked with the survey totals for each of the twelve regions. They then used regression methods to explain the regional expenditure totals by means of covariates. With twelve regions they limited themselves to five covariates so as to preserve more than a minimal number of degrees of freedom. The resulting regression models allowed them to produce fitted values for total expenditure in each region. It was also possible to calculate the variances of both estimates of total expenditure, those arising from the survey data and those generated by the regression models. Making the assumption that the two estimates for each consumption category and region were independent of each other, it was then possible to produce weighted means of the two estimates, with the weights being the inverses of the respective variances as advocated by Fay & Herriot (1979). The weighted means have variances lower than either the survey or the fitted estimates; as we have discussed, the procedure "borrows strength" from the regression models to enhance the accuracy of the estimates.

Our focus has been on the variance of estimates of the democratic inflation rate for each region. We have found that, even when working to the COICOP87 level the resulting standard deviations from uncertainty in the mean expenditure shares were not large- and that this was also true for plutocratic price indices. We have limited our investigation of variances to 2016 because that is the year for which we have income covariates. But Dawber et al. (2022) draw attention to considerable instability in the regional weights from one year to the next and suggest that this is responsible for erratic movements in regional indices. Thus they point out that, for the North-East a zero entry is shown for electricity (category 4.5.1- Item 26) in 2018 after 2-4% expenditure weights in earlier years. The data set that they used also had no entries for owner occupiers' housing costs (4.2- Item 21); the version of LCFS scaled to be coherent with

	North East	Northt West	Yorks	E Mid	W Mid	East	London	South East	South West	Wales
Jan2012	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Feb2012	100.6	100.5	100.5	100.5	100.5	100.5	100.5	100.5	100.5	100.5
Mar2012	100.8	100.7	100.8	100.7	100.7	100.7	100.7	100.7	100.7	100.8
Apr2012	101.4	101.3	101.3	101.2	101.3	101.3	101.3	101.3	101.3	101.4
May2012	101.3	101.3	101.3	101.2	101.2	101.2	101.3	101.2	101.2	101.2
Jun2012	100.9	100.9	100.9	100.9	100.9	100.8	101.1	100.9	100.8	100.8
Jul2012	100.8	100.9	100.8	100.9	100.8	100.8	101.2	100.9	100.9	100.7
Aug2012	101.2	101.2	101.1	101.2	101.2	101.1	101.6	101.2	101.3	101.1
Sep2012	101.7	101.8	101.7	101.7	101.7	101.7	101.8	101.7	101.6	101.7
Oct2012	102.0	102.1	102.0	102.1	101.9	101.9	102.3	101.9	101.9	101.9
Nov2012	102.2	102.3	102.3	102.4	102.2	102.2	102.6	102.2	102.1	102.1
Dec2012	102.7	102.8	102.7	102.9	102.7	102.7	103.2	102.7	102.6	102.6
Jan2013	102.4	102.5	102.4	102.6	102.4	102.4	102.7	102.4	102.3	102.4
Feb2013	103.1	103.1	103.0	103.2	103.0	103.0	103.4	103.0	102.9	103.0
Mar2013	103.5	103.5	103.4	103.5	103.4	103.4	103.7	103.3	103.3	103.4
Apr2013	103.8	103.8	103.7	103.8	103.6	103.6	104.1	103.6	103.6	103.6
May2013	104.0	104.0	103.8	103.9	103.9	103.8	104.4	103.7	103.7	103.8
Jun2013	103.8	103.8	103.7	103.8	103.7	103.6	104.2	103.6	103.6	103.5
Jul2013	103.8	103.8	103.7	103.8	103.7	103.6	104.4	103.6	103.6	103.5
Aug2013	104.2	104.3	104.1	104.2	104.2	104.1	104.8	104.0	104.0	103.9
Sep2013	104.6	104.5	104.4	104.4	104.4	104.3	104.9	104.3	104.3	104.2
Oct2013	104.5	104.5	104.3	104.4	104.4	104.3	104.9	104.3	104.2	104.2
Nov2013	104.6	104.6	104.4	104.5	104.4	104.3	104.9	104.3	104.3	104.3
Dec2013	105.3	105.2	105.0	105.1	105.1	104.9	105.6	104.8	104.8	104.9
Jan2014	104.6	104.6	104.5	104.6	104.5	104.4	104.9	104.3	104.3	104.3
Feb2014	105.0	105.1	104.9	105.1	104.9	104.8	105.3	104.8	104.7	104.7
Mar2014	105.2	105.3	105.1	105.3	105.1	105.0	105.5	105.0	104.9	104.9
Apr2014	105.4	105.6	105.4	105.5	105.4	105.3	106.1	105.3	105.2	105.1
May2014	105.3	105.5	105.3	105.4	105.3	105.2	105.9	105.2	105.1	105.0
Jun2014	105.5	105.6	105.5	105.6	105.4	105.3	106.1	105.4	105.2	105.1
Jul2014	105.2	105.4	105.1	105.3	105.1	105.1	106.0	105.1	104.9	104.7
Aug2014	105.5	105.7	105.4	105.6	105.4	105.3	106.4	105.5	105.2	104.9
Sep2014	105.5	105.8	105.6	105.6	105.5	105.4	106.3	105.4	105.3	105.1
Oct2014	105.6	105.8	105.6	105.6	105.6	105.4	106.3	105.4	105.3	105.2
Nov2014	105.3	105.6	105.4	105.4	105.4	105.2	106.1	105.2	105.1	105.0
Dec2014	105.4	105.6	105.4	105.4	105.4	105.2	106.3	105.3	105.1	104.8
Jan2015	104.5	104.8	104.5	104.5	104.5	104.3	105.5	104.4	104.3	103.9
Feb2015	104.7	105.0	104.8	104.7	104.7	104.6	105.8	104.6	104.5	104.1
Mar2015	104.8	105.1	104.9	104.8	104.9	104.7	105.9	104.8	104.7	104.2
Apr2015	105.0	105.3	105.1	105.0	105.1	105.0	106.2	105.1	104.9	104.5
May2015	105.1	105.5	105.2	105.2	105.2	105.1	106.4	105.2	105.1	104.7
Jun2015	105.2	105.5	105.3	105.2	105.2	105.2	106.4	105.3	105.1	104.7
Jul2015	104.9	105.3	105.1	105.0	105.0	105.1	106.6	105.3	105.0	104.5
Aug2015	105.2	105.5	105.3	105.2	105.2	105.3	107.0	105.5	105.2	104.7
Sep2015	105.1	105.5	105.2	105.1	105.1	105.2	106.6	105.3	105.1	104.5
Oct2015	105.2	105.6	105.3	105.2	105.2	105.2	106.6	105.3	105.2	104.6
Nov2015	105.3	105.6	105.4	105.2	105.2	105.2	106.6	105.3	105.2	104.7
Dec2015	105.3	105.6	105.4	105.2	105.2	105.3	107.1	105.5	105.2	104.7

Table 6: Regional Price Indices: Jan 2012=100

	North East	North West	Yorks	E Mid	W Mid	East	London	South East	South West	Wales
Jan2016	104.6	105.0	104.7	104.6	104.6	104.6	106.0	104.7	104.5	104.0
Feb2016	104.8	105.2	105.0	104.8	104.8	104.8	106.3	104.9	104.7	104.2
Mar2016	105.0	105.4	105.2	105.0	104.9	105.0	106.6	105.2	105.0	104.4
Apr2016	105.3	105.6	105.4	105.3	105.3	105.3	106.8	105.4	105.2	104.7
May2016	105.4	105.8	105.6	105.5	105.5	105.5	107.0	105.7	105.4	105.0
Jun2016	105.6	106.0	105.8	105.7	105.7	105.7	107.2	105.9	105.6	105.2
Jul2016	105.5	105.9	105.8	105.6	105.5	105.7	107.3	106.0	105.6	105.0
Aug2016	105.8	106.2	106.1	105.8	105.8	105.9	107.7	106.3	105.8	105.3
Sep2016	106.1	106.5	106.3	106.1	106.1	106.2	107.8	106.4	106.1	105.7
Oct2016	106.2	106.6	106.4	106.2	106.3	106.3	107.8	106.5	106.2	105.9
Nov2016	106.4	106.8	106.6	106.5	106.6	106.5	108.0	106.6	106.4	106.2
Dec2016	106.7	107.1	107.0	106.8	106.8	106.9	108.6	107.1	106.8	106.5
Jan2017	106.3	106.7	106.5	106.5	106.5	106.5	107.9	106.6	106.4	106.2
Feb2017	107.1	107.4	107.2	107.2	107.3	107.2	108.7	107.3	107.1	106.9
Mar2017	107.5	107.8	107.7	107.6	107.7	107.6	109.0	107.7	107.4	107.3
Apr2017	108.2	108.5	108.2	108.2	108.3	108.2	109.9	108.4	108.1	107.9
May2017	108.6	108.9	108.7	108.6	108.7	108.6	110.1	108.7	108.5	108.3
Jun2017	108.6	108.9	108.7	108.6	108.7	108.6	110.3	108.8	108.6	108.3
Jul2017	108.6	108.9	108.6	108.6	108.6	108.6	110.4	108.9	108.7	108.2
Aug2017	109.2	109.5	109.2	109.1	109.2	109.1	111.1	109.5	109.2	108.7
Sep2017	109.5	109.8	109.6	109.5	109.6	109.5	111.1	109.7	109.4	109.2
Oct2017	109.7	109.9	109.7	109.6	109.7	109.6	111.1	109.7	109.5	109.4
Nov2017	110.0	110.3	110.1	110.0	110.1	110.0	111.3	110.0	109.8	109.8
Dec2017	110.3	110.6	110.4	110.3	110.4	110.3	112.0	110.5	110.3	110.0
Jan2018	109.9	110.1	109.9	109.8	109.9	109.8	111.1	109.8	109.7	109.6
Feb2018	110.3	110.5	110.3	110.2	110.3	110.2	111.5	110.3	110.1	110.1
Mar2018	110.5	110.6	110.4	110.3	110.4	110.3	111.5	110.4	110.2	110.2
Apr2018	111.0	111.2	111.0	110.9	111.0	110.9	112.2	111.0	110.8	110.9
May2018	111.4	111.6	111.4	111.3	111.3	111.3	112.6	111.4	111.1	111.2
Jun2018	111.4	111.7	111.5	111.3	111.4	111.3	112.7	111.5	111.2	111.3
Jul2018	111.5	111.7	111.5	111.3	111.4	111.3	112.9	111.6	111.2	111.2
Aug2018	112.3	112.4	112.2	112.0	112.1	112.0	113.8	112.3	111.8	111.8
Sep2018	112.3	112.5	112.3	112.2	112.3	112.2	113.4	112.2	112.0	112.1
Oct2018	112.4	112.6	112.4	112.3	112.4	112.3	113.4	112.3	112.1	112.2
Nov2018	112.6	112.8	112.6	112.5	112.5	112.4	113.6	112.4	112.3	112.4
Dec2018	112.8	113.0	112.8	112.6	112.7	112.6	114.0	112.7	112.4	112.5
Jan2019	111.8	112.1	111.9	111.7	111.7	111.7	113.0	111.9	111.6	111.5
Feb2019	112.3	112.6	112.4	112.2	112.1	112.2	113.5	112.4	112.1	112.0
Mar2019	112.5	112.8	112.6	112.4	112.4	112.4	113.6	112.6	112.3	112.2
Apr2019	113.6	114.1	113.7	113.5	113.5	113.5	115.0	113.7	113.3	113.4
May2019	113.9	114.4	114.1	113.9	113.9	113.9	115.1	114.0	113.7	113.7
Jun2019	113.9	114.4	114.1	113.9	113.9	113.8	115.2	114.0	113.7	113.7
Jul2019	113.9	114.4	114.1	113.8	113.9	113.8	115.2	114.1	113.8	113.7
Aug2019	114.3	114.9	114.5	114.3	114.3	114.2	115.8	114.5	114.2	114.1
Sep2019	114.4	114.9	114.6	114.4	114.4	114.3	115.7	114.5	114.2	114.2
Oct2019	114.0	114.5	114.3	114.0	114.0	114.0	115.3	114.2	113.9	113.8
Nov2019	114.3	114.8	114.5	114.3	114.3	114.3	115.5	114.5	114.2	114.1
Dec2019	114.4	114.9	114.6	114.3	114.4	114.3	115.7	114.6	114.2	114.2
Jan2020	114.2	114.5	114.3	114.0	114.2	114.1	115.3	114.3	114.0	113.9

Table 7: Regional Price Indices: Jan 2012=100

the macroeconomic consumption data does not suffer from this defect.

We also find, looking at a time series of regional price indices, produced using national COICOP87 price indices, that there are few erratic movements. Our indices are produced by single chain linking, in January, rather than double-chain linking in January and February since we have not been able to replicate the second approach. But it is hard to believe that double chain linking has a substantial impact on volatility. An anonymous referee has reported that,

"the instability identified by Dawber et al. (2022) arose came from some “inflated” adjusted weights of some COICOP categories (e.g., for North East, “Electricity”, “Telephone and telefax equipment and services”, “Major durables for in and outdoor recreation”, and “Education”) – if those COICOPs were eliminated, the regional indices where more stable and comparable to the regional index computed with national weights.”

We seem to have avoided these problems by using the version of the LCFS adjusted by ONS to be consistent with the national accounts.

8 Conclusions

We have explored the role of small area methods in the estimation of weights for consumer price indices. A key restriction is that the sum of estimated weights should equal one. This is shown to impose a strong restriction on models estimated with random effects, and we find that, in the particular case of interest to us, the restrictions are strongly rejected even when tested for a model estimated over twelve broad COICOP categories.

A fixed effects model offers an alternative approach. In this case estimation of equations explaining the weights results in coefficients which add to one, provided only that the same explanatory variables are used in all the expenditure share equations. If dummy variables are used to represent regional or other characteristics, then the estimated shares for the characteristics in question will exactly match the data. As a result it is not possible to produce mixed estimates as variance-weighted means of the fitted values and data. With suitable explanatory variables, the fitted values may nevertheless be more precise than those calculated from the data alone. The models are also likely to be helpful when estimating expenditure patterns for categories which cut across those associated with the dummy variables.

In our particular context model-based and data based estimates are generally very similar. Nor is it generally the case that the model-based estimates are more precise than the data-based estimates.

This study complements other work for ONS in a number of ways. First of all, producing indices using regional weights applied to national price indices for 87 COICOP categories, does not result in material instability problems. That suggests that the instability identified by Dawber et al. (2022) is probably generated by instability in regional prices rather than by erratic movements in weights. If the problems with these are addressed, perhaps by raising the sample size, it should then be reasonably straightforward to produce regional price indices. If there are any concerns about instability, arising from local price collection issues, these can probably be addressed by publishing them on a quarterly rather than monthly basis. On the other hand quarterly aggregation will not reduce errors arising from uncertainty in expenditure shares.

References

Aitken, A. & Weale, M. (2020), 'A democratic measure of household income growth: Theory and application to the united kingdom', *Economica* **87**, 589–610.

Barten, A. (1969), 'Maximum-likelihood Estimation of a Complete System of Demand Equations', *European Economic Review* **1**, 7–73.

Dawber, J., Würz, N., Smith, P. A., Flower, T., Thomas, H., Schmid, T. & Tzavidis, N. (2022), 'Experimental uk regional consumer price inflation with model-based expenditure weights', *Journal of Official Statistics* **38**, 213–237.

Deaton, A. & Muellbauer, J. (1976), 'An almost ideal demand system', *American Economic Review* **70**, 312–326.

Fay, R. & Herriot, R. (1979), 'Estimates of income for small places: an application of james-stein procedures to census data', *Journal of the American Statistical Association* **74**, 269–277.

Harville, D. (1977), 'Maximum likelihood approaches to variance component estimation and to related problems', *Journal of the American Statistical Association* **72**, 320–338.

Office for National Statistics (2021), Consumer price inflation time series: April 2021. https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceindices/current.

Prais, S. (1958), 'Whose Cost of Living?', *Review of Economic Studies* **26**, 126–134.

Stuart, A. & Ord, J. (1986), *Kendall's Advanced Theory of Statistics: Vol I*, Charles Griffin and Co. Ltd., London.

A Models with Regional Random and Fixed Effects estimated for COICOP12

	Food	Drink	Clothing	Housing	Furnishings	Health
Mean(log hhld inc)	-0.088***	-0.010	-0.013	-0.080*	0.008	0.015***
	(-5.16)	(-1.28)	(-1.34)	(-2.39)	(0.67)	(3.39)
log σ_u	-5.322***	-5.227***	-5.648***	-3.775***	-5.750***	-5.413***
	(-24.04)	(-23.19)	(-9.90)	(-9.61)	(-7.01)	(-19.15)
log σ_e	-2.508***	-2.647***	-2.545***	-1.983***	-2.524***	-2.808***
	(-86.19)	(-76.82)	(-82.33)	(-44.26)	(-68.81)	(-30.18)
	Transport	Comms	Recreation	Education	Hotels etc.	Misc.
Mean (log hhld inc)	0.050*	-0.017***	0.074***	-0.006	0.047*	0.014
	(2.38)	(-4.39)	(6.67)	(-0.61)	(2.26)	(1.08)
log σ_u	-5.007***	-6.865***	-4.396***	-6.169***	-5.686***	-5.698***
	(-11.31)	(-13.57)	(-12.73)	(-11.99)	(-8.50)	(-7.54)
log σ_e	-2.111***	-3.655***	-2.116***	-3.048***	-2.349***	-2.557***
	(-115.47)	(-73.36)	(-62.79)	(-44.18)	(-127.08)	(-98.22)
4151 observations						

t statistics in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 8: The Coefficient on Income in the Random Effects Model

	Food	Drink	Clothing	Housing	Furnishings	Health
Mean(log hhld inc)	-0.089***	-0.007	-0.015	-0.083***	0.010	0.015*
	(-9.46)	(-0.86)	(-1.64)	(-5.27)	(1.09)	(2.19)
log σ_e	-2.513	-2.645	-2.538	-1.981	-2.526	-2.813
	Transport	Comms	Recreation	Education	Hotels etc.	Misc.
Mean(log hhld inc)	0.051***	-0.018***	0.077***	-0.008	0.049***	0.018*
	(3.69)	(-6.12)	(5.57)	(-1.53)	(4.46)	(1.97)
log σ_e	-2.112	-3.650	-2.112	-3.058	-2.353	-2.551
4151 observations						

t statistics in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 9: The Coefficient on Income in the Fixed Effects Model