Inference for stable isotope mixing models: a study of the diet of dunlin

Erik Barry Erhardt and Edward J. Bedrick

University of New Mexico, Albuquerque, USA

Summary. Stable isotope sourcing is used to estimate proportional contributions of sources to a mixture, such as in the analysis of animal diets and plant nutrient use. Statistical methods for inference on the diet proportions by using stable isotopes have focused on the linear mixing model. Existing frequentist methods assume that the diet proportion vector can be uniquely solved for in terms of one or two isotope ratios. We develop large sample methods that apply to an arbitrary number of isotope ratios, assuming that the linear mixing model has a unique solution or is overconstrained. We generalize these methods to allow temporal modelling of the population mean diet, assuming that isotope ratio response data are collected over time. The methodology is motivated by a study of the diet of dunlin, a small migratory seabird.

Keywords: Animal ecology; Estimating equations; Least squares; Resource utilization

1. Introduction

Stable isotope analysis of a consumer animal's tissues (the mixture) and their potential prey and diet (the sources) is a powerful means of quantifying relative contributions of isotopically distinct dietary components, providing many benefits in comparison with traditional methods for quantifying diet, namely analysis of stomach and faecal contents (Hobson and Wassenaar, 1999). Some background information will help in understanding this approach. Elements can exist in both stable and unstable or radioactive forms. Each form of an element, or isotope, has the same number of protons but different numbers of neutrons. Elements of biological interest such as hydrogen, carbon, nitrogen, oxygen and sulphur have two or more stable isotopes with the lightest of these present in much greater abundance than the others. For example, carbon and nitrogen each have one heavy stable isotope (^{13}C and ^{15}N) with a natural abundance of about 1% or less and one light stable isotope (^{12}C and ^{14}N) that makes up the remainder. Carbon also has a radioactive isotope: ^{14}C. The isotope ratio (IR) of an element, $\delta = 1000(R_{sample}/R_{standard} - 1)$‰, is a normalized ratio of the number of rarer to the most common stable isotope in a sample, R_{sample}, relative to an international standard, $R_{standard}$, given in parts per thousand (Kendall and McDonnell, 1998). A mass spectrometer is typically used to measure IRs from tissue or blood samples after vaporization and ionization.

Stable isotopes provide quantitative information on diet because animal tissues carry an isotopic imprint of the processes and resources used for their synthesis (DeNiro and Epstein, 1978). Carbon and nitrogen IRs have found widespread use in studies of animal diets (Rundel *et al.*, 1989). Carbon, $\delta^{13}C$, varies among plant and algae primary producers with different

Address for correspondence: Edward J. Bedrick, Division of Epidemiology and Biostatistics, Department of Internal Medicine, University of New Mexico Health Sciences Center, MS C10 5550 Albuquerque, NM 87131, USA.
E-mail: EBedrick@salud.unm.edu

photosynthetic pathways, but changes little with trophic transfer, and so is useful for identifying the sources of dietary carbon (DeNiro and Epstein, 1981). Nitrogen, δ^{15}N, increases stepwise with trophic transfer and is thus used to estimate position in a food web (Minagawa and Wada, 1984). Similarly to carbon, sulphur, δ^{34}S, varies among primary producers but changes little with trophic transfer; sulphur is especially informative in marine systems where it exists in higher abundances (Currin *et al.*, 1995). Additional information can be found in Newsome *et al.* (2007).

Inference on diet by using stable isotopes has focused on the simplest mass balance model, which is called the basic mixing model (BMM). The BMM states that the population mean IR in the mixture or consumer, β, is a convex combination of the mean IRs δ_s in the source populations after correcting for discrimination, Δ_s. The weights in the convex combination are the proportional contributions of the sources to the consumer's diet. Assuming that diet consists of S sources and that I isotopes are measured,

$$\beta = \sum_{s=1}^{S} \pi_s \delta'_s = A\pi,$$

where $\pi = (\pi_1, \ldots, \pi_S)^{\mathrm{T}}$ is the vector of diet proportions. Matrix A has columns $\delta'_s = \delta_s + \Delta_s$ for source $s = 1, \ldots, S$, which are the discrimination-corrected source IRs or the source IR values as found in the tissues of the consumer. The source-specific discrimination terms account for the consumer's ingestion, metabolization and excretion of their diet, which altogether are also called 'trophic fractionation' (Minagawa and Wada, 1984). Letting $\mathbf{1}_S$ be an $S \times 1$ vector of 1s, re-express A and β in the BMM to include the constraint $\mathbf{1}_S^{\mathrm{T}}\pi = 1$ in the equivalent form $\beta_* = A_*\pi$, where $A_*^{\mathrm{T}} = (\mathbf{1}_S \ A^{\mathrm{T}})$ and $\beta_*^{\mathrm{T}} = (1 \ \beta^{\mathrm{T}})$.

The diet proportions $\pi_1, \ldots, \pi_S$ represent the proportion of ingested dry weight biomass from the S sources on assuming that the concentration of each element and the percentage of an element that is assimilated into the tissue (the assimilation efficiency) are identical across sources. Elemental concentrations and digestive efficiencies of consumers can vary widely across food sources, suggesting that the more complex extended mixing model (Koch and Phillips, 2002) may be more appropriate in certain settings. We focus on the BMM.

The goal is to estimate π accounting for estimation of β and A. A frequentist approach requires that π is uniquely determined from β and A. In the absence of plausible constraints on π, this typically requires that the number of isotopes $I \geqslant S - 1$, the number of free parameters in π. For example, $\pi = A_*^{-1}\beta_*$ when $I + 1 = S$ and A_* is invertible. In contrast, a Bayesian approach applies to arbitrary I and S but the posterior distribution of π is strongly influenced by the prior for π when the BMM is underconstrained ($I + 1 < S$). Moore and Semmens (2008), Parnell and Jackson (2008), Parnell *et al.* (2010), Ward *et al.* (2010) and Erhardt and Bedrick (2012) considered Bayesian estimation in the BMM.

We consider a frequentist approach based on consistent estimators of β and A, say $\bar{\mathbf{b}}$ and $\hat{A}$. We propose two weighted least squares (WLS) methods assuming $I + 1 \geqslant S$ and identifiability, i.e. at least S of the I isotope constraints are non-collinear. The first method applies non-linear WLS in the 'working model' $\bar{\mathbf{b}} = \hat{A}\,\pi(\alpha) + \mathbf{e}$, where π is parameterized in terms of an $(S-1) \times 1$ vector α. The second approach considers linear WLS in the working model $\bar{\mathbf{b}}_* = \hat{A}_*\pi + \mathbf{e}_*$, where $\bar{\mathbf{b}}_*^{\mathrm{T}} = (1 \ \bar{\mathbf{b}}^{\mathrm{T}})$ and $\hat{A}_*^{\mathrm{T}} = (\mathbf{1}_S \ \hat{A}^{\mathrm{T}})$ are the augmented data summaries. Linear WLS incorporates the sum constraint on π but ignores the non-negativity constraint, thus simplifying the computation. Linear WLS is 'direct' when $I + 1 = S$ because the 'empirical BMM' $\bar{\mathbf{b}}_* = \hat{A}_*\pi$ has a unique solution $\hat{A}_*^{-1}\bar{\mathbf{b}}_*$ provided that $\hat{A}_*$ is invertible. For simplicity, we refer to a BMM with $I + 1 = S$ as having a unique solution whereas the BMM with $I + 1 > S$ is overconstrained.

Existing frequentist methods are limited to a BMM with a unique solution and $I \leqslant 2$ (Phillips and Gregg, 2001). These methods do not apply in our motivating example, where the model is

overconstrained. Sections 2 and 3 provide a general asymptotic framework for WLS estimation in models that are overconstrained or have a unique solution. Section 3 also considers small sample properties of the WLS estimators and the importance of proper weighting in the least squares procedures. Much of the discussion is motivated by a study of the diet of dunlin, which is a small migratory seabird. Section 4 considers a more complex BMM, where the temporal sampling of the dunlin is taken into consideration, allowing for the estimation of the population diet as a function of time. Section 5 gives concluding remarks. Technical results are presented in Appendix A.

2. Preliminaries

2.1. Data

We assume that an independent and identically distributed sample $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_J$ of consumers is selected with $E(\mathbf{b}_j) = \beta$ and $\mathrm{cov}(\mathbf{b}_j) = \Sigma_b$. Similarly, independent and identically distributed samples $\{\mathbf{d}_{sj}, \text{ for } j = 1, \ldots, K_s\}$ for $s = 1, \ldots, S$, with $E(\mathbf{d}_{sk}) = \delta_s$ and $\mathrm{cov}(\mathbf{d}_{sk}) = \Sigma_s$, are available to estimate the source mean IRs δ_s. Let $\bar{\mathbf{b}}$ and $\bar{\mathbf{d}}_s$ be the $I \times 1$ sample mean IR vectors for the consumer and for the sth source respectively.

To be consistent with our motivating example, we assume that discrimination is estimated from single-source diet experiments. In a single-source diet experiment, consumers are fed for an extended period of time a controlled diet that is representative of a source. Discrimination is the mean difference between the IRs in the source of diet and in tissues of the consumer at diet equilibrium when there is no residual effect of previous diet on the IRs (Caut *et al.*, 2009). Assuming that a diet experiment is performed for the sth source, we observe independent and identically distributed IR samples for the proxy diet source $\{\mathbf{d}_{\mathrm{D}sj}, \text{ for } j = 1, \ldots, K_{\mathrm{D}s}\}$ and for the consumer tissue $\{\mathbf{d}_{\mathrm{T}sj}, \text{ for } j = 1, \ldots, K_{\mathrm{T}s}\}$. Here $E(\mathbf{d}_{\mathrm{D}sj}) = \delta_{\mathrm{D}s}$ and $\mathrm{cov}(\mathbf{d}_{\mathrm{D}sj}) = \Sigma_{\mathrm{D}s}$, whereas $E(\mathbf{d}_{\mathrm{T}sj}) = \delta_{\mathrm{T}s}$ and $\mathrm{cov}(\mathbf{d}_{\mathrm{T}sj}) = \Sigma_{\mathrm{T}s}$. The discrimination $\boldsymbol{\Delta}_s = \delta_{\mathrm{T}s} - \delta_{\mathrm{D}s}$ is estimated by $\hat{\boldsymbol{\Delta}}_s = \bar{\mathbf{d}}_{\mathrm{T}s} - \bar{\mathbf{d}}_{\mathrm{D}s}$, which is the difference in sample means. Another strategy to estimate discrimination which is not considered here uses a regression over a range of sources (Felicetti *et al.*, 2003).

Given these summaries, β and A are estimated with $\bar{\mathbf{b}}$ and $\hat{A}$, where the sth column of $\hat{A}$ is $\bar{\mathbf{d}}_s + \hat{\boldsymbol{\Delta}}_s$. Discrimination is often assumed to be identical for similar sources and, if so, then one single-source diet experiment suffices to estimate their common discrimination. In general, the number of diet experiments is $S_0 \leqslant S$, so there are S_0 estimated discrimination values $\hat{\boldsymbol{\Delta}}_s$ in $\hat{A}$.

2.2. Motivating example: dunlin diet

We use data from Evans Ogden *et al.* (2005) to quantify the proportional use that 174 *Calidris alpina pacifica* (dunlin: a small migratory seabird) made of farmland and marine resources on the Fraser River Delta, British Columbia, Canada, from January to April 2000. The two sources ($S = 2$) in the dunlin diet represent protein from invertebrates that feed on plants with distinct photosynthetic pathways and IR ranges (Hobson, 1999). The terrestrial source represents C3 plants, which are so called because the first organic carbon compound made in photosynthesis contains three carbon atoms, whereas the marine source represents C4 plants.

The data set contains the carbon and nitrogen IRs (i.e. $\delta^{13}\mathrm{C}$ and $\delta^{15}\mathrm{N}$, so $I = 2$) for the 174 dunlin and for samples of 20 marine and 16 terrestrial invertebrates from the two sources. To estimate discrimination, Evans Ogden *et al.* (2004) conducted a diet experiment with foods of terrestrial C3 origin on the blood of four control dunlin. Ideally, a diet experiment would be performed for all isotopes simultaneously for a sample of each source paired with a sample of consumers, resulting in different discrimination estimates for each source. However, because

Table 1.   Dunlin data summary statistics†

	Results for carbon		Results for nitrogen		Correlation
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	
Mixture					
Dunlin ($J = 174$)	−16.40	3.43	11.85	1.34	0.67
Sources					
Terrestrial ($s = 1$; $K_1 = 16$)	−25.36	1.31	6.05	1.26	0.45
Marine ($s = 2$; $K_2 = 20$)	−13.60	2.82	11.09	1.85	−0.35
Diet experiment					
Dunlin tissue ($K_{T_s} = 4$)	−23.28	0.32	6.50	0.11	0.45
Diet ($K_{D_s} = 29$)	−24.66	0.41	3.50	0.40	0.34
Discrimination (dunlin − diet)	1.38		3.00		

†174 observations of dunlin blood as a mixture of $S = 2$ sources using $I = 2$ isotopes of carbon and nitrogen. Summaries are sample sizes, means, standard deviations, and correlations for mixture, sources and diet experiment.

of great time, cost and care of conducting such a set of experiments, this is rarely performed. We follow their recommendation and assume that discrimination is the same for terrestrial and marine sources, and we use the difference between the sample mean IRs in the control dunlin and 29 diet samples, $\hat{\Delta}_1 = \bar{\mathbf{d}}_{T1} - \bar{\mathbf{d}}_{D1}$, as the estimated discrimination. The data are summarized in Table 1.

Fig. 1 is a plot of carbon and nitrogen IRs for the dunlin and source samples. The source data plotted are the discrimination-corrected IR pairs $(\delta^{13}C, \delta^{15}N)^T_{sk} + \hat{\Delta}_1$. The sample means are included in the plot. Thus, the plotted dunlin mean is $\bar{b}$ and the discrimination corrected source means are the two columns of $\hat{A}$. If the BMM held and the means were estimated without error, then the dunlin mean would lie on the line segment joining the two discrimination-corrected source means, thus determining the population mean proportion vector $\boldsymbol{\pi} = (\pi_1, \pi_2)^T$.

3. Estimation methods

3.1. Non-linear weighted least squares estimation

To obtain some generality, assume that $A = A(\boldsymbol{\theta})$ is a differentiable function of a vector $\boldsymbol{\theta}$ of population means and set $\hat{A} = A(\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the corresponding vector of sample means. Let Q be a fixed $I \times I$ positive definite matrix. Consider estimating $\boldsymbol{\pi}$ by minimizing $\mathbf{F}(\boldsymbol{\alpha}) = (\bar{\mathbf{b}} - \hat{A}\boldsymbol{\pi}(\boldsymbol{\alpha}))^T Q (\bar{\mathbf{b}} - \hat{A}\boldsymbol{\pi}(\boldsymbol{\alpha}))$, where $\pi_s(\boldsymbol{\alpha}) = \exp(\alpha_s)/\Sigma^S_{l=1}\exp(\alpha_l)$ for $s = 1, \ldots, S$. Here $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{S-1})^T$ and $\alpha_S = 0$. Let $\tilde{\boldsymbol{\pi}} = \boldsymbol{\pi}(\tilde{\boldsymbol{\alpha}})$, where $\tilde{\boldsymbol{\alpha}}$ minimizes $\mathbf{F}(\boldsymbol{\alpha})$ and is typically obtained by solving the non-linear WLS estimating equations $\dot{\mathbf{F}}(\boldsymbol{\alpha}) = \partial \mathbf{F}/\partial \boldsymbol{\alpha} = -2(\bar{\mathbf{b}} - \hat{A}\boldsymbol{\pi}(\boldsymbol{\alpha}))^T Q \hat{A}\dot{\boldsymbol{\pi}} = \mathbf{0}^T_{S-1}$, where $\dot{\boldsymbol{\pi}}_{S \times (S-1)} = \partial \boldsymbol{\pi}/\partial \boldsymbol{\alpha} = (D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T)M$ with $D(\boldsymbol{\pi}) = \text{diag}(\pi_1, \ldots, \pi_S)$ and $M^T = (I_{S-1}\ \mathbf{0}_{S-1})$. Here I_{S-1} is the rank $S - 1$ identity matrix and $\mathbf{0}_{S-1}$ is an $(S - 1) \times 1$ vector of 0s.

We assume that $\sqrt{J}(\bar{\mathbf{b}} - \boldsymbol{\beta}, \hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ has a limiting normal distribution with mean 0, which implies by the delta method that $\sqrt{J}(\bar{\mathbf{b}} - \hat{A}\boldsymbol{\pi})$ has a limiting $N(\mathbf{0}_I, \Sigma)$ distribution for some Σ. At a minimum, this requires that J and the sample sizes that are used to estimate $\boldsymbol{\theta}$ increase at the same rate. We show in Appendix A that $\sqrt{J}(\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}) \sim N(\mathbf{0}_S, W_Q)$ asymptotically, where

$$W_Q = \dot{\boldsymbol{\pi}}(\dot{\boldsymbol{\pi}}^T A^T Q A \dot{\boldsymbol{\pi}})^{-1}\dot{\boldsymbol{\pi}}^T A^T Q \Sigma Q A \dot{\boldsymbol{\pi}}(\dot{\boldsymbol{\pi}}^T A^T Q A \dot{\boldsymbol{\pi}})^{-1}\dot{\boldsymbol{\pi}}^T.$$
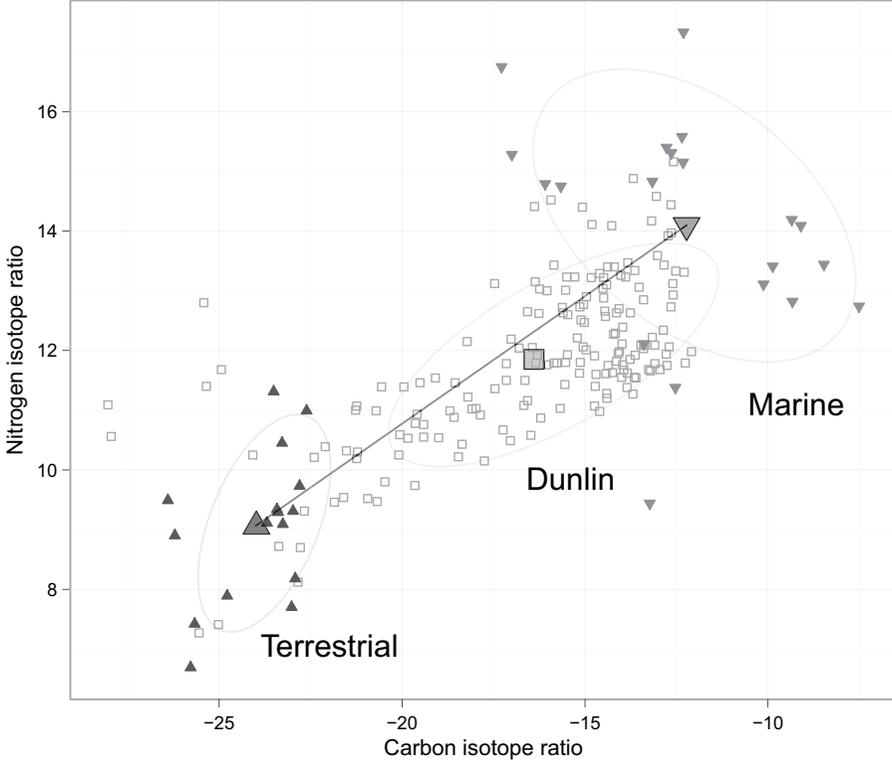
Fig. 1. Dunlin (□) mixture and terrestrial (▲) and marine (▼) discrimination-corrected source δ^{13}C and δ^{15}N observations with means and bivariate normal 75% data ellipses for reference: the line segment connects the discrimination-corrected source means

Although we assumed for simplicity that $\hat{\theta}$ was a vector of averages, the limiting result applies to more general consistent estimators.

We note that W_Q does not depend on the logistic parameterization of π as $\dot{\pi}$ can be replaced by any full rank $S \times (S-1)$ matrix Γ whose columns are orthogonal to $\mathbf{1}_S$. Furthermore, Σ has a simple form for the dunlin data because $\bar{\mathbf{b}} - \hat{A}\pi$ is a linear combination of independent sample means:

$$\text{cov}(\bar{\mathbf{b}} - \hat{A}\pi) = \text{cov}\{\bar{\mathbf{b}} - \pi_1\bar{\mathbf{d}}_1 - \pi_2\bar{\mathbf{d}}_2 - (\bar{\mathbf{d}}_{\text{T1}} - \bar{\mathbf{d}}_{\text{D1}})\}$$
$$= \frac{\Sigma_b}{J} + \frac{\pi_1^2\Sigma_1}{K_1} + \frac{\pi_2^2\Sigma_2}{K_2} + \frac{\Sigma_{\text{T1}}}{K_{\text{T1}}} + \frac{\Sigma_{\text{D1}}}{K_{\text{D1}}}. \tag{1}$$

3.2. Linear estimation procedure

A simpler but asymptotically equivalent estimator is obtained by ignoring the non-negativity of the probabilities and applying linear WLS in the working model $\bar{\mathbf{b}}_* = \hat{A}_*\pi + \mathbf{e}_*$. Kalbfleisch and Lawless (1984) used essentially this approach for estimating Markov chain transition probabilities. To implement this idea, we minimize $(\bar{\mathbf{b}}_* - \hat{A}_*\pi)^{\text{T}} Q_*(\bar{\mathbf{b}}_* - \hat{A}_*\pi)$ subject to $\mathbf{1}_S^{\text{T}}\pi = 1$, where the first rows of each $\bar{\mathbf{b}}_*$ and $\hat{A}_*$ are all 1s and

$$Q_* = \begin{pmatrix} 1 & \mathbf{0}_I^{\text{T}} \\ \mathbf{0}_I & Q \end{pmatrix}.$$

The minimizer is the (sum) constrained WLS estimator

$$\hat{\pi} = \hat{\pi}_U - \frac{(\hat{A}_*^T Q_* \hat{A}_*)^{-1} \mathbf{1}_S (\mathbf{1}_S^T \hat{\pi}_U - 1)}{\mathbf{1}_S^T (\hat{A}_*^T Q_* \hat{A}_*)^{-1} \mathbf{1}_S},$$

where $\hat{\pi}_U = (\hat{A}_*^T Q_* \hat{A}_*)^{-1} \hat{A}_*^T Q_* \bar{\mathbf{b}}_*$ is the unconstrained WLS estimator.

This formulation allows us to consider simultaneously the unique solution case $I + 1 = S$ and the overconstrained model where $I + 1 > S$. If $I + 1 = S$, the constraint $\mathbf{1}_S^T \hat{\pi}_U = 1$ is automatically satisfied and Q is irrelevant as $\hat{\pi} = \hat{\pi}_U = \hat{A}_*^{-1} \bar{\mathbf{b}}_*$. Alternatively, when $I + 1 > S$, it is unnecessary to account for the sum constraint in the objective function as $\hat{\pi}$ can be obtained by minimizing $(\bar{\mathbf{b}} - \hat{A}\pi)^T Q(\bar{\mathbf{b}} - \hat{A}\pi)$ subject to $\mathbf{1}_S^T \pi = 1$, giving the alternative equivalent expression

$$\hat{\pi} = \hat{\pi}_U - \frac{(\hat{A}^T Q \hat{A})^{-1} \mathbf{1}_S (\mathbf{1}_S^T \hat{\pi}_U - 1)}{\mathbf{1}_S^T (\hat{A}^T Q \hat{A})^{-1} \mathbf{1}_S},$$

where $\hat{\pi}_U = (\hat{A}^T Q \hat{A})^{-1} \hat{A}^T Q \bar{\mathbf{b}}$. Our development provides one result for both settings even though the natural reduced forms of the estimators just given appear functionally distinct.

A direct calculation shows that $\hat{\pi} - \pi = \hat{L}_* (\bar{\mathbf{b}}_* - \hat{A}_* \pi)$, where

$$\hat{L}_* = L_*(\hat{A}, Q) = \left\{ I_S - \frac{(\hat{A}_*^T Q_* \hat{A}_*)^{-1} \mathbf{1}_S \mathbf{1}_S^T}{\mathbf{1}_S^T (\hat{A}_*^T Q_* \hat{A}_*)^{-1} \mathbf{1}_S} \right\} (\hat{A}_*^T Q_* \hat{A}_*)^{-1} \hat{A}_*^T Q_*. \tag{2}$$

If, as assumed, $\sqrt{J}(\bar{\mathbf{b}} - \beta, \hat{\theta} - \theta)$ has a limiting normal distribution with mean $\mathbf{0}$ so that $\sqrt{J}(\bar{\mathbf{b}} - \hat{A}\pi)$ has a limiting $N(\mathbf{0}_I, \Sigma)$ distribution, then $\sqrt{J}(\bar{\mathbf{b}}_* - \hat{A}_* \pi) \sim N(\mathbf{0}_{I+1}, \Sigma_*)$ asymptotically, where

$$\Sigma_* = \begin{pmatrix} 0 & \mathbf{0}_I^T \\ \mathbf{0}_I & \Sigma \end{pmatrix}.$$

Assuming that $\hat{L}_*$ converges in probability to $L_* = L_*(A, Q)$, as J increases we find that $\sqrt{J}(\hat{\pi} - \pi)$ has an asymptotic $N(\mathbf{0}_S, V_Q)$ distribution, where $V_Q = L_* \Sigma_* L_*^T$.

3.3. *Further properties of estimators and discussion of results*

The following five results are relatively straightforward to show.

(a) If the linear WLS estimator $\hat{\pi}$ has non-negative elements, then $\hat{\pi} = \pi(\tilde{\alpha})$ for some $\tilde{\alpha}$ that satisfies $\dot{\mathbf{F}}(\tilde{\alpha}) = \mathbf{0}$. Consequently, $\hat{\pi} = \tilde{\pi}$ when $\hat{\pi}$ is a probability vector.
(b) The limiting distributions of $\tilde{\pi}$ and $\hat{\pi}$ are identical, i.e. $V_Q = W_Q$.
(c) The asymptotically optimal choice for Q is $Q = \Sigma^{-1}$, i.e. $V_Q - V_{\Sigma^{-1}}$ is positive semidefinite for each Q.
(d) The same limiting distribution is obtained if Q is replaced by a consistent estimator $\hat{Q}$.
(e) $V_Q = V = A_*^{-1} \Sigma_* A_*^{-T}$ for the unique solution case $I + 1 = S$.

Our result for the unique solution case ($I + 1 = S$) extends Phillips and Gregg (2001) in two important ways. First, Phillips and Gregg (2001) considered only $I \leqslant 2$. Their derivation expresses π as a linear combination of elements of A and β and then uses the delta method to derive variances. Separate derivations are provided for $I = 1$ and $I = 2$. We use a simpler matrix approach that applies to an arbitrary number of isotopes. Second, we allow for general estimation of discrimination, whereas Phillips and Gregg (2001) treated discrimination as known.

Our focus on moment-based estimators allows π to be estimated from published summary data (mean vectors and covariance matrices). However, the non-linear WLS estimator $\tilde{\pi}$ is also a pseudo-maximum-likelihood estimator (Gong and Samaniego, 1981) under certain conditions. In particular, suppose that $L(\alpha, \theta, \Sigma_b)$ is the likelihood function for the consumer model assuming $\mathbf{b}_j \sim N(\beta, \Sigma_b)$, where $\beta = A(\theta)\pi(\alpha)$. A pseudo-maximum-likelihood estimator for (α, Σ_b) maximizes $L(\alpha, \hat{\theta}, \Sigma_b)$ for some consistent estimator $\hat{\theta}$. The pseudo-maximum-likelihood estimator of α is a non-linear WLS estimator with weight matrix given by the inverse of the consumer sample covariance matrix. Interestingly, this estimator cannot be more efficient than the non-linear WLS estimator with the optimal weights $Q = \Sigma^{-1}$, which accounts for uncertainty in $\hat{\theta}$.

3.4. Estimates for the dunlin study

For the dunlin study, π_1 and $\pi_2 = 1 - \pi_1$ are the population diet proportions that are associated with terrestrial and marine plant invertebrates respectively. The linear WLS estimate of π_1 using estimated optimal weights is $\hat{\pi}_1 = 0.385$ with estimated standard deviation $\text{SD} = 0.036$. The weights are based on the ordinary least squares (OLS) estimate of π_1 and sample covariance matrices. An approximate 95% confidence interval for π_1 is $(0.315, 0.455)$. The linear OLS estimate is 0.369 with estimated standard deviation 0.038. As $0 < \hat{\pi}_1 < 1$, the linear and non-linear estimates are identical.

3.5. Small sample properties of estimators

Large sample inferences might be inaccurate in the dunlin study because the diet experiment estimates discrimination from four dunlin. However, the variation in the dunlin IRs in the diet experiment is very small relative to source variation (see Table 1), so this small sample size may not be a serious concern. To address this and related issues, we designed a study to examine the small sample properties of $\pi(\tilde{\alpha})$ and $\hat{\pi}$ in the dunlin study. We simulated data from multivariate normal distributions for the sources, the blood and diet components of discrimination and for the IR responses. The distributional assumptions are reasonably consistent with the observed distributions. Population means, population covariance matrices and sample sizes were set to the observed summaries, except that the mean for the IR dunlin response distribution was $\hat{A}\pi$, where π was allowed to vary, and the covariance matrix for the distribution of the dunlin blood samples in the diet experiment was defined to be a multiple $r \geqslant 1$ of the observed covariance matrix. These modifications provide a simple way to assess the effect of applying the asymptotic results when the discrimination is highly variable and estimated from a small sample.

Table 2 gives the estimated biases and root-mean-square errors of $\tilde{\pi}_1$ and $\hat{\pi}_1$ for $r = 1, 16, 64$ and $\pi_1 = 0.03, 0.06, 0.12, 0.20, 0.40$, along with the proportion of samples for which $\hat{\pi}_1 < 0$. Results are for optimal WLS estimators and 2500 simulated samples. The non-linear WLS estimator $\tilde{\pi}_1$ was 0 whenever $\hat{\pi}_1 < 0$. The proportion of samples with $\hat{\pi}_1 < 0$ increases as either π_1 decreases or the variation in $\hat{A}$ increases. In general, $\tilde{\pi}_1$ has a smaller root-mean-square error but larger bias than $\hat{\pi}_1$, which is nearly unbiased. Table 2 also gives the estimated percentage coverage of the nominal 95% large sample confidence interval $\max(0, \hat{\pi}_1) \pm 1.96\,\text{SD}$. The standard errors of the estimated percentages are at most 1%. As $\hat{\pi}_1$ is truncated at 0, the coverages closely agree with coverages of large sample intervals based on $\tilde{\pi}_1$. The coverages are close to the nominal level, except when the variation in the diet experiment ($r = 64$) is too large for the given sample size. The overall differences between the linear and non-linear WLS estimators are relatively small here but could be substantial in settings with three or more isotopes when at least two of the proportions are small.

Table 2. Estimated bias and $\sqrt{}$MSE for $\tilde{\pi}_1$ and $\hat{\pi}_1$, estimated $\Pr(\hat{\pi}_1 < 0)$ and percentage coverage of the nominal 95% confidence interval based on $\hat{\pi}_1$ for selected combinations of π_1 and variance multiplier r†

r	π_1	$\widehat{Pr}(\hat{\pi}_1 < 0)$	$\hat{\pi}_1$		$\tilde{\pi}_1$		*Estimated confidence interval coverage*
			Bias	$\sqrt{}$*MSE*	*Bias*	$\sqrt{}$*MSE*	
1	0.03	0.25	0.000	0.045	0.007	0.036	96.1
	0.06	0.08	0.001	0.042	0.003	0.039	95.8
	0.12	0.00	0.001	0.040	0.001	0.040	92.8
	0.20	0.00	0.001	0.039	0.001	0.039	93.9
	0.40	0.00	0.001	0.039	0.001	0.039	93.9
16	0.03	0.30	0.000	0.063	0.014	0.047	94.1
	0.06	0.16	0.002	0.061	0.007	0.052	94.2
	0.12	0.02	0.001	0.060	0.002	0.059	91.9
	0.20	0.00	0.002	0.060	0.002	0.060	91.6
	0.40	0.00	0.002	0.060	0.002	0.060	91.6
64	0.03	0.37	−0.000	0.098	0.026	0.070	91.6
	0.06	0.26	−0.002	0.095	0.015	0.074	92.0
	0.12	0.10	−0.000	0.097	0.005	0.087	90.0
	0.20	0.02	0.005	0.096	0.005	0.094	85.0
	0.40	0.02	0.005	0.096	0.005	0.094	85.0

†The results are based on 2500 simulated samples.

We concluded that the large sample approximations were tenable for the dunlin study but could be an issue in other settings. If needed, bootstrap methods are an alternative to the large sample delta methods that are presented here. For the dunlin study, the basic, percentile and Studentized bootstrap confidence intervals (Davison and Hinkley (1997), page 199) had similar coverages and lengths to those of the large sample delta method based interval $\max(0, \hat{\pi}_1) \pm 1.96\,\mathrm{SD}$.

3.6. The importance of weighting

The asymptotic variance $V_{\Sigma^{-1}}$ of the optimal WLS estimator cannot be reduced by omitting isotopes from the estimation. To be specific, suppose that π is estimated by $\hat{\pi}_R$, which is a WLS estimator from a reduced set of $I_R < I$ isotopes ($I_R + 1 \geqslant S$). Then, $\mathrm{cov}(\hat{\pi}_R) - V_{\Sigma^{-1}}$ is positive semidefinite. In the dunlin study, π can be estimated by using carbon and nitrogen individually or by using both stable isotopes. However, the asymptotic variances of the two single-isotope estimators cannot be smaller than the variance of the optimal WLS estimator based on both isotopes. There is no distinction between WLS and OLS in the single-isotope analyses because they are equivalent for the unique solution case ($I + 1 = S$, with $I = 1$ and $S = 2$). We find that $(\hat{\pi}_1, \mathrm{SD}) = (0.355, 0.045)$ by using carbon only and $(\hat{\pi}_1, \mathrm{SD}) = (0.445, 0.060)$ by using nitrogen only, whereas $(\hat{\pi}_1, \mathrm{SD}) = (0.385, 0.036)$ by using both carbon and nitrogen. There is a noticeable reduction in the estimated uncertainty when both isotopes are used together.

This reduction in variance is only guaranteed when Q is chosen optimally. A poor choice of Q may increase the uncertainty in the estimated diet proportions when isotopes are added to the estimation. To illustrate, Table 3 gives asymptotic SDs for the dunlin data, fixing $\pi_1 = \hat{\pi}_1$ and setting all population means and covariance matrices at the observed values, with the exception that $\widehat{\mathrm{var}}(\bar{b}_2)$ and $\widehat{\mathrm{cov}}(\bar{b}_1, \bar{b}_2)$ were replaced by $\omega^2 \widehat{\mathrm{var}}(\bar{b}_2)$ and $\omega \widehat{\mathrm{cov}}(\bar{b}_1, \bar{b}_2)$ respectively. Values of $\omega > 1$ correspond to increasing the variability of the nitrogen IR in the consumer, leaving all

Table 3. Loss of efficiency in the OLS estimator as a function of consumer nitrogen isotope ratio variance inflation factor ω†

ω	$WLS(C+N)$	$OLS(C+N)$	C only	N only
1	0.036	0.038	0.044	0.062
10	0.044	0.056	0.044	0.211
20	0.043	0.083	0.044	0.408
30	0.042	0.112	0.044	0.609
40	0.042	0.142	0.044	0.811
50	0.042	0.173	0.044	1.012

†Values are the asymptotic SD as a function of ω for four possible estimators of π_1, the population diet proportion associated with the terrestrial source.

other summaries fixed. Estimators of π_1 that use either the carbon (labelled C only) or nitrogen IR (N only) individually were considered in addition to the optimal WLS estimator and the OLS estimator that use both IRs simultaneously (labelled WLS(C+N) and OLS(C+N)). As ω increases, the importance of proper weighting becomes more apparent as the OLS estimator based on both isotopes becomes considerably more variable than the estimator based on carbon only.

4. Temporal modelling of diet

Temporal changes in dunlin diet might be expected over the January–April 2000 sampling period, which would be reflected in temporal trends in dunlin IRs. Fig. 2(a) shows time trends in dunlin IRs as a function of sampling date. Evans Ogden *et al.* (2005) supported the position that diet is temporally dependent and predicted an increase in terrestrial habitat use by dunlin during periods of heavy rain and an increase in field feeding during lower temperatures and higher wind speeds. This suggests that π should be modelled as a function of time, in contrast with our earlier analysis that treated dunlin diet as static. A typical diet may also depend on age and sex. However, the mean source IRs should remain relatively constant over this short sampling period but could depend on dunlin covariates.

To generalize the BMM, we focus on the temporal component of diet and assume that $\beta(t) = A\pi(t)$, where the consumer IR mean $\beta(t)$ and the diet proportion vector $\pi(t)$ are functions of time t, but A is constant. The observed IR responses and sampling times $(\mathbf{b}_j, t_j)$ are assumed to be independent across individuals with $E(\mathbf{b}_j|t_j) = \beta(t_j)$ and $\mathrm{cov}(\mathbf{b}_j|t_j) = \Sigma_b$ for $j = 1, \ldots, J$. In the direct estimation method, we assume a model $\pi(t) = \pi(t; \alpha)$ and estimate α by minimizing $\mathbf{F}(\alpha) = \Sigma_{j=1}^{J} (\mathbf{b}_j - \hat{A}\pi(t_j; \alpha))^{\mathrm{T}} Q (\mathbf{b}_j - \hat{A}\pi(t_j; \alpha))$. A natural choice is the logistic model $\pi(t; \alpha) = (\pi_1(t; \alpha), \ldots, \pi_S(t; \alpha))^{\mathrm{T}}$, where

$$\pi_j(t; \alpha) = \exp\{\mathbf{x}_j(t)^{\mathrm{T}} \alpha_j\} \Big/ \sum_{l=1}^{S} \exp\{\mathbf{x}_l(t)^{\mathrm{T}} \alpha_l\}$$

for $j = 1, \ldots, S$ and $\alpha = (\alpha_1, \ldots, \alpha_{S-1})$ with $\alpha_S = \mathbf{x}_S(t) = \mathbf{0}$. The asymptotic distributions of the WLS estimators $\tilde{\alpha}$ and $\pi(t; \tilde{\alpha})$ are presented in Appendix A. The results apply if Q is based on a consistent estimator.
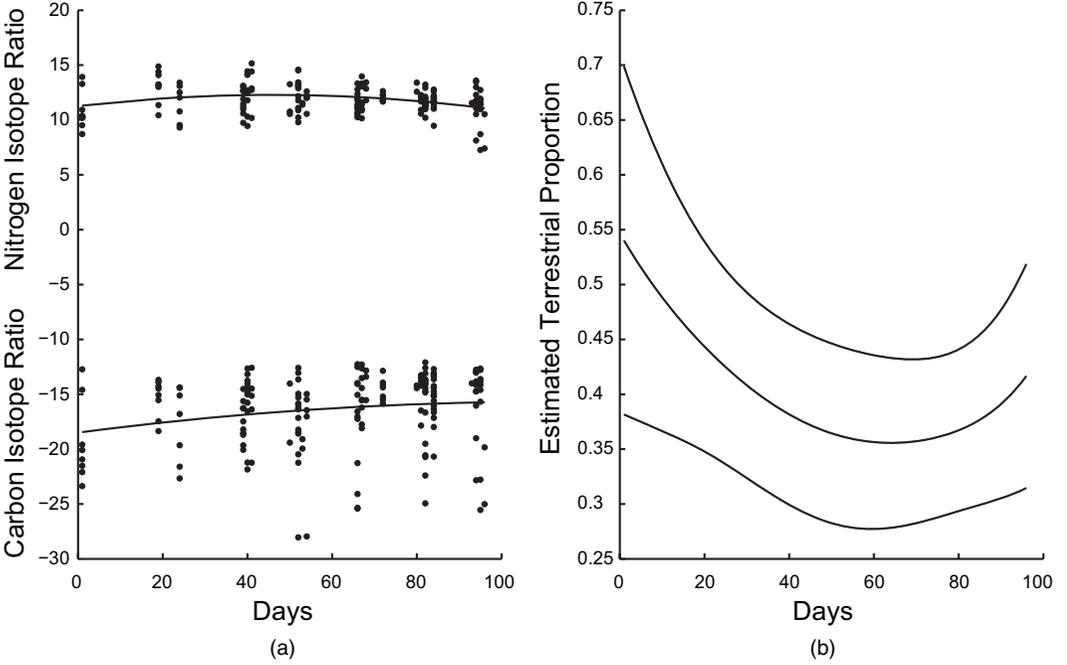
Fig. 2.   (a) Observed carbon and nitrogen IRs with quadratic fits and (b) indirectly estimated population diet proportion associated with terrestrial source as a function of time, with pointwise 95% confidence intervals based on a quadratic fit

An alternative approach is to estimate $\pi(t)$ indirectly by fitting the working model $\hat{\beta}_*(t) = \hat{A}_* \pi(t) + \mathbf{e}_*(t)$ by linear WLS, where $\hat{\beta}_*(t)^{\mathrm{T}} = (1 \ \hat{\beta}(t)^{\mathrm{T}})$ for some estimator $\hat{\beta}(t)$ of $\beta(t)$. This is a two-step process where the first step is to compute $\hat{\beta}(t)$ for each t followed by the fitting of the working model to obtain $\hat{\pi}(t)$. For simplicity, we consider the OLS estimator $\hat{\beta}(t) = B^{\mathrm{T}} U (U^{\mathrm{T}} U)^{-1} \mathbf{u}(t)$ from the multivariate regression model $\beta(t) = \Lambda \mathbf{u}(t)$, where Λ is an $I \times p$ matrix of parameters and $\mathbf{u}(t)$ is a $p \times 1$ predictor, e.g. a low order polynomial. Here, U and B have rows $\mathbf{u}(t_j)^{\mathrm{T}}$ and $\mathbf{b}_j^{\mathrm{T}}$ for $j = 1, \ldots, J$ respectively. Assuming regularity conditions on the design points (Sen and da Motta Singer (1993), chapter 7), $\sqrt{J}\{\hat{\beta}(t) - \beta(t)\}$ has a limiting normal distribution with mean $\mathbf{0}_I$ and thus typically $\sqrt{J}\{\hat{\beta}(t) - \hat{A}\pi(t)\} \sim N(\mathbf{0}_I, \Sigma(t))$ asymptotically for each t. If so, then our limiting result in Section 3.2 for the linear WLS estimator $\hat{\pi}$ applies to $\sqrt{J}\{\hat{\pi}(t) - \pi(t)\}$ with $\Sigma_*(t)$ defined in terms of $\Sigma(t) = \mathrm{cov}[\sqrt{J}\{\hat{\beta}(t) - \hat{A}\pi(t)\}]$. For the dunlin study, $\Sigma(t)$ has the form of equation (1) with Σ_b/J replaced by $\mathbf{u}(t)^{\mathrm{T}} (U^{\mathrm{T}} U)^{-1} \mathbf{u}(t)\Sigma_b$. Non-linear WLS can be used in the second step of the two-step process without changing the limiting result. Splines and local polynomial smoothers (Ruppert *et al.*, 2003) offer a different approach to smoothing IR responses.

4.1. *Dunlin temporal analysis*

Fig. 2(a) plots the observed and fitted dunlin IRs as a function of sampling date assuming that $\beta(t)$ follows a quadratic model ($\mathbf{u}(t)^{\mathrm{T}} = (1 \ t \ t^2)$). The two fitted curves follow the IR trends reasonably well. Given $\hat{\beta}(t)$, Fig. 2(b) plots the indirectly estimated proportion of diet attributable to terrestrial sources $\hat{\pi}_1(t)$ along with pointwise 95% confidence intervals. Optimal WLS was used for each t. The estimated proportion tends to decrease during the first month,
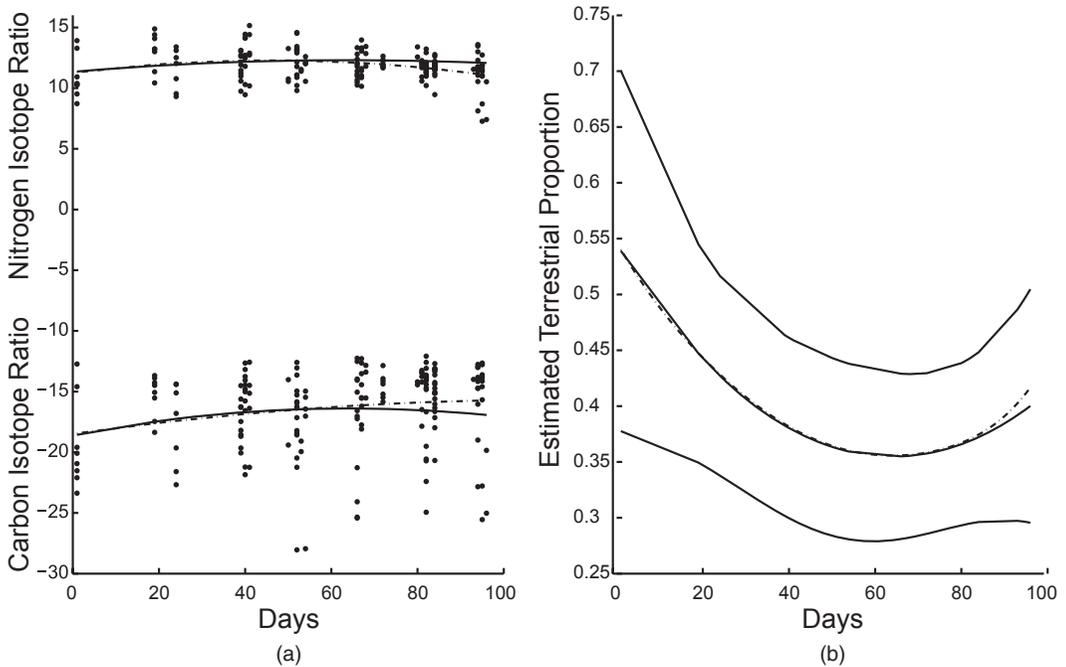
Fig. 3. (a) Observed and estimated (————) mean carbon and nitrogen IRs based on a quadratic fit to the logit of the population diet proportion associated with terrestrial source ($\cdot\text{--}\cdot\text{--}$, estimate based on quadratic models for IRs) and (b) estimated population diet proportion associated with terrestrial source as a function of time, with pointwise 95% confidence interval, based on a quadratic logit model ($\cdot\text{--}\cdot\text{--}$, indirect estimate based on quadratic models for IRs)

followed by a smaller increase in the proportion at later sampling times. The estimate shows a marked difference from the 'static' estimate of 0.385 based on ignoring time.

Alteratively, we used non-linear WLS to estimate $\pi_1(t)$ directly, assuming that $\text{logit}\{\pi_1(t)\}$ was a quadratic function of time. Fig. 3(b) plots $\pi_1(t; \tilde{\alpha})$ along with pointwise 95% confidence intervals. The weighting matrix used optimal weights assuming that $\pi(t)$ was constant. Fig. 3(a) plots the dunlin IRs with the estimated IR means $\beta(t; \tilde{\alpha}) = \hat{A}\pi(t; \tilde{\alpha})$. The plots include the point estimates from the indirect method for comparison. Summaries from the two estimation methods including confidence intervals for $\pi_1(t)$ are nearly identical for $t < 75$ days. Differences between the two methods are only noticeable near the ends of the sampling period, where $\beta(t; \tilde{\alpha})$ does not follow the IR responses as well as $\hat{\beta}(t)$.

4.2. Does dunlin diet change with time?

The direct and indirect methods are two distinct approaches to account for time. We made no attempt in our analysis to have the parametric models for $\beta(t)$ and $\pi_1(t; \alpha)$ compatible. In particular, if we had assumed that $\pi_1(t)$ was a quadratic function of time, rather than $\text{logit}\{\pi_1(t)\}$, then $\beta(t)$ would also be a quadratic in time assuming that the BMM holds. The similarity of the two fits is largely because $\text{logit}\{\tilde{\pi}_1(t)\}$ is a roughly linear function of $\tilde{\pi}_1(t)$ over the range of $\tilde{\pi}_1(t)$.

There is one important distinction between the analyses. The estimated constant, linear and quadratic coefficients in the directly estimated logistic model are $(-0.597, -0.025, 0.126)$ with standard errors of $(0.169, 0.097, 0.078)$ respectively. The Wald statistic for testing the significance of the linear and quadratic coefficients is 4.19 with a p-value of 0.12. This suggests that the

marked difference between the quadratic logistic trend and the static estimate that ignores time is not statistically significant. However, the four linear and quadratic effects in the parametric model for the mean dunlin IRs are simultaneously highly significant ($p < 0.001$), leading to a conclusion that the dunlin mean IRs change over time. If the BMM holds, then the diet proportions must also change with time: a conclusion that is reached without having to estimate either A or $\pi_1(t)$. A possible explanation for the discrepancy in conclusions is that $\text{var}(\tilde{\alpha})$ accounts for A being estimated, thus possibly resulting in a less sensitive test of whether mean diet changes with time.

4.3.   Model specification and checking

A challenge with using the direct estimation method is that the responses $\mathbf{b}_i$ do not provide direct visual information about plausible models $\pi(t; \alpha)$. Although the residuals $\mathbf{b}_i - \beta(t_i; \tilde{\alpha})$ highlight lack of fit, they do not distinguish misspecification of the BMM from misspecification of $\pi(t; \alpha)$. In essence, the simplicity of the direct method comes at the expense of having to assume that the BMM holds. In contrast, the indirect method estimates $\beta(t)$ independently of the BMM and thus provides a formal check on the BMM, in the spirit of classical lack-of-fit tests. However, an important consideration here is that parametric models for $\beta(t)$ that are used with the indirect method must be sufficiently flexible to allow the BMM to hold. In the dunlin data, the BMM stipulates that $\beta(t)$ lies on the line segment joining the discrimination-corrected source means. A polynomial model for $\beta(t)$, such as that considered in our analysis, allows this condition to be satisfied though constraints on the polynomial coefficients are implied by the BMM. These constraints are not used when fitting the parametric model for $\beta(t)$; rather the fits from the parametric model are used to estimate $\pi(t)$ under the BMM. The indirectly estimated proportion $\hat{\pi}(t)$ may be badly biased if the parametric functions that are used with $\beta(t)$ are incompatible with the relationships between the components of $\beta(t)$ implied by the BMM.

To test lack of fit, we consider the indirect method residuals $\mathbf{R}_*(t) = \hat{\beta}_*(t) - \hat{A}_* \hat{\pi}(t)$, which have an approximate mean of $\mathbf{0}$ assuming that the BMM and the model for $\beta(t)$ hold. Following equation (2),

$$\mathbf{R}_*(t) = \hat{\beta}_*(t) - \hat{A}_* \pi(t) - \hat{A}_*(\hat{\pi}(t) - \pi(t)) = (I_{I+1} - \hat{A}_* \hat{L}_*)(\hat{\beta}_*(t) - \hat{A}_* \pi(t))$$

and hence asymptotically

$$\text{cov}\{\sqrt{J}\mathbf{R}_*(t)\} = (I_{I+1} - A_* L_*)\Sigma_*(t)(I_{I+1} - A_* L_*)^{\mathrm{T}} = \Sigma_{\mathbf{R}_*}(t).$$

This suggests that $\text{LOF}(t) = J \mathbf{R}_*(t)^{\mathrm{T}} \hat{\Sigma}_{\mathbf{R}_*}^{-}(t) \mathbf{R}_*(t) \overset{\sim}{\sim} \chi^2_{I+1-S}$ to test model adequacy, where $\hat{\Sigma}_{\mathbf{R}_*}(t)$ is a consistent estimator of $\Sigma_{\mathbf{R}_*}(t)$. This check requires $I + 1 > S$ as $\mathbf{R}_*(t) = \mathbf{0}$ for the unique solution case with $I + 1 = S$.

In the dunlin study, $\text{LOF}(t) > \chi^2_{1,.95}$ for $t > 75$ days with the quadratic model for $\beta(t)$. As the fitted quadratic model tracks the temporal trend in the IRs, lack of fit is likely to be due to inadequacy of the BMM for $t > 75$, possibly because of a sampling bias or a temporal change in the discrimination-corrected mean source IRs (i.e. A). Recall that the main differences between the indirect and direct estimates of $\pi(t)$ occurred at the ends of the sampling period.

5.   Concluding remarks

We presented two extensions of existing methodology for estimating source proportions based on the BMM. First, we generalized Phillips and Gregg (2001) to overconstrained systems with an arbitrary number of isotopes, allowing for fairly general estimation of the discrimination-corrected mean source matrix A. Second, we proposed methods for modelling the diet proportion

vector as a function of time based on cross-sectional IR response data. The procedures proposed are easy to implement by using either linear or non-linear WLS. Extensions that allow the diet proportion vector to depend on fixed or time varying covariates are straightforward. Another natural generalization of the BMM would allow the source means to depend on time, e.g. $\beta(t) = A(t)\pi(t)$, which at a minimum would require sampling sources over time. Similar developments for the extended mixing model that allow variation in elemental concentrations and assimilation efficiency (Koch and Phillips, 2002) are also possible.

Our temporal analysis based on a quadratic fit suggests that the proportion of dunlin diet that is attributable to terrestrial sources changes non-linearly over the sampling period. Although a quadratic model probably does not best represent the biological and ecological process that govern dunlin diet, we believe that our conclusions are fairly robust to the choice of this model. In particular, an alternative approach is to estimate indirectly the proportion of dunlin diet attributable to terrestrial sources from a non-parametric local linear or spline fit to the carbon and nitrogen IRs. The local linear fit fairly closely agreed with the quadratic fit that is illustrated in Fig. 2(a). Furthermore, a plot of the indirectly estimated proportion with confidence bands as a function of time exhibited some oscillation about a quadratic trend but was fairly consistent with the summaries that are presented in Fig. 2(b). Additional research into the use of smoothing methods for the estimation of temporal trends is warranted.

Acknowledgements

Appendix A: Technical results

A.1.   Asymptotic distribution of weighted least squares estimator $\pi(t; \tilde{\alpha})$ in temporal model

Let $\dot{\pi}_j$ be the $S \times d$ matrix of partial derivatives of $\pi_j = (\pi_{j1}, \ldots, \pi_{jS})^{\mathrm{T}} = \pi(t_j; \alpha)$ with respect to α, where $d = \dim(\alpha)$. For the logistic model

$$\pi_{js} = \exp\{\mathbf{x}_s(t_j)^{\mathrm{T}}\alpha_s\} \Big/ \sum_{l=1}^{S} \exp\{\mathbf{x}_l(t_j)^{\mathrm{T}}\alpha_l\}$$

for $s = 1, \ldots, S$ with $\alpha^{\mathrm{T}} = (\alpha_1, \ldots, \alpha_{S-1})$ and $\alpha_S = \mathbf{x}_S(t) = 0$, we find that $\dot{\pi}_j = (D(\pi_j) - \pi_j \pi_j^{\mathrm{T}})M(t_j)$, where $M(t)$ is the block diagonal matrix $\mathrm{BlkDiag}\{\mathbf{x}_1(t)^{\mathrm{T}}, \ldots, \mathbf{x}_{S-1}(t)^{\mathrm{T}}\}$ stacked on $\mathbf{0}_d^{\mathrm{T}}$.

Consider the estimating equation $R(\alpha; \theta)^{\mathrm{T}} = (R_1(\alpha; \theta)^{\mathrm{T}}, R_2(\alpha; \theta)^{\mathrm{T}}) = \mathbf{0}_{d+f}^{\mathrm{T}}$, where

$$R_1(\alpha; \theta) = \frac{1}{J} \sum_{j=1}^{J} \dot{\pi}_j^{\mathrm{T}} A(\theta)^{\mathrm{T}} Q(\mathbf{b}_j - A(\theta)\pi(t_j; \alpha)),$$

$R_2(\alpha, \theta) = \hat{\theta} - \theta$ and $f = \dim(\theta)$. Note that $R_1(\alpha, \hat{\theta}) = \mathbf{0}_d$ is the estimating equation that corresponds to minimizing the objective function $\mathbf{F}(\alpha)$ in Section 4. Thus, the asymptotic distribution of the solution $(\tilde{\alpha}, \hat{\theta})$ to $R(\alpha, \theta) = 0$ provides the asymptotic distribution for the minimizer of $\mathbf{F}(\alpha)$. Although Q is assumed to be fixed, this limiting distribution is not affected by using a consistent estimator of Q in $\mathbf{F}(\alpha)$ provided that Q is estimated independently of the estimating equation.

Let

$$\dot{R}(\alpha; \theta) = \begin{pmatrix} \dfrac{\partial R_1}{\partial \alpha} & \dfrac{\partial R_1}{\partial \theta} \\ \dfrac{\partial R_2}{\partial \alpha} & \dfrac{\partial R_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial R_1}{\partial \alpha} & \dfrac{\partial R_1}{\partial \theta} \\ \mathbf{0}_{f \times d} & -I_f \end{pmatrix},$$

where

$$J \frac{\partial R_1}{\partial \alpha_a} = -\sum_{j=1}^{J} \dot{\pi}_j^{\mathrm{T}} A(\theta)^{\mathrm{T}} Q A(\theta)\dot{\pi}_{j(a)} + \sum_{j=1}^{J} \ddot{\pi}_{j(a)}^{\mathrm{T}} A(\theta)^{\mathrm{T}} Q(\mathbf{b}_j - A(\theta)\pi_j)$$

and

$$J\frac{\partial R_1}{\partial \boldsymbol{\theta}_b} = \sum_{j=1}^{J} \dot{\boldsymbol{\pi}}_j^{\mathrm{T}} \dot{A}_b^{\mathrm{T}} Q(\mathbf{b}_j - A(\boldsymbol{\theta})\boldsymbol{\pi}_j) - \sum_{j=1}^{J} \dot{\boldsymbol{\pi}}_j^{\mathrm{T}} A(\boldsymbol{\theta})^{\mathrm{T}} Q \dot{A}_b \boldsymbol{\pi}_j.$$

Here, $\dot{\boldsymbol{\pi}}_{j(a)}$ is the column vector that is obtained by differentiating each element of $\boldsymbol{\pi}_j$ with respect to $\boldsymbol{\alpha}_a$, whereas $\ddot{\boldsymbol{\pi}}_{j(a)}$ is the matrix that is obtained by differentiating each element of $\dot{\boldsymbol{\pi}}_{j(a)}$ with respect to $\boldsymbol{\alpha}_a$. Similarly, $\dot{A}_b$ is the matrix resulting from differentiating each element of $A(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}_b$.

We assume sufficient regularity conditions so that standard theory for solutions to estimating equations is appropriate; see, for example, Aerts *et al.* (2002), chapter 5. For example, noting that $E\{\mathbf{b}_j - A(\boldsymbol{\theta})\boldsymbol{\pi}_j\} = \mathbf{0}_I$, we assume that laws of large numbers apply so that $\partial R_1/\partial \boldsymbol{\alpha}$ converges to

$$R_{11} = -\lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \dot{\boldsymbol{\pi}}_j^{\mathrm{T}} A(\boldsymbol{\theta})^{\mathrm{T}} Q A(\boldsymbol{\theta})\dot{\boldsymbol{\pi}}_j \equiv -\lim \mathrm{ave}\{\dot{\boldsymbol{\pi}}_j^{\mathrm{T}} A(\boldsymbol{\theta})^{\mathrm{T}} Q A(\boldsymbol{\theta})\dot{\boldsymbol{\pi}}_j\}$$

and $\partial R_1/\partial \boldsymbol{\theta}$ converges to the $d \times f$ matrix R_{12}, which has bth column

$$\mathbf{R}_{12(b)} = -\lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \dot{\boldsymbol{\pi}}_j^{\mathrm{T}} A(\boldsymbol{\theta})^{\mathrm{T}} Q \dot{A}_b \boldsymbol{\pi}_j \equiv -\lim \mathrm{ave}\{\dot{\boldsymbol{\pi}}_j^{\mathrm{T}} A(\boldsymbol{\theta})^{\mathrm{T}} Q \dot{A}_b \boldsymbol{\pi}_j\} \qquad b = 1, \dots, f.$$

We assume that $\hat{\boldsymbol{\theta}}$ is independent of the $\mathbf{b}_j$s and that $\sqrt{J}(R_1(\boldsymbol{\alpha}; \boldsymbol{\theta}), R_2(\boldsymbol{\alpha}; \boldsymbol{\theta}))$ has a limiting normal distribution with mean $\mathbf{0}$ and diagonal block covariance matrix $\Sigma(\boldsymbol{\alpha}, \boldsymbol{\theta})$ with blocks

$$\Sigma_1(\boldsymbol{\alpha}; \boldsymbol{\theta}) = \lim \mathrm{ave}[\mathrm{cov}\{\dot{\boldsymbol{\pi}}_j^{\mathrm{T}} A(\boldsymbol{\theta})^{\mathrm{T}} Q \mathbf{b}_j\}] = \lim \mathrm{ave}\{\dot{\boldsymbol{\pi}}_j^{\mathrm{T}} A(\boldsymbol{\theta})^{\mathrm{T}} Q \Sigma_b Q A(\boldsymbol{\theta})\dot{\boldsymbol{\pi}}_j\}$$

and $\Sigma_2(\boldsymbol{\theta})$. Recall that $\hat{\boldsymbol{\theta}}$ is a vector of sample means, so establishing the asymptotic joint normality is simpler. Following Aerts *et al.* (2002), page 84, the limiting joint distribution of $\sqrt{J}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ and $\sqrt{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is normal with mean $\mathbf{0}$ and covariance matrix $(-\dot{R}_{\lim})^{-1}\Sigma(\boldsymbol{\alpha}, \boldsymbol{\theta})(-\dot{R}_{\lim})^{-\mathrm{T}}$, where $\dot{R}_{\lim} = \lim_{J\to\infty} \dot{R}(\boldsymbol{\alpha}; \boldsymbol{\theta})$ with

$$\dot{R}_{\lim}^{-1} = \begin{pmatrix} R_{11} & R_{12} \\ \mathbf{0}_{d\times f} & I_d \end{pmatrix}^{-1} = \begin{pmatrix} R_{11}^{-1} & -R_{11}^{-1} R_{12} \\ \mathbf{0}_{d\times f} & I_d \end{pmatrix}.$$

The asymptotic distribution of $(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\sqrt{J}$ has covariance matrix

$$R_{11}^{-1}(\Sigma_1(\boldsymbol{\alpha}; \boldsymbol{\theta}) + R_{12}\Sigma_2(\boldsymbol{\theta})R_{12}^{\mathrm{T}})R_{11}^{-\mathrm{T}}.$$

The delta method then implies that $\sqrt{J}\{\boldsymbol{\pi}(t; \tilde{\boldsymbol{\alpha}}) - \boldsymbol{\pi}(t; \boldsymbol{\alpha})\} \sim N_S\{\mathbf{0}_S, W_Q(t)\}$ asymptotically, where

$$W_Q(t) = \dot{\boldsymbol{\pi}}(t)R_{11}^{-1}(\Sigma_1(\boldsymbol{\alpha}; \boldsymbol{\theta}) + R_{12}\Sigma_2(\boldsymbol{\theta})R_{12}^{\mathrm{T}})R_{11}^{-\mathrm{T}}\dot{\boldsymbol{\pi}}(t)^{\mathrm{T}}.$$

A.2. Asymptotic distribution of the non-temporal weighted least squares $\pi(\tilde{\alpha})$ in Section 3.1

The results in Appendix A.1 apply to the simpler model in Section 3.1 with

$$R_1(\boldsymbol{\alpha}, \boldsymbol{\theta}) = (\bar{\mathbf{b}} - A(\boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{\alpha}))^{\mathrm{T}} Q A(\boldsymbol{\theta})\dot{\boldsymbol{\pi}}$$

and $R_{11} = \dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q A \dot{\boldsymbol{\pi}}$. Recalling that Σ is the asymptotic covariance matrix of $\sqrt{J}\{\bar{\mathbf{b}} - A(\hat{\boldsymbol{\theta}})\boldsymbol{\pi}(\boldsymbol{\alpha})\}$ and noting that $\mathrm{cov}[\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q \sqrt{J}\{\bar{\mathbf{b}} - A(\hat{\boldsymbol{\theta}})\boldsymbol{\pi}(\boldsymbol{\alpha})\}]$ is asymptotically equivalent to $\Sigma_1(\boldsymbol{\alpha}; \boldsymbol{\theta}) + R_{12}\Sigma_2(\boldsymbol{\theta})R_{12}^{\mathrm{T}}$ and $\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q \Sigma Q A \dot{\boldsymbol{\pi}}$, it follows that

$$\sqrt{J}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \sim N\{\mathbf{0}_{S-1}, (\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q A \dot{\boldsymbol{\pi}})^{-1}\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q \Sigma Q A \dot{\boldsymbol{\pi}}(\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q A \dot{\boldsymbol{\pi}})^{-1}\}$$

asymptotically, and thus $\sqrt{J}(\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}) \sim N(\mathbf{0}_S, W_Q)$ asymptotically, where

$$W_Q = \dot{\boldsymbol{\pi}}(\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q A \dot{\boldsymbol{\pi}})^{-1}\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q \Sigma Q A \dot{\boldsymbol{\pi}}(\dot{\boldsymbol{\pi}}^{\mathrm{T}} A^{\mathrm{T}} Q A \dot{\boldsymbol{\pi}})^{-1}\dot{\boldsymbol{\pi}}^{\mathrm{T}}.$$

References

Aerts, M., Geys, H., Molenberghs, G. and Ryan, L. (2002) *Topics in Modelling of Clustered Data*. London: Chapman and Hall.

Caut, S., Angulo, E. and Courchamp, F. (2009) Variation in discrimination factors ($\Delta^{15}N$ and $\Delta^{13}C$): the effect of diet isotopic values and applications for diet reconstruction. *J. Appl. Ecol.*, **46**, 443–453.

Currin, C., Newell, S. and Paerl, H. (1995) The role of standing dead *spartina alterniflora* and benthic microalgae in salt marsh food webs: considerations based on multiple stable isotope analysis. *Mar. Ecol. Prog. Ser. Oldndrf*, **121**, 99–116.

Davison, A. and Hinkley, D. (1997) *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

DeNiro, M. and Epstein, S. (1978) Influence of the diet on the distribution of carbon isotopes in animals. *Geochim. Cosmochim. Acta*, **42**, 495–506.

DeNiro, M. and Epstein, S. (1981) Influence of the diet on the distribution of nitrogen isotopes in animals. *Geochim. Cosmochim. Acta*, **48**, 341–351.

Erhardt, E. and Bedrick, E. (2012) A Bayesian framework for stable isotope mixing models. *Environ. Ecol. Statist.*, **19**, 1–21.

Evans Ogden, L., Hobson, K. and Lank, D. (2004) Blood isotopic ($\delta^{13}C$ and $\delta^{15}N$) turnover and diet-tissue fractionation factors in captive dunlin (Calidris alpina pacifica). *Auk*, **121**, 170–177.

Evans Ogden, L. J., Hobson, K. A., Lank, D. B. and Bittman, S. (2005) Stable isotope analysis reveals that agricultural habitat provides an important dietary component for nonbreeding Dunlin. *Avn Conservn Ecol.*, **1**, no. 3.

Felicetti, L., Schwartz, C., Rye, R., Haroldson, M., Gunther, K., Phillips, D. and Robbins, C. (2003) Use of sulfur and nitrogen stable isotopes to determine the importance of whitebark pine nuts to Yellowstone grizzly bears. *Can. J. Zool.*, **81**, 763–770.

Gong, G. and Samaniego, F. (1981) Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.*, **9**, 861–869.

Hobson, K. (1999) Tracing origins and migration of wildlife using stable isotopes: a review. *Oecologia*, **120**, 314–326.

Hobson, K. and Wassenaar, L. (1999) Stable isotope ecology: an introduction. *Oecologia*, **120**, 312–313.

Kalbfleisch, J. and Lawless, J. (1984) Least-squares estimation of transition probabilities from aggregate data. *Can. J. Statist.*, **12**, 169–182.

Kendall, C. and McDonnell, J. (1998) *Isotope Tracers in Catchment Hydrology*. New York: Elsevier Science.

Koch, P. and Phillips, D. (2002) Incorporating concentration dependence in stable isotope mixing models: a reply to Robbins, Hilderbrand and Farley (2002). *Oecologia*, **133**, 14–18.

Minagawa, M. and Wada, E. (1984) Stepwise enrichment of ^{15}N along food chains: further evidence and the relation between $\delta^{15}N$ and animal age. *Geochim. Cosmochim. Acta*, **48**, 1135–1140.

Moore, J. and Semmens, B. (2008) Incorporating uncertainty and prior information into stable isotope mixing models. *Ecol. Lett.*, **11**, 470–480.

Newsome, S., Martinez del Rio, C., Bearhop, S. and Phillips, D. (2007) A niche for isotopic ecology. *Front. Ecol. Environ.*, **5**, 429–436.

Parnell, A., Inger, R., Bearhop, S. and Jackson, A. (2010) Source partitioning using stable isotopes: coping with too much variation. *PLOS ONE*, **5**, no. 3, article e9672.

Parnell, A. and Jackson, A. (2008) siar: Stable Isotope Analysis in *R. R Package Version 3.3*.

Phillips, D. and Gregg, J. (2001) Uncertainty in source partitioning using stable isotopes. *Oecologia*, **127**, 171–179.

Rundel, P., Ehleringer, J. and Nagy, K. (1989) *Stable Isotopes in Ecological Research*. New York: Springer.

Ruppert, D., Wand, M. and Carroll, R. (2003) *Semiparametric Regression*, vol. 12. New York: Cambridge University Press.

Sen, P. and da Motta Singer, J. (1993) *Large Sample Methods in Statistics: an Introduction with Applications*. Boca Raton: Chapman and Hall–CRC.

Ward, E., Semmens, B. and Schindler, D. (2010) Including source uncertainty and prior information in the analysis of stable isotope mixing models. *Environ. Sci. Technol.*, **44**, 4645–4650.