ISSN 1440-771X
ISBN 0 7326 1071 0

# MONASH UNIVERSITY

AUSTRALIA

## An EM Algorithm for Modelling Variably-Aggregated Demand

**Simone Grose and Keith McLaren**

**Working Paper 2/2000**

**March 2000**

## DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

# AN EM ALGORITHM FOR MODELLING VARIABLY-AGGREGATED DEMAND

SIMONE GROSE*

AND

KEITH MCLAREN

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

MONASH UNIVERSITY, AUSTRALIA.

MARCH 2000

* Corresponding author: Dr. S. D. Grose, Department of Econometrics and Business Statistics, Monash University, Clayton, Australia 3800. Email: Simone.Grose@BusEco.monash.edu.au, Tel: +61-3-9905-2964, Fax: +61-3-9905-5474.

# ABSTRACT

The response of consumer demand to prices, income, and other characteristics is important for a range of policy issues. Naturally, the level of detail for which consumer behaviour can be estimated depends on the level of disaggregation of the available data. However, it is often the case that the available data is differently aggregated in different time periods, with the information available in later time periods usually being more detailed. The applied researcher is thus faced with choosing between detail, in which case the more highly aggregated data is ignored; or duration, in which case the data must be aggregated up to the “lowest common denominator”. Furthermore, since parametric demand systems invariably involve a large number of parameters, with the number increasing at least linearly with the number of expenditure categories, it may well be that only the second option is feasible. That is, there is simply not enough data available at the finer aggregation level for the chosen model to be estimated.

This paper develops an EM algorithm for the estimation of a consumer demand system involving variably aggregated data. The methodology is based on the observation that more highly aggregated data does in fact contain information on the finer subcategories. It is therefore possible, under certain simplifying assumptions, to derive the distribution of the unobserved fine-level expenditures conditional on the observed but more highly aggregated data. The expectation of the log-likelihood is then taken with respect to this conditional distribution. Under the assumption of multivariate normality both these steps can be performed analytically, resulting in an EM criterion that can be maximised iteratively at comparatively little cost. The technique is applied to an ABS dataset containing historical information relating to private final consumption expenditures on up to 18 commodities.



KEYWORDS: EM Algorithm, Singular demand systems, Linear expenditure system, Missing data.

JEL classification: C32, C51, D12, E21

# 1. Introduction.

Grose and McLaren (1999) considered the problem of estimating a consumer demand system in a situation in which data is not available on all commodity categories in all time periods. Such a situation can arise quite routinely, since it is common practice to collect, or at least publish, consumption expenditure and price data for more highly disaggregated commodity sets over time, with the result that data is available at differing levels of disaggregation in different time periods. For example, expenditure data may initially be collected for categories “Food”, “Durables” and “Other”; where “Other” is later split into “Other goods” and “Other services”. It is evident, however, that not only does the expenditure on “Other” equal the sum of expenditures on “Other goods” and “Other services”; but that the *expected* expenditure on “Other” also equals the sum of *expected* component expenditures. Furthermore, if an additive stochastic component is assumed, then this also obeys a similar “summing” rule. In other words, the specification of an economic and statistical model for the most disaggregated data naturally implies a corresponding model applying to the data at any level of aggregation.

A known form of aggregation thus implies an “aggregated” model, in which the *observed* expenditure shares are expressed in terms of the model for the complete system. The conventional assumption of additive, normally distributed disturbances then allows the straightforward derivation of the likelihood function implied by differing degrees of commodity disaggregation[1]. Nonetheless, straightforward derivation of the likelihood function does not necessarily equate with straightforward estimation. It turns out that it is no longer possible to “concentrate” the likelihood, leaving all the model parameters, including the $\frac{1}{2}n(n+1)$ parameters of the covariance matrix, to be estimated numerically. The result is an optimization problem found by Grose and McLaren to be infeasible for any realistic sample size.

[1] The “aggregated” likelihood is reviewed in more detail in Section 2.

The strategy considered in Grose and McLaren (1999) (hereafter GM) was to reduce the dimension of the optimization problem by resorting to a reduced-order parameterization of the covariance matrix. This paper considers, instead, a "data augmentation" approach, in which the EM algorithm is used to iteratively maximize the aggregated likelihood, without the need to restrictively reparameterize the model covariance structure.

## 2. The "aggregated" likelihood.

Consider the standard time series specification of an $N$-commodity demand system, in which the $t^{\text{th}}$ observation on the $N$-vector of expenditure shares is modelled as a function of the $N$-vector of prices $\tilde{\mathbf{p}}_t'$, income (assumed equal to total expenditure) in the $t^{\text{th}}$ period $m_t$, a $k$-vector of parameters θ, and an additive, serially independent, zero mean disturbance with constant $N \times N$ positive-semidefinite variance-covariance matrix $\tilde{\mathbf{\Sigma}}$ of rank $n = N - 1$. That is,

$$\tilde{\mathbf{w}}_t' = \tilde{\mathcal{M}}'(\tilde{\mathbf{p}}_t', m_t, \theta) + \tilde{\mathbf{u}}_t' \; ; \quad \tilde{\mathbf{u}}_t \sim (0, \tilde{\mathbf{\Sigma}}) \, , \tag{2.1}$$

where $\tilde{\mathcal{M}}(\tilde{\mathbf{p}}_t', m_t, \theta)$ is the $N$-vector of *expected* expenditure shares, conditional on prices and income, for given θ[2]. The "full rank" system, involving $n$ of the $N$ categories, is then written as

$$\mathbf{w}_t' = \mathcal{M}'(\tilde{\mathbf{p}}_t', m_t, \theta) + \mathbf{u}_t' \; ; \quad \mathbf{u}_t \sim (0, \mathbf{\Sigma}) \, ; \tag{2.2}$$

where we assume, without loss of generality, that $\underset{n\times 1}{\mathbf{w}_t} = \mathbf{J} \underset{N\times 1}{\tilde{\mathbf{w}}_t}$, where $\underset{n\times N}{\mathbf{J}} = \left[\mathbf{I_n} \;\vdots\; \mathbf{0}\right]$. $\mathbf{u}_t$ and $\mathcal{M}$ are defined analogously.

The further assumption that $\mathbf{u}_t$ is serially independent and distributed $n$-variate normal[3] then implies the conventional log-likelihood[4]

---

[2] Quantities pertaining to the complete $T \times N$ system will be indicated by a "~" over the symbol for the corresponding "full rank" quantity.

[3] The multivariate normal (MVN) assumption is quite standard, even after the transformation to expenditure share form. Other possibilities that take the bounds on the expenditure share disturbance into account are naturally rather less tractable. See, for instance, Fry, Fry and McLaren (1996).

[4] Constants will generally be ignored when writing down likelihood functions.

$$\ell(\theta,\Sigma) = -\frac{T}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{U}'\mathbf{U}), \tag{2.3}$$

where
$$\underset{T\times n}{\mathbf{U}} = \underset{T\times n}{\mathbf{W}} - \mathcal{M}(\underset{T\times N}{\tilde{\mathbf{P}}}, \underset{T\times 1}{\mathbf{m}}, \theta) \tag{2.4}$$

is the $T \times n$ matrix of disturbances, **W** is the $T \times n$ matrix of observed expenditure shares, and $\mathcal{M}(\cdot)$ is the $T \times n$ matrix of expected expenditure shares, conditional on the $T \times N$ matrix of prices $\tilde{\mathbf{P}}$, the $T \times 1$ vector of total expenditures **m**, and the vector of "mean" parameters, θ.

Suppose now that the data is divided according to the differing degrees of expenditure category aggregation in each of $S > 1$ subperiods $\mathcal{T}_1,\ldots,\mathcal{T}_S$. With a slight change in notation, let the *N*-vector of complete, but possibly partially unobserved, expenditure shares[5] be now denoted by $\tilde{\mathbf{x}}$, let $\tilde{\mathbf{w}}$ denote the vector of *observed*, but more highly aggregated, expenditure shares, and note that $\tilde{\mathbf{w}}_t$ is naturally a linear combination of the elements of $\tilde{\mathbf{x}}_t$. It therefore follows (see GM for details) that, after selecting a category common to all subperiods for omission, we have, for the $r^{\text{th}}$ subperiod $\mathcal{T}_r$, $r$ = 1,…,*S*, the full-rank "aggregated" model

$$\mathbf{w}_t' = \mathbf{x}_t'\mathbf{A}_\mathbf{r}' = \mathcal{M}(\tilde{\mathbf{p}}_t', m_t, \theta)\mathbf{A}_\mathbf{r}' + \mathbf{u}_t'\mathbf{A}_\mathbf{r}' \; ; \quad \mathbf{A}_\mathbf{r}\mathbf{u}_t \sim \mathrm{N}(0, \mathbf{A}_\mathbf{r}\boldsymbol{\Sigma}\mathbf{A}_\mathbf{r}') \; ,$$

where $\tilde{\mathcal{M}}(\cdot)$ denotes the expectation of the *complete* *N*-commodity expenditure set $\tilde{\mathbf{x}}$, $\mathbf{x} = \mathbf{J}\,\tilde{\mathbf{x}}$ is the underlying *n*-vector of expenditure shares after exclusion of the common commodity, $\mathbf{w} = \mathbf{A}_\mathbf{r}\mathbf{x}$ is the $n_r$-vector of expenditure shares actually observed in the $r^{\text{th}}$ subperiod (except for the excluded common commodity), and $\mathbf{A}_\mathbf{r}$ is the $n_r \times n$ full row rank "aggregation matrix" taking **x** into **w**. Assuming $T_r$ such observations then yields the $r^{th}$ subperiod log-likelihood:

$$\ell_r = -\frac{T_r}{2}\ln|\boldsymbol{\Sigma}_\mathbf{r}| - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_\mathbf{r}^{-1}\mathbf{U}_\mathbf{r}'\mathbf{U}_\mathbf{r}), \tag{2.5}$$

where
$$\underset{n_r\times n_r}{\boldsymbol{\Sigma}_\mathbf{r}} = \mathbf{A}_\mathbf{r}\boldsymbol{\Sigma}\mathbf{A}_\mathbf{r}' \; ,$$

[5] It will be convenient, in the following, to treat expenditures and expenditure shares interchangeably, assuming throughout that likelihood (2.3) is appropriate in either case. The actual choice of endogenous

$$\underset{T_r \times n_r}{\mathbf{U_r}} = \underset{T_r \times n_r}{\mathbf{W_r}} - \mathcal{M}(\underset{T_r \times N}{\tilde{\mathbf{P}}_\mathbf{r}}, \underset{T_r \times 1}{\mathbf{m_r}}, \theta) \underset{n \times n_r}{\mathbf{A}'_\mathbf{r}}, \tag{2.6}$$

$\mathbf{W_r}$ and $\mathbf{U_r}$ are the $T_r \times n_r$ matrices of observed expenditure shares and their disturbances for the $r^{th}$ subperiod, and $\mathcal{M}(\tilde{\mathbf{P}}_\mathbf{r}, \mathbf{m_r}, \theta) \equiv \mathcal{M}_\mathbf{r}(\theta)$ is the $T_r \times n$ matrix of expected expenditure shares, conditional on the $T_r \times N$ matrix of prices[6] $\tilde{\mathbf{P}}_\mathbf{r}$, the $T_r \times 1$ vector of total expenditures $\mathbf{m_r}$, and the $k$-vector of mean parameters θ. The "aggregated" log-likelihood follows:

$$\ell(\theta, \Sigma) = \sum_{r=1}^{S} \ell_r = -\frac{1}{2}\sum_{r=1}^{S} T_r \ln\left|\mathbf{A_r}\Sigma\mathbf{A}'_\mathbf{r}\right| - \frac{1}{2}\sum_{r=1}^{S} \operatorname{tr}\left(\mathbf{U}'_\mathbf{r}\mathbf{U_r}(\mathbf{A_r}\Sigma\mathbf{A}'_\mathbf{r})^{-1}\right). \tag{2.7}$$

Now, in the case of likelihood (2.3) there is a well-known closed-form expression for the MLE of **Σ**; namely $\hat{\Sigma} = T^{-1}\mathbf{U}'\mathbf{U}$. The FOC[7] for **Σ** in likelihood (2.7), on the other hand, has no such closed form solution; and hence there is no "aggregated" analogue of the usual profile likelihood that might be feasibly maximized with respect to θ. Grose and McLaren dealt with this problem by reparameterizing **Σ** as per De Boer and Harkema (1986), resulting in a feasible estimation procedure at the price of a somewhat restrictive covariance structure. We now consider a method of maximizing (2.7) as it stands.

## 3. "Modelling" the missing data: the EM algorithm.

We are aware that, if expenditure data were observed for all commodity categories in all time periods, then the covariance matrix can be concentrated out of the log-likelihood. This results in an estimation problem that, while still infeasible for "small" sample sizes (such as, it must be noted, the 18 category, 27 annual observations example of Grose and McLaren (1999)), *is* practical in larger samples, such as those that become available if we are willing to use quarterly data. In other words, if we had a complete expenditure set, spanning an adequate timeframe, our model could be

variable makes no difference to the present argument, and thus may as well be reserved for the model-specification stage.

[6] Note that it is assumed that, although expenditure data is not available for all *N* commodities in all time periods, price data is.

estimated without imposing assumptions on the cross-commodity covariance matrix. The main problem with the estimation of such a model when some of the expenditure data are missing is due to our inability to “remove” the covariance matrix from the consequent, more complex, likelihood.

This suggests an alternative strategy based on the idea of *data augmentation* and the expectation-maximization (EM) algorithm. Briefly, this involves using the chosen model to simulate, or *impute* values to, the missing expenditure shares, conditional on given values of the parameters θ and Σ, and the data actually observed. This “completes” the expenditure dataset, enabling maximization of the likelihood, conditional on the imputed data. The log-likelihood function of the parameters given the *observed* data is then just the expectation of the “augmented” log-likelihood, taken with respect to the conditional density of the missing data; and can therefore be estimated by averaging over multiple such imputed datasets.

Let us denote the complete (but only partially observed) expenditure set by **X**. **W** will denote the observed expenditure data, and **Z** the unobserved remainder. That is,

$$\mathbf{X} = (\mathbf{W},\ \mathbf{Z}) = (observed, missing)\,.$$

Also let $\omega = (\theta, \Sigma)$ denote the complete set of unknown parameters.

Obviously, the “complete” log-likelihood $\ell(\omega\,|\mathbf{X})$ is as per (2.3) for a model of the form (2.2) with MVN errors; that is,

$$\ell(\omega\,|\mathbf{X}) = const - \frac{T}{2}\ln|\mathbf{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\mathbf{U}(\theta)\mathbf{\Sigma}^{-1}\mathbf{U}'(\theta)\right); \qquad (3.1)$$

where $\mathbf{U}(\theta) = \mathbf{X} - \mathcal{M}(\theta)$, and expected expenditure $\mathcal{M}(\theta)$ is assumed known[8] for given θ. We note that, in our missing data context, $\ell(\omega\,|\mathbf{X})$ involves not only the

[7] See Grose and McLaren (1999), §3.1.

[8] Strictly speaking, $\mathcal{M}(\theta)$ is the expected expenditure *conditional* on the observed values of any explanatory variables (such as price, and total expenditure). This conditioning is thus implicit in (3.1); allowing us to evade, in the usual manner, the regression aspect of the model. It is not made explicit for reasons of notational convenience and clarity. We note, however, that the requirement that the expectation of **X** be known except for θ *does* imply that the complete $T \times N$ matrix of prices and the *T*-vector of total expenditures must be available.

unknown parameters θ and Σ, but also the unknown data **Z**; and thus could not in general be maximized consistently with respect to **Z** even if this were thought desirable.

Accordingly, let $Q(\omega,\omega^*)$ denote the *expected* log-likelihood of the parameters ω; with the expectation taken with respect to the missing data **Z**, conditional on the observed data **W**, *and* some given $\omega^*$. That is,

$$Q(\omega,\omega^*) = E_{\mathbf{Z}|\mathbf{W},\omega^*}\big(\ell(\omega|\mathbf{X})\big) = \int_{\mathbf{Z}} \ell(\omega|\mathbf{X})\mathcal{P}(\mathbf{Z}|\mathbf{W},\omega^*)d\mathbf{Z}\,; \qquad (3.2)$$

where "$\mathcal{P}$(.)" denotes a probability density function, and $\mathcal{P}(\mathbf{Z}|\mathbf{W},\omega^*)$ is the density of **Z** conditional on **W** and $\omega^*$.

Now, assuming some pre-specified $\omega^*$, $Q(\omega,\omega^*)$ can (in principle) be maximized with respect to ω, yielding a "first approximation" $\hat{\omega}$; which can then be used to refine the conditional density of **Z**. A new expectation $Q(\omega,\hat{\omega})$ and a new maximizing $\hat{\omega}$ follow. In summary, beginning with some plausible starting value $\hat{\omega}^{(0)}$, the iterative EM scheme is based on the construction, and subsequent maximization, of $Q(\omega,\hat{\omega}^{(i)})$ with respect to ω, yielding the next iterate $\hat{\omega}^{(i+1)}$, the corresponding maximum $Q(\hat{\omega}^{(i+1)},\hat{\omega}^{(i)})$, and a "new" criterion $Q(\omega,\hat{\omega}^{(i+1)})$.

It can be shown[9] that $Q(\hat{\omega}^{(i+1)},\hat{\omega}^{(i)}) \geq Q(\hat{\omega}^{(i)},\hat{\omega}^{(i)})$ implies $\ell(\hat{\omega}^{(i+1)}|\mathbf{W}) \geq \ell(\hat{\omega}^{(i)}|\mathbf{W})$; that is, the scheme results in a non-decreasing sequence of log-likelihood values; and that, subject to certain regularity conditions, this sequence converges to a maximum of the likelihood function. Hence the EM algorithm (eventually) maximizes (or at least finds a maximum of) the observed log-likelihood.

The crucial point here is that it be possible to maximize $Q(\omega,\omega^*)$, where it is not possible (or is very difficult) to maximize $\ell(\omega|\mathbf{W})$ directly. The latter, as we noted in §2, is the situation with regard to our "aggregated" log-likelihood if we do not wish to reparameterize Σ, because we cannot concentrate Σ out of the likelihood. Equally, however, maximization of the expected log-likelihood will itself be feasible only if it turns out to be possible to concentrate Σ out of $Q(\omega,\omega^*)$. On this point observe, for

[9] See Dempster, Laird and Rubin (1977) for the proofs of these results in a general setting. A somewhat more accessible summary can be found in McLachlan and Krishnan (1997).

later reference, that for a criterion of the form (3.1) the $n \times n$ matrix of scores with respect to the elements of $\mathbf{\Sigma}$ is

$$\frac{\partial \ell}{\partial \Sigma} = \Sigma^{-1}\{\mathbf{U}'\mathbf{U} - T\Sigma\}\Sigma^{-1} - \frac{1}{2}\mathbf{D}_{\Sigma^{-1}\{\mathbf{U}'\mathbf{U}-T\Sigma\}\Sigma^{-1}}; \tag{3.3}$$

where $\mathbf{D_Q}$ denotes the diagonal matrix with nonzero elements set equal to the diagonal elements of $\mathbf{Q}$. The corresponding first order condition (FOC) for $\mathbf{\Sigma}$ is thus

$$\mathbf{U}'\mathbf{U} - T\Sigma = 0. \tag{3.4}$$

## 3.1 The conditional density of the missing data

Suppose, for any period *t*, the complete expenditure model is written

$$\underset{n\times 1}{\mathbf{x}_t}|\omega \sim \mathrm{N}(\mu_t(\theta), \underset{n\times n}{\Sigma}); \tag{3.5}$$

where, in the $r^{th}$ subperiod, only $\underset{n_r\times 1}{\mathbf{w}_t} = \mathbf{A_r} \underset{n\times 1}{\mathbf{x}_t}$, $n_r \le n$, is observed.

For convenience dropping the time subscript, partition the *n*-vector $\mathbf{x}$ as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \begin{matrix} m_r\times 1 \\ n_r\times 1 \end{matrix} \sim \mathrm{N}\left( \begin{pmatrix} \mu_1(\theta) \\ \mu_2(\theta) \end{pmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

where the commodities are ordered such that the $m_r = (n - n_r)$-vector $\mathbf{x}_1$ equals the unknown, or "missing", expenditure set $\mathbf{z}$. The remainder $\mathbf{x}_2$, though not directly observed, could be determined from knowledge of $\mathbf{x}_1$ and $\mathbf{w}$. The $r^{th}$ subperiod transformation from $\mathbf{x}$ to ($\mathbf{z}$, $\mathbf{w}$),

$$\underset{n\times 1}{\mathbf{x}} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \to \begin{pmatrix} \mathbf{z} \\ \mathbf{w} \end{pmatrix} \begin{matrix} m_r\times 1 \\ n_r\times 1 \end{matrix}$$

is therefore defined by $\mathbf{z} = \mathbf{x}_1 = \mathbf{J_r x}$ and $\mathbf{w} = \mathbf{A_r x}$; where $\mathbf{J_r}$ denotes the $m_r \times n$ matrix $\begin{bmatrix} \mathbf{I}_{\mathbf{m}_r} & \mathbf{0} \end{bmatrix}$. We thus have

$$\begin{pmatrix} \mathbf{z} \\ \mathbf{w} \end{pmatrix} = \mathbf{B_r x}, \text{ where } \mathbf{B_r} = \begin{bmatrix} \mathbf{J_r} \\ \mathbf{A_r} \end{bmatrix},$$

implying

$$\begin{pmatrix} \mathbf{z} \\ \mathbf{w} \end{pmatrix} \sim \mathrm{N}\left(\mathbf{B_r}\mu , \mathbf{B_r}\Sigma\mathbf{B_r'}\right) \equiv \mathrm{N}\left(\begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}\right),$$

with $\upsilon_1 = \mathbf{J_r}\mu \equiv \mu_1$, $\upsilon_2 = \mathbf{A_r}\mu$, $\Omega_{11} = \mathbf{J_r}\Sigma\mathbf{J_r'} \equiv \Sigma_{11}$, $\Omega_{12} = \mathbf{J_r}\Sigma\mathbf{A_r'}$, and $\Omega_{22} = \mathbf{A_r}\Sigma\mathbf{A_r'} \equiv \Sigma_{\mathbf{r}}$. The standard result regarding the conditional distribution of a subvector of normal variates then instantly yields

$$\mathbf{z} \,|\, \mathbf{w},\omega \sim \mathrm{N}\left(\upsilon_1 + \Omega_{12}\Omega_{22}^{-1}(\mathbf{w} - \upsilon_2)\,,\; \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right),$$

where $\left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1} = \left(\Omega^{-1}\right)_{11}$ is the top-left $m_r \times m_r$ partition of $\Omega^{-1} = \mathbf{B_r}^{-1\prime}\Sigma^{-1}\mathbf{B_r}^{-1}$. Hence, for $t \in \mathcal{T}_r$,

$$\mathbf{z}_t \,|\, \mathbf{w}_t,\omega \sim \mathrm{N}\left(\mu_{\mathbf{z}_t|\mathbf{w}_t}, \Sigma_{\mathbf{z|w}}\right), \tag{3.6}$$

where $\mu_{\mathbf{z}_t|\mathbf{w}_t} = \mathbf{J_r}\mu_t + \mathbf{J_r}\Sigma\mathbf{A_r'}(\mathbf{A_r}\Sigma\mathbf{A_r'})^{-1}(\mathbf{w}_t - \mathbf{A_r}\mu_t)$ is the conditional mean of $\mathbf{z}_t$, the $m_r \times m_r$ matrix $\Sigma_{\mathbf{z|w}} = \mathbf{J_r}\Sigma\mathbf{J_r'} + \mathbf{J_r}\Sigma\mathbf{A_r'}(\mathbf{A_r}\Sigma\mathbf{A_r'})^{-1}\mathbf{A_r}\Sigma\mathbf{J_r'}$ is its conditional variance, and, for the model of §2 $\mu_t \equiv \mathcal{M}(\tilde{\mathbf{p}}_t', m_t, \theta)$. The density of the $T_r \times m_r$ matrix of missing data $\mathbf{Z_r}$, conditional on the $T_r \times n_r$ matrix of observed data $\mathbf{W_r}$ and $\omega = (\theta, \Sigma)$, then follows:

$$\begin{aligned}
\mathcal{P}(\mathbf{Z_r} \,|\, \mathbf{W_r}, \omega) &= \prod_{t \in \mathcal{T}_r} \mathcal{P}(\mathbf{z}_t \,|\, \mathbf{w}_t, \omega) \\
&= (2\pi)^{-m_r T_r/2} \left| \Sigma_{\mathbf{z|w}}^{-1} \right|^{T_r/2} \exp\left\{ -\frac{1}{2} \sum_{t \in \mathcal{T}_r} (\mathbf{z}_t - \mu_{\mathbf{z}_t|\mathbf{w}_t})' \, \Sigma_{\mathbf{z|w}}^{-1} (\mathbf{z}_t - \mu_{\mathbf{z}_t|\mathbf{w}_t}) \right\} \\
&= (2\pi)^{-m_r T_r/2} \left| \Sigma_{\mathbf{z|w}}^{-1} \right|^{T_r/2} \exp\left\{ -\frac{1}{2} \mathrm{tr}\left\{ \left(\mathbf{Z_r} - \mathcal{M}_{\mathbf{Z_r}|\mathbf{W_r}}\right) \Sigma_{\mathbf{z|w}}^{-1} \left(\mathbf{Z_r} - \mathcal{M}_{\mathbf{Z_r}|\mathbf{W_r}}\right)' \right\} \right\}
\end{aligned}$$

where $\mathcal{M}_{\mathbf{Z_r}|\mathbf{W_r}}$ is the $T_r \times m_r$ matrix with $t^{\text{th}}$ row equal to $\mu'_{\mathbf{z}_t|\mathbf{w}_t}$.

### 3.2 MCEM

Simulation of the missing data in each subperiod is thus quite straightforward for given $\mathbf{W_r}$, $r = 1,\ldots,S$, and $\omega^* = (\theta^*, \Sigma^*)$. Having obtained an imputed $\mathbf{Z_r}$ (denoted $\mathbf{Z}_\mathbf{r}^*$ to emphasize its dependence on $\omega^*$), the augmented $T_r \times n$ expenditure set for $t \in \mathcal{T}_r$, $\mathbf{X}_\mathbf{r}^*$, then follows via the reverse transform

$$\mathbf{x}_t = \mathbf{B}_\mathbf{r}^{-1}\begin{pmatrix}\mathbf{z}_t \\ \mathbf{w}_t\end{pmatrix} \quad \Rightarrow \quad \underset{T_r\times n}{\mathbf{X}_\mathbf{r}^*} = \left[\underset{T_r\times m_r}{\mathbf{Z}_\mathbf{r}^*} \quad \underset{T_r\times n_r}{\mathbf{W}_\mathbf{r}}\right]\mathbf{B}_\mathbf{r}^{-1\prime},\ r = 1,\ldots,S.$$

Finally, combining $\mathbf{X}_\mathbf{1}^*,\ \ldots,\ \mathbf{X}_\mathbf{S}^*$ yields the corresponding "realisation", $\mathbf{X}^*$, of the complete $T \times n$ expenditure set.

The expected log-likelihood $Q(\omega,\omega^*)$ can now be estimated by the *average* log-likelihood, taken over $M$ such independently realised augmented datasets $\mathbf{X}_1^*,\ldots,\mathbf{X}_M^*$,

$$\overline{Q}(\omega,\omega^*) = \frac{1}{M}\sum_{j=1}^{M}\ell(\omega|\mathbf{X}_j^*);$$

leading to the Monte-Carlo EM (MCEM) algorithm as follows.

For $i = 0,1,2,\ldots$, some starter $\hat{\omega}^{(0)} = (\hat{\theta}^{(0)},\hat{\Sigma}^{(0)})$, and prespecified tolerance $\varepsilon$ :

(1) Draw $\mathbf{Z}_{\mathbf{r}1},\mathbf{Z}_{\mathbf{r}2},\ldots,\mathbf{Z}_{\mathbf{r}M} \overset{iid}{\sim} \mathcal{P}(\mathbf{Z_r}|\mathbf{W_r},\hat{\omega}^{(i)})$ for each subperiod $r = 1,2,\ldots,S$.

(2) Combine $\mathbf{Z}_{\mathbf{r}1},\mathbf{Z}_{\mathbf{r}2},\ldots,\mathbf{Z}_{\mathbf{r}M}$ with $\mathbf{W_r}$ to generate $\mathbf{X}_{\mathbf{r}1},\mathbf{X}_{\mathbf{r}2},\ldots,\mathbf{X}_{\mathbf{r}M}$, and hence $M$ realizations of the augmented expenditure data, $\mathbf{X}_1^*,\mathbf{X}_2^*,\ldots,\mathbf{X}_M^*$.

(3) Maximize the average log-likelihood $\overline{Q}(\omega,\hat{\omega}^{(i)})$ with respect to $\omega$, so obtaining $\hat{\omega}^{(i+1)}$.

(4) Repeat (1) – (3) until $\left|\overline{Q}(\hat{\omega}^{(i+1)},\hat{\omega}^{(i)}) - \overline{Q}(\hat{\omega}^{(i)},\hat{\omega}^{(i)})\right| < \varepsilon$.

Notice, however, that

$$\frac{\partial\overline{Q}(\omega,\omega^*)}{\partial\Sigma} = \frac{1}{M}\sum_{j=1}^{M}\frac{\partial\ell(\omega|\mathbf{X}_j^*)}{\partial\Sigma};$$

where $\ell(\omega|\mathbf{X}_j^*)$ is as per (3.1). The required derivative is thus of the form (3.3), with **U** replaced by $\mathbf{U}_j = \mathbf{X}_j^* - \mathcal{M}(\theta)$, implying

$$\frac{\partial \overline{Q}(\omega,\omega^*)}{\partial \Sigma} = \frac{1}{M}\sum_{j=1}^{M}\left(\Sigma^{-1}\left\{\mathbf{U}_j'\mathbf{U}_j - T\Sigma\right\}\Sigma^{-1}\right) - \frac{1}{2M}\sum_{j=1}^{M}\mathbf{D}_{\Sigma^{-1}\left\{\mathbf{U}_j'\mathbf{U}_j - T\Sigma\right\}\Sigma^{-1}}$$
$$= \Sigma^{-1}\left\{\overline{\mathbf{S}}^* - T\Sigma\right\}\Sigma^{-1} - \frac{1}{2}\mathbf{D}_{\Sigma^{-1}\left\{\overline{\mathbf{S}}^* - T\Sigma\right\}\Sigma^{-1}}$$

where $$\overline{\mathbf{S}}^* = \frac{1}{M}\sum_{j=1}^{M}\mathbf{U}_j'\mathbf{U}_j \; .$$

Thus $\dfrac{\partial \overline{Q}(\omega,\omega^*)}{\partial \Sigma} = 0$ yields $\hat{\Sigma} = \overline{\mathbf{S}}^* / T$; implying the "concentrated" criterion

$$\overline{Q}_c(\theta,\omega^*) = \frac{1}{M}\sum_{j=1}^{M}\ell_c(\theta|\mathbf{X}_j^*) = const - \frac{T}{2}\ln\left|\overline{\mathbf{S}}^*(\theta)\right|. \qquad (3.7)$$

Consequently, Step (3), which we recall in our context is likely to be infeasible, would now be replaced by maximization of $\overline{Q}_c(\theta,\hat{\omega}^{(i)})$ (or, equivalently, minimisation of $\ln\left|\overline{\mathbf{S}}^*(\theta)\right|$) with respect to θ, yielding $\hat{\theta}^{(i+1)}$, $\overline{\mathbf{S}}^{(i+1)}$, and hence $\hat{\Sigma}^{(i+1)}$.

## 4. EM: performing the E-step analytically.

MCEM is thus feasible, because it allows us to concentrate **Σ** out of the criterion, and relatively simple to implement. However, it is, as it happens, extremely time-consuming, due to the very large[10] number of imputation draws needed in the MC expectation stage to achieve convergence with an acceptable degree of precision. We would clearly prefer an analytic expression for the expected log-likelihood.

We accordingly require a closed form solution for (3.2), where $\ell(\omega\,|\,\mathbf{X})$ is given by (3.1), and

$$\mathcal{P}(\mathbf{Z}\,|\,\mathbf{W},\omega) = \prod_{r=1}^{S}\mathcal{P}(\mathbf{Z}_\mathbf{r}\,|\,\mathbf{W}_\mathbf{r},\omega).$$

## 4.1 "Vectorizing" the conditional density

To rewrite the conditional density in a form more suited to our purpose, let us define the "*vecr*" operator[11], and the accompanying notation:

$$\underset{nT\times 1}{\underset{\sim}{\mathbf{u}}} \triangleq \text{vecr}(\mathbf{U}) = \text{vec}(\mathbf{U}') \,;$$

so that $\text{tr}\left(\mathbf{U}\Sigma^{-1}\mathbf{U}'\right) \equiv \underset{\sim}{\mathbf{u}}' \left\lfloor \mathbf{I_T} \otimes \Sigma^{-1} \right\rfloor \underset{\sim}{\mathbf{u}}$ . Similarly defining $\underset{\sim}{\mathbf{x}} = \text{vecr}(\mathbf{X})$ and $\underset{\sim}{\mu}_{\mathbf{x}}(\theta) = \text{vecr}(\mathcal{M}(\theta))$ ; so that $\underset{\sim}{\mathbf{u}} = \underset{\sim}{\mathbf{x}} - \underset{\sim}{\mu}_{\mathbf{x}}(\theta)$ , then implies the "complete" model

$$\mathcal{P}(\underset{\sim}{\mathbf{x}}\,|\,\omega) = (2\pi)^{-nT/2}\ |\Sigma|^{-T/2}\ \exp\left\{-\frac{1}{2}(\underset{\sim}{\mathbf{x}} - \underset{\sim}{\mu}_{\mathbf{x}}(\theta))'\left(\mathbf{I_T} \otimes \Sigma^{-1}\right)(\underset{\sim}{\mathbf{x}} - \underset{\sim}{\mu}_{\mathbf{x}}(\theta))\right\}. \quad (4.1)$$

We further define

$$\underset{\sim}{\mathbf{w}} = \begin{pmatrix} \underset{\sim}{\mathbf{w}}_1 \\ \vdots \\ \underset{\sim}{\mathbf{w}}_S \end{pmatrix} = \begin{pmatrix} \text{vecr}(\mathbf{W}_1) \\ \vdots \\ \text{vecr}(\mathbf{W}_S) \end{pmatrix}\begin{matrix} {\scriptstyle n_1T_1\times 1} \\ \\ {\scriptstyle n_ST_S\times 1} \end{matrix} \quad \text{and} \quad \underset{\sim}{\mathbf{z}} = \begin{pmatrix} \underset{\sim}{\mathbf{z}}_1 \\ \vdots \\ \underset{\sim}{\mathbf{z}}_S \end{pmatrix} = \begin{pmatrix} \text{vecr}(\mathbf{Z}_1) \\ \vdots \\ \text{vecr}(\mathbf{Z}_S) \end{pmatrix}\begin{matrix} {\scriptstyle m_1T_1\times 1} \\ \\ {\scriptstyle m_ST_S\times 1} \end{matrix} \,;$$

so that the division of **X** into "observed" and "missing" (**W**, **Z**) is equivalent to the transformation

$$\underset{nT\times 1}{\underset{\sim}{\mathbf{x}}} \rightarrow \begin{pmatrix} \underset{\sim}{\mathbf{z}} \\ \underset{\sim}{\mathbf{w}} \end{pmatrix}\begin{matrix} {\scriptstyle \sum_{r=1}^{S} m_rT_r\times 1} \\ {\scriptstyle \sum_{r=1}^{S} n_rT_r\times 1} \end{matrix} . \quad (4.2)$$

Transformation (4.2) is necessarily accomplished in two stages. The first of these is "aggregation"; that is,

---

[10] In the model under consideration at least $10^4$ imputation draws in each iteration were needed to achieve an adequately accurate MC expectation.

[11] The "observation-by-observation" operator "vecr" is here preferred to the more commonly employed "vec" as it allows us to extrapolate easily from any given subperiod *r* to the complete sample.

$$\underset{\sim}{\mathbf{x}}^{\ddagger} = \tilde{\mathbf{B}}\,\underset{\sim}{\mathbf{x}}\,,$$

where $$\underset{nT\times nT}{\tilde{\mathbf{B}}} = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1}\otimes\mathbf{B_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{I}_{\mathbf{T}_S}\otimes\mathbf{B_S} \end{bmatrix}.$$

$\tilde{\mathbf{B}}\,\underset{\sim}{\mathbf{x}}$ yields a $nT$-vector $\underset{\sim}{\mathbf{x}}^{\ddagger}$ comprised explicitly of the "observed" and "missing" data; but still co-mingled together, as we see below. The second stage, from $\underset{\sim}{\mathbf{x}}^{\ddagger}$ to $\underset{\sim}{\mathbf{x}}^{\dagger}$, therefore requires a permutation operation, in which the missing data is moved to the "top" of the vector. This is accomplished by the operation

$$\underset{\sim}{\mathbf{x}}^{\dagger} = \tilde{\mathbf{P}}\,\underset{\sim}{\mathbf{x}}^{\ddagger} = \begin{pmatrix} \underset{\sim}{\mathbf{z}} \\ \underset{\sim}{\mathbf{w}} \end{pmatrix};$$

where $$\underset{nT\times nT}{\tilde{\mathbf{P}}} = \begin{pmatrix} \tilde{\mathbf{P}}_{\mathbf{z}} \\ \tilde{\mathbf{P}}_{\mathbf{w}} \end{pmatrix} \begin{matrix} \sum_{r=1}^{S} m_r T_r \times nT \\ \sum_{r=1}^{S} n_r T_r \times nT \end{matrix},$$

$$\tilde{\mathbf{P}}_{\mathbf{z}} = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1}\otimes\mathbf{J_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{I}_{\mathbf{T}_S}\otimes\mathbf{J_S} \end{bmatrix},$$

and $$\tilde{\mathbf{P}}_{\mathbf{w}} = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1}\otimes\bar{\mathbf{J}}_{\mathbf{1}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{I}_{\mathbf{T}_S}\otimes\bar{\mathbf{J}}_{\mathbf{S}} \end{bmatrix},$$

in which $\mathbf{J_r}$ and $\bar{\mathbf{J}}_{\mathbf{r}}$ denote the $m_r \times n$ and $n_r \times n$ "selection matrices" $\mathbf{J_r} = \begin{bmatrix}\mathbf{I}_{m_r} & \mathbf{0}\end{bmatrix}$ and $\bar{\mathbf{J}}_{\mathbf{r}} = \begin{bmatrix}\mathbf{0} & \mathbf{I}_{n_r}\end{bmatrix}$, respectively, $r = 1,\ldots,S$.

It can be confirmed, by direct computation, that $\underset{\sim}{\mathbf{x}}^{\dagger} = \tilde{\mathbf{P}}\,\underset{\sim}{\mathbf{x}}^{\ddagger} = \tilde{\mathbf{P}}\tilde{\mathbf{B}}\,\underset{\sim}{\mathbf{x}}$ does indeed equal $(\underset{\sim}{\mathbf{z}}'\;\;\underset{\sim}{\mathbf{w}}')'$. That is, with $\underset{\sim}{\mathbf{x}} = \text{vecr}(\underset{T\times n}{\mathbf{X}})$, $\underset{\sim}{\mathbf{x}}^{\ddagger} = \tilde{\mathbf{B}}\,\underset{\sim}{\mathbf{x}}$, and $\underset{\sim}{\mathbf{x}}^{\dagger} = \tilde{\mathbf{P}}\,\underset{\sim}{\mathbf{x}}^{\ddagger}$, we have

$$
\underset{\sim}{\mathbf{x}} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{T_1} \\ \mathbf{x}_{T_1+1} \\ \vdots \\ \mathbf{x}_{T_1+T_2} \\ \vdots \\ \mathbf{x}_{1+\Sigma_r^{S-1}T_r} \\ \vdots \\ \mathbf{x}_T \end{pmatrix} \begin{matrix} n\times 1 \\ \\ n\times 1 \\ n\times 1 \\ \\ n\times 1 \\ \\ n\times 1 \\ \\ n\times 1 \end{matrix}, \quad \underset{\sim}{\mathbf{x}}^{\ddagger} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{w}_1 \\ \vdots \\ \mathbf{z}_{T_1} \\ \mathbf{w}_{T_1} \\ \mathbf{z}_{T_1+1} \\ \mathbf{w}_{T_1+1} \\ \vdots \\ \mathbf{z}_{T_1+T_2} \\ \mathbf{w}_{T_1+T_2} \\ \vdots \\ \mathbf{z}_{1+\Sigma_r^{S-1}T_r} \\ \mathbf{w}_{1+\Sigma_r^{S-1}T_r} \\ \vdots \\ \mathbf{z}_T \\ \mathbf{w}_T \end{pmatrix} \begin{matrix} m_1\times 1 \\ n_1\times 1 \\ \\ m_1\times 1 \\ n_1\times 1 \\ m_2\times 1 \\ n_2\times 1 \\ \\ m_2\times 1 \\ n_2\times 1 \\ \\ m_S\times 1 \\ n_S\times 1 \\ \\ m_S\times 1 \\ n_S\times 1 \end{matrix}, \quad \textit{and} \quad \underset{\sim}{\mathbf{x}}^{\dagger} = \begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_{T_1} \\ \mathbf{z}_{T_1+1} \\ \vdots \\ \mathbf{z}_{T_1+T_2} \\ \vdots \\ \mathbf{z}_{1+\Sigma_r^{S-1}T_r} \\ \vdots \\ \mathbf{z}_T \\ \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_{T_1} \\ \mathbf{w}_{T_1+1} \\ \vdots \\ \mathbf{w}_{T_1+T_2} \\ \vdots \\ \mathbf{w}_{1+\Sigma_r^{S-1}T_r} \\ \vdots \\ \mathbf{w}_T \end{pmatrix} \begin{matrix} m_1\times 1 \\ \\ m_1\times 1 \\ m_2\times 1 \\ \\ m_2\times 1 \\ \\ m_S\times 1 \\ \\ m_S\times 1 \\ n_1\times 1 \\ \\ n_1\times 1 \\ n_2\times 1 \\ \\ n_2\times 1 \\ \\ n_S\times 1 \\ \\ n_S\times 1 \end{matrix} = \begin{pmatrix} \underset{\sim}{\mathbf{z}}_1 \\ \underset{\sim}{\mathbf{z}}_2 \\ \vdots \\ \underset{\sim}{\mathbf{z}}_S \\ \underset{\sim}{\mathbf{w}}_1 \\ \underset{\sim}{\mathbf{w}}_2 \\ \vdots \\ \underset{\sim}{\mathbf{w}}_S \end{pmatrix}.
$$

Now, $\widetilde{\mathbf{P}}$ is clearly $nT \times nT$ orthogonal, implying $\left|\widetilde{\mathbf{P}}\right| = 1$. Furthermore, it is generally possible to order our commodities so that $\mathbf{A_r} = \left[\mathbf{L_r} \quad \mathbf{I_{n_r}}\right] \;\forall\; r = 1,\ldots,S$ with $\mathbf{L_r}$ an $n_r \times m_r$ matrix of ones and zeros. It therefore follows that

$$
\mathbf{B_r} = \begin{bmatrix} \mathbf{I_{m_r}} & \mathbf{0} \\ \mathbf{L_r} & \mathbf{I_{n_r}} \end{bmatrix}, \; \mathbf{B_r}^{-1} = \begin{bmatrix} \mathbf{I_{m_r}} & \mathbf{0} \\ -\mathbf{L_r} & \mathbf{I_{n_r}} \end{bmatrix}, \text{ and } \left|\mathbf{B_r}\right| = 1,
$$

implying $\left|\widetilde{\mathbf{B}}\right| = 1$ also. The density of $\underset{\sim}{\mathbf{x}}^{\dagger} \triangleq \begin{pmatrix} \underset{\sim}{\mathbf{z}} \\ \underset{\sim}{\mathbf{w}} \end{pmatrix} = \widetilde{\mathbf{P}}\widetilde{\mathbf{B}}\underset{\sim}{\mathbf{x}}$ is therefore simply

$$
\mathcal{P}(\underset{\sim}{\mathbf{x}}^{\dagger}|\omega) = (2\pi)^{-nT/2}\; |\Sigma|^{-T/2}\; \exp\left\{-\frac{1}{2}(\underset{\sim}{\mathbf{x}}^{\dagger} - \underset{\sim}{\mu}_{\mathbf{x}}^{\dagger}(\theta))'\widetilde{\Omega}^{-1}(\underset{\sim}{\mathbf{x}}^{\dagger} - \underset{\sim}{\mu}_{\mathbf{x}}^{\dagger}(\theta))\right\}, \qquad (4.3)
$$

with mean $$\underset{\sim}{\mu}_{\mathbf{x}}^{\dagger} = \tilde{\mathbf{P}}\tilde{\mathbf{B}}\underset{\sim}{\mu}_{\mathbf{x}} \equiv \begin{pmatrix} \underset{\sim}{\mu}_{\mathbf{z}} \\ \underset{\sim}{\mu}_{\mathbf{w}} \end{pmatrix}$$

and variance $$\tilde{\Omega} = \tilde{\mathbf{P}}\tilde{\mathbf{B}}\left[\mathbf{I}_{\mathbf{T}} \otimes \Sigma\right]\tilde{\mathbf{B}}'\tilde{\mathbf{P}}' = \begin{bmatrix} \tilde{\Omega}_{\mathbf{zz}} & \tilde{\Omega}_{\mathbf{zw}} \\ \tilde{\Omega}_{\mathbf{wz}} & \tilde{\Omega}_{\mathbf{ww}} \end{bmatrix}.$$

The conditional pdf of the missing data, $\mathcal{P}(\mathbf{Z}|\mathbf{W},\omega)$, is now just

$$\mathcal{P}(\underset{\sim}{\mathbf{z}}\,|\,\underset{\sim}{\mathbf{w}},\omega) = (2\pi)^{-\frac{1}{2}\sum_{r}^{S} m_r T_r} \left|\left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}}\right|^{1/2} \exp\left\{-\frac{1}{2}(\underset{\sim}{\mathbf{z}} - \underset{\sim}{\mu}_{\mathbf{z|w}})' \left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}} (\underset{\sim}{\mathbf{z}} - \underset{\sim}{\mu}_{\mathbf{z|w}})\right\}, \quad (4.4)$$

where $$\underset{\sim}{\mu}_{\mathbf{z|w}} = \underset{\sim}{\mu}_{\mathbf{z|w}}(\omega) = \underset{\sim}{\mu}_{\mathbf{z}}(\theta) + \tilde{\Omega}_{\mathbf{zw}}\tilde{\Omega}_{\mathbf{ww}}^{-1}(\underset{\sim}{\mathbf{w}} - \underset{\sim}{\mu}_{\mathbf{w}}(\theta)) \quad (4.5)$$

is the conditional mean of $\underset{\sim}{\mathbf{z}}$ given $\underset{\sim}{\mathbf{w}}$ and $\omega = (\theta, \Sigma)$, and

$$\left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}} = \left(\tilde{\Omega}_{\mathbf{zz}} - \tilde{\Omega}_{\mathbf{zw}}\,\tilde{\Omega}_{\mathbf{ww}}^{-1}\,\tilde{\Omega}_{\mathbf{wz}}\right)^{-1}$$

is its conditional variance.

On the computational side, since $\tilde{\mathbf{P}}^{-1} = \tilde{\mathbf{P}}'$ it follows that

$$\tilde{\Omega}^{-1} = \tilde{\mathbf{P}}\tilde{\mathbf{B}}'^{-1}\left[\mathbf{I}_{\mathbf{T}} \otimes \Sigma^{-1}\right]\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{P}}' = \left[\begin{array}{c:c} \tilde{\mathbf{P}}_{\mathbf{z}}\tilde{\mathbf{B}}'^{-1}\left[\mathbf{I}_{\mathbf{T}} \otimes \Sigma^{-1}\right]\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{P}}_{\mathbf{z}}' & \tilde{\mathbf{P}}_{\mathbf{z}}\tilde{\mathbf{B}}'^{-1}\left[\mathbf{I}_{\mathbf{T}} \otimes \Sigma^{-1}\right]\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{P}}_{\mathbf{w}}' \\ \hdashline \tilde{\mathbf{P}}_{\mathbf{w}}\tilde{\mathbf{B}}'^{-1}\left[\mathbf{I}_{\mathbf{T}} \otimes \Sigma^{-1}\right]\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{P}}_{\mathbf{z}}' & \tilde{\mathbf{P}}_{\mathbf{w}}\tilde{\mathbf{B}}'^{-1}\left[\mathbf{I}_{\mathbf{T}} \otimes \Sigma^{-1}\right]\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{P}}_{\mathbf{w}}' \end{array}\right];$$

implying $$\left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}} = \tilde{\mathbf{P}}_{\mathbf{z}}\tilde{\mathbf{B}}'^{-1}\left[\mathbf{I}_{\mathbf{T}} \otimes \Sigma^{-1}\right]\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{P}}_{\mathbf{z}}'.$$

The latter simplifies, on partitioning $\mathbf{B}_{\mathbf{r}}^{-1}$ as $\left[\overline{\mathbf{L}}_{\mathbf{r}}' \quad \overline{\mathbf{J}}_{\mathbf{r}}'\right]$ with $\overline{\mathbf{L}}_{\mathbf{r}} = \left[\mathbf{I}_{\mathrm{m_r}} \quad -\mathbf{L}_{\mathbf{r}}'\right]$, as the block-diagonal matrix

$$\left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}} = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1} \otimes \overline{\mathbf{L}}_{\mathbf{1}}\Sigma^{-1}\overline{\mathbf{L}}_{\mathbf{1}}' & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{I}_{\mathbf{T}_S} \otimes \overline{\mathbf{L}}_{\mathbf{S}}\Sigma^{-1}\overline{\mathbf{L}}_{\mathbf{S}}' \end{bmatrix}. \quad (4.6)$$

Similarly, substituting $\underset{\sim}{\mu}_{\mathbf{z}} = \tilde{\mathbf{P}}_{\mathbf{z}}\tilde{\mathbf{B}}\,\underset{\sim}{\mu}_{\mathbf{x}} \equiv \tilde{\mathbf{P}}_{\mathbf{z}}\underset{\sim}{\mu}_{\mathbf{x}}$ and $\underset{\sim}{\mu}_{\mathbf{w}} = \tilde{\mathbf{P}}_{\mathbf{w}}\tilde{\mathbf{B}}\,\underset{\sim}{\mu}_{\mathbf{x}}$ into (4.5) and exploiting the usual partition of $\mathbf{B_r}$ (see §3.1) yields

$$\underset{\sim}{\mu}_{\mathbf{z}|\mathbf{w}} = \tilde{\mathbf{D}}(\Sigma)\,\underset{\sim}{\mu}_{\mathbf{x}}(\theta) \ + \ \tilde{\mathbf{G}}(\Sigma)\,\underset{\sim}{\mathbf{w}}\,;$$

where

$$\tilde{\mathbf{G}}(\Sigma) = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1} \otimes \ \mathbf{J}_1\Sigma\mathbf{A}_1'(\mathbf{A}_1\Sigma\mathbf{A}_1')^{-1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{I}_{\mathbf{T}_S} \otimes \ \mathbf{J}_S\Sigma\mathbf{A}_S'(\mathbf{A}_S\Sigma\mathbf{A}_S')^{-1} \end{bmatrix},$$

and

$$\tilde{\mathbf{D}}(\Sigma) = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1} \otimes \left(\mathbf{J}_1 - \ \mathbf{J}_1\Sigma\mathbf{A}_1'(\mathbf{A}_1\Sigma\mathbf{A}_1')^{-1}\mathbf{A}_1\right) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{I}_{\mathbf{T}_S} \otimes \left(\mathbf{J}_S - \ \mathbf{J}_S\Sigma\mathbf{A}_S'(\mathbf{A}_S\Sigma\mathbf{A}_S')^{-1}\mathbf{A}_S\right) \end{bmatrix}.$$

## 4.2 E-step: the expected log-likelihood

We are now in a position to establish the following closed-form solution for $Q\,(\omega,\omega^*)$.

PROPOSITION 1. Let $\underset{\sim}{\mathbf{x}}_a^*$ denote the "mean-augmented" expenditure set

$$\underset{\sim}{\mathbf{x}}_a^* \equiv \underset{\sim}{\mathbf{x}}_a(\omega^*) = \begin{pmatrix} \underset{\sim}{\mu}_{\mathbf{z}|\mathbf{w}}^* \\ \underset{\sim}{\mathbf{w}} \end{pmatrix},$$

where $$\underset{\sim}{\mu}_{\mathbf{z}|\mathbf{w}}^* = \underset{\sim}{\mu}_{\mathbf{z}|\mathbf{w}}(\omega^*) = \tilde{\mathbf{D}}(\Sigma^*)\,\underset{\sim}{\mu}_{\mathbf{x}}(\theta^*) \ + \ \tilde{\mathbf{G}}(\Sigma^*)\,\underset{\sim}{\mathbf{w}}\,.$$

Also let $\underset{\sim}{\mathbf{u}}_a^*$ denote the vector of "augmented residuals"

$$\underset{\sim}{\mathbf{u}}_a^* = (\tilde{\mathbf{P}}\tilde{\mathbf{B}})^{-1}(\underset{\sim}{\mathbf{x}}_a^* - \underset{\sim}{\mu}_{\mathbf{x}}^\dagger) = (\tilde{\mathbf{P}}\tilde{\mathbf{B}})^{-1}\underset{\sim}{\mathbf{x}}_a^* \ - \underset{\sim}{\mu}_{\mathbf{x}}(\theta)\,,$$

and let $\mathbf{U}^*(\theta)$ denote the $T \times n$ matrix constructed such that $vecr(\mathbf{U}^*) = \underset{\sim}{\mathbf{u}}_a^*$ .

Finally, let $\mathbf{U}^{*\prime}\mathbf{U}^{*} = \mathbf{S}^{*}(\theta)$ and $\overline{\mathbf{L}}'_{\mathbf{r}}\left(\overline{\mathbf{L}}_{\mathbf{r}}\Sigma^{*-1}\overline{\mathbf{L}}'_{\mathbf{r}}\right)^{-1}\overline{\mathbf{L}}_{\mathbf{r}} = \mathbf{V}^{*}_{\mathbf{r}}$. Then the EM criterion for the problem at hand is

$$Q(\omega, \omega^{*}) \;=\; -\frac{T}{2}\ln|\Sigma| \;-\; \frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}\widetilde{\mathbf{S}}^{*}(\theta)\right\}; \tag{4.7}$$

where $$\widetilde{\mathbf{S}}^{*}(\theta) \;=\; \mathbf{S}^{*}(\theta) \;+\; \sum_{r=1}^{S} T_r \mathbf{V}^{*}_{\mathbf{r}} \,.$$

Proof of the proposition requires the following preliminary Lemma.

LEMMA 1. Let $\mathbf{x}$ denote an $n$-vector of independent variables partitioned as $\begin{pmatrix}\mathbf{z}\\ \mathbf{w}\end{pmatrix}$, where $\mathbf{z}$ is $m \times 1$, $m < n$. Let the $n$-vector $\mathbf{a}$ and the $n \times n$ symmetric matrix $\mathbf{A}$ be partitioned comformably; ie; $\mathbf{a} = \begin{pmatrix}\mathbf{a}_{\mathbf{z}}\\ \mathbf{a}_{\mathbf{w}}\end{pmatrix}$, and $\mathbf{A} = \begin{bmatrix}\mathbf{A}_{\mathbf{zz}} & \mathbf{A}_{\mathbf{zw}}\\ \mathbf{A}_{\mathbf{wz}} & \mathbf{A}_{\mathbf{ww}}\end{bmatrix}$. Then, for $\mathbf{b}$ an $m$-vector of constants, and matrix $\mathbf{B}$ $m \times m$ symmetric positive definite:

$$\int\cdots\int(\mathbf{x}-\mathbf{a})'\mathbf{A}(\mathbf{x}-\mathbf{a})e^{-(\mathbf{z}-\mathbf{b})'\mathbf{B}(\mathbf{z}-\mathbf{b})/2}dz_1\cdots dz_n$$
$$= (2\pi)^{m/2}|\mathbf{B}|^{-1/2}\left\{\mathrm{tr}\left(\mathbf{A}_{\mathbf{zz}}\mathbf{B}^{-1}\right) + (\mathbf{b}^{*}-\mathbf{a})'\mathbf{A}(\mathbf{b}^{*}-\mathbf{a})\right\};$$

where $\mathbf{b}^{*} = \begin{pmatrix}\mathbf{b}\\ \mathbf{w}\end{pmatrix}$.

PROOF OF LEMMA 1: The result proceeds directly from Graybill (1983, p. 342, Thm. 10.5.1), on rewriting the quadratic forms involved as

$$\begin{aligned}(\mathbf{x}-\mathbf{a})'\mathbf{A}(\mathbf{x}-\mathbf{a}) &= (\mathbf{z}-\mathbf{a}_{\mathbf{z}})'\mathbf{A}_{\mathbf{zz}}(\mathbf{z}-\mathbf{a}_{\mathbf{z}}) + 2(\mathbf{z}-\mathbf{a}_{\mathbf{z}})'\mathbf{A}_{\mathbf{zw}}(\mathbf{w}-\mathbf{a}_{\mathbf{w}})\\ &\quad + (\mathbf{w}-\mathbf{a}_{\mathbf{w}})'\mathbf{A}_{\mathbf{ww}}(\mathbf{w}-\mathbf{a}_{\mathbf{w}})\\ &= \mathbf{z}'\mathbf{A}_{\mathbf{zz}}\mathbf{z} + 2\mathbf{z}'\left\{\mathbf{A}_{\mathbf{zw}}(\mathbf{w}-\mathbf{a}_{\mathbf{w}}) - \mathbf{A}_{\mathbf{zz}}\mathbf{a}_{\mathbf{z}}\right\}\\ &\quad + \left\{\mathbf{a}'_{\mathbf{z}}\mathbf{A}_{\mathbf{zz}}\mathbf{a}_{\mathbf{z}} - 2\mathbf{a}'_{\mathbf{z}}\mathbf{A}_{\mathbf{zw}}(\mathbf{w}-\mathbf{a}_{\mathbf{w}}) + (\mathbf{w}-\mathbf{a}_{\mathbf{w}})'\mathbf{A}_{\mathbf{ww}}(\mathbf{w}-\mathbf{a}_{\mathbf{w}})\right\}\end{aligned}$$

and $$(\mathbf{z}-\mathbf{b})'\mathbf{B}(\mathbf{z}-\mathbf{b})/2 \;=\; \tfrac{1}{2}\mathbf{z}'\mathbf{B}\mathbf{z} - \mathbf{z}'\mathbf{B}\mathbf{b} + \tfrac{1}{2}\mathbf{b}'\mathbf{B}\mathbf{b}\,.$$

PROOF OF PROPOSITION 1.

In the notation of §4.1 the EM criterion $Q(\omega,\omega^*)$ is now

$$Q(\omega,\omega^*) \triangleq \int_{\underset{\sim}{\mathbf{z}}} \ell(\omega\,|\underset{\sim}{\mathbf{x}}^{\dagger})\mathcal{P}(\underset{\sim}{\mathbf{z}}\,|\underset{\sim}{\mathbf{w}},\omega^*)\,d\underset{\sim}{\mathbf{z}},$$

with $\ell(\omega\,|\underset{\sim}{\mathbf{x}}^{\dagger}) = \ln \mathcal{P}(\underset{\sim}{\mathbf{x}}^{\dagger}|\omega)$. This becomes, on substituting (4.4) and (4.3),

$$Q(\omega,\omega^*) = c \;-\; \frac{T}{2}\ln|\Sigma| \;-\; \frac{1}{2}(2\pi)^{-\frac{1}{2}\sum_r^S m_r T_r}\left|\left(\tilde{\Omega}^{-1}\right)^*_{\mathbf{zz}}\right|^{\frac{1}{2}}\mathcal{I}(\omega,\omega^*),$$

with $c = -nT\ln(2\pi)/2$, $\left(\tilde{\Omega}^{-1}\right)^*_{\mathbf{zz}} = \left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}}\Big|_{\omega^*}$, and

$$\mathcal{I}(\omega,\omega^*) = \int_{\underset{\sim}{\mathbf{z}}} (\underset{\sim}{\mathbf{x}}^{\dagger} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}})'\tilde{\Omega}^{-1}(\underset{\sim}{\mathbf{x}}^{\dagger} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}})\exp\left\{-\tfrac{1}{2}(\underset{\sim}{\mathbf{z}} - \underset{\sim}{\mu}^*_{\mathbf{z|w}})'\left(\tilde{\Omega}^{-1}\right)^*_{\mathbf{zz}}(\underset{\sim}{\mathbf{z}} - \underset{\sim}{\mu}^*_{\mathbf{z|w}})\right\}d\underset{\sim}{\mathbf{z}}.$$

Applying Lemma 1 to $\mathcal{I}(\omega,\omega^*)$ then yields

$$(2\pi)^{-\frac{1}{2}\sum_r^S m_r T_r}\left|\left(\tilde{\Omega}^{-1}\right)^*_{\mathbf{zz}}\right|^{\frac{1}{2}}\mathcal{I}(\omega,\omega^*) \;=\; \mathrm{tr}\left\{\left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}}\left(\tilde{\Omega}^{-1}\right)^{*-1}_{\mathbf{zz}}\right\} \;+\; (\underset{\sim}{\mathbf{x}}^*_{a} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}})'\tilde{\Omega}^{-1}(\underset{\sim}{\mathbf{x}}^*_{a} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}}),$$

where the quadratic form simplifies as

$$\begin{aligned}(\underset{\sim}{\mathbf{x}}^*_{a} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}})'\tilde{\Omega}^{-1}(\underset{\sim}{\mathbf{x}}^*_{a} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}}) &= (\underset{\sim}{\mathbf{x}}^*_{a} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}})'(\tilde{\mathbf{P}}\tilde{\mathbf{B}})'^{-1}\left(\mathbf{I_T}\otimes\Sigma^{-1}\right)(\tilde{\mathbf{P}}\tilde{\mathbf{B}})^{-1}(\underset{\sim}{\mathbf{x}}^*_{a} - \underset{\sim}{\mu}^{\dagger}_{\mathbf{x}})\\ &= \underset{\sim}{\mathbf{u}}'^*_{a}\left(\mathbf{I_T}\otimes\Sigma^{-1}\right)\underset{\sim}{\mathbf{u}}^*_{a}\\ &\equiv \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{U}^{*\prime}\mathbf{U}^*\right).\end{aligned}$$

Furthermore, $\left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}}$ is block-diagonal, with typical block $\mathbf{I}_{\mathrm{T_r}}\otimes\overline{\mathbf{L}}_{\mathbf{r}}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{L}}'_{\mathbf{r}}$, so that

$$\mathrm{tr}\left\{\left(\tilde{\Omega}^{-1}\right)_{\mathbf{zz}}\left(\tilde{\Omega}^{-1}\right)^{*-1}_{\mathbf{zz}}\right\} \;=\; \sum_{r=1}^{S} T_r\,\mathrm{tr}\left(\overline{\mathbf{L}}_{\mathbf{r}}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{L}}'_{\mathbf{r}}\left(\overline{\mathbf{L}}_{\mathbf{r}}\boldsymbol{\Sigma}^{*-1}\overline{\mathbf{L}}'_{\mathbf{r}}\right)^{-1}\right).$$

Thus we finally have (ignoring the constant)

$$Q(\omega,\omega^*) \;=\; -\frac{T}{2}\ln\left|\boldsymbol{\Sigma}\right| \;-\; \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{U}^{*\prime}\mathbf{U}^{*}\right) \;-\; \frac{1}{2}\sum_{r=1}^{S} T_r\ \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\overline{\mathbf{L}}_{\mathbf{r}}'\left(\overline{\mathbf{L}}_{\mathbf{r}}\boldsymbol{\Sigma}^{*-1}\overline{\mathbf{L}}_{\mathbf{r}}'\right)^{-1}\overline{\mathbf{L}}_{\mathbf{r}}\right).$$

Collecting the trace terms then establishes the Proposition.

### 4.3 M-step: the concentrated expected log-likelihood

It is evident, from (4.7), that $Q(\omega,\omega^*)$ takes the familiar form of a Gaussian "log-likelihood". The FOC for $\Sigma$ will, accordingly, be of the form (3.4), implying

$$\hat{\boldsymbol{\Sigma}} = T^{-1}\tilde{\mathbf{S}}^{*}(\theta).$$

Hence we can, as before, construct a *concentrated* expected log-likelihood

$$Q_c(\theta,\omega^*) \;=\; -\frac{T}{2}\ln\left|\tilde{\mathbf{S}}^{*}(\theta)\right|, \qquad (4.8)$$

requiring maximization with respect to θ only. We now have the EM algorithm for the maximization of the observed log-likelihood in a variably-aggregated demand system as simply:

For $i$ = 0,1,2,…, some starting value $\hat{\omega}^{(0)} = (\hat{\theta}^{(0)},\hat{\boldsymbol{\Sigma}}^{(0)})$, and prespecified tolerance $\varepsilon$ :

(1) Compute $\underset{\sim}{\mu}_{\mathbf{z}|\mathbf{w}}(\hat{\omega}^{(i)})$, and hence $\underset{\sim}{\mathbf{x}}_g^{(i)} = \underset{\sim}{\mathbf{x}}_g(\hat{\omega}^{(i)})$.

(2) Construct $\underset{\sim}{\mathbf{u}}_g^{(i)}(\theta) = \underset{\sim}{\mathbf{u}}_g(\theta,\hat{\omega}^{(i)})$, $\tilde{\mathbf{S}}^{(i)}(\theta) = \tilde{\mathbf{S}}(\theta,\hat{\omega}^{(i)})$ and hence $Q_c(\theta,\hat{\omega}^{(i)})$.

(3) Maximize $Q_c(\theta,\hat{\omega}^{(i)})$ with respect to θ, so obtaining $\hat{\theta}^{(i+1)}$, $\tilde{\mathbf{S}}(\hat{\theta}^{(i+1)},\hat{\omega}^{(i)})$, and hence $\hat{\boldsymbol{\Sigma}}^{(i+1)} = T^{-1}\tilde{\mathbf{S}}(\hat{\theta}^{(i+1)},\hat{\omega}^{(i)})$.

(4) Repeat (1) – (3) until $\left|Q_c(\hat{\omega}^{(i+1)},\hat{\omega}^{(i)}) - Q_c(\hat{\omega}^{(i)},\hat{\omega}^{(i)})\right| < \varepsilon$.

## 4.4 Practical issues

### *4.4.1 No missing data in the final subperiod*

A few practical issues now arise. To begin with, notice that, in contrast to the treatment of GM, there is no explicit need, in our EM setup, for a complete set of expenditure data to exist in one of the subperiods. The only obvious limit on the number of expenditure categories that could be "modelled" in this way is the availability of the corresponding price data. In practice, as one might expect, there is a limit on the amount of data which can be said to be "missing", in this case imposed by the effect on the rate of convergence of increasing the proportion of "missing" to "observed".

Typically, however, the observed expenditures will be held to be "complete" for one of the subperiods, generally the last; in which case $n_S = n$, $m_S = 0$, $\mathbf{B_S} = \mathbf{I_n}$, the lower-right submatrix of $\tilde{\mathbf{P}}_\mathbf{w}$ reduces to the $nT_S \times nT_S$ identity matrix, while the corresponding rows of $\tilde{\mathbf{P}}_\mathbf{z}$, and subsequent matrices such as $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{G}}$, disappear entirely. Thus, for the 3-subperiod example of §5 following we will have

$$\tilde{\mathbf{P}}_\mathbf{z} = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1} \otimes \mathbf{J_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\mathbf{T}_2} \otimes \mathbf{J_2} & \mathbf{0} \end{bmatrix};$$

and

$$\tilde{\mathbf{P}}_\mathbf{w} = \begin{bmatrix} \mathbf{I}_{\mathbf{T}_1} \otimes \bar{\mathbf{J}}_\mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\mathbf{T}_2} \otimes \bar{\mathbf{J}}_\mathbf{2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{\mathrm{nT}_3} \end{bmatrix}.$$

### *4.4.2 Standard errors*

We now turn to the matter of obtaining appropriate standard errors at the end of the EM estimation. It can be shown (see Tanner (1993), p.47) that the "observed" Hessian,

$$H(\omega|\mathbf{W}) = \frac{\partial^2 \ell(\omega|\mathbf{W})}{\partial\omega\partial\omega'},$$

is the sum of the expected augmented Hessian and the variance of the augmented score; ie,

$$H(\omega|\mathbf{W}) = E_{\mathbf{Z}|\mathbf{W},\omega}\{H(\omega|\mathbf{X})\} + V_{\mathbf{Z}|\mathbf{W},\omega}\{q(\omega|\mathbf{X})\}$$
$$\equiv \left.\frac{\partial^2 Q(\omega,\omega^*)}{\partial\omega\partial\omega'}\right|_{\omega=\omega^*} + E_{\mathbf{Z}|\mathbf{W},\omega}\{q(\omega|\mathbf{X})q(\omega|\mathbf{X})'\};$$

so that the Hessian matrix naturally returned at the end of the EM estimation need only be "adjusted" by the expectation, taken with respect to $\mathcal{P}(\mathbf{Z}|\mathbf{W},\hat{\omega})$, of the outer product of the score, also evaluated at $\hat{\omega}$. The latter, in our case, is most simply "estimated" by the MC method discussed in §3.2. That is, we generate, with $\omega = \hat{\omega}$, $M$ augmented expenditure sets $\mathbf{X}_j$, compute the corresponding $q(\omega|\mathbf{X}_j)q(\omega|\mathbf{X}_j)'$, and obtain the average of the latter. $q(\omega|\mathbf{X})$ would in this case be given by Grose and McLaren (1999) equations (2.4) and (A.1).

Alternatively, since we in fact already have a functional form for $\ell(\omega|\mathbf{W})$ (equation (2.7)), the observed Hessian $H(\omega|\mathbf{W})$ could (in theory) be obtained directly, either by differentiation (analytic or numeric) of the analytic observed score $q(\omega|\mathbf{W})$ [12], or by simply computing the numeric second derivatives of $\ell(\omega|\mathbf{W})$.

### *4.4.3 Choice of demand system for estimation*

A final practical issue involves the feasibility of estimating any given demand system with the available data, even after augmentation. In short, for the algorithm to be feasible it must be possible to maximize the full $N$-commodity likelihood using the augmented dataset. In our case, with up to 18 commodities, this means that we must restrict our attention to either the extremely parsimonious LES, or to models, such as linearized AIDS, that can be estimated by iterative GLS (see Grose and McLaren (2000)). It must be noted, however, that the latter is already a fairly computer intensive procedure, and its insertion into an EM "loop" would be time consuming in the extreme.

---

[12] The derivatives of the "aggregated" log-likelihood with respect to $\Sigma$ and $\theta$ are given in GM.

## 5. EM estimation of the LES.

The linear expenditure system (LES) was estimated for Grose and McLaren (1999)'s quarterly data with 3 subperiods of differing aggregation. The data consists of 95 quarterly observations on up to 18 expenditure categories, covering the period 1974(3$^{rd}$ quarter) – 1998(1$^{st}$ quarter). The 1$^{st}$ subperiod (1974(3) – 1985(3)) comprises 45 quarterly observations on 12 expenditure categories; the 2$^{nd}$ (1985(4) – 1989(3)) involves 16 quarterly observations on 16 categories; and the 3$^{rd}$ (1989(4) – 1998(1)) involves 34 quarterly observations on 18 categories. The 12, 16, and 18 commodity sets are as listed in Grose and McLaren (1999, §4.3). All expenditures are in A$ per capita; prices are measured by the IPD for each expenditure category, and equal unity in 1989/90. The "omitted" category in all subperiods was, once again, Food.

The LES with "free" covariance matrix was estimated by EM as described above, with the expectation stage performed analytically as per §4, and the maximization stage performed by numerically maximizing (using Newton-Raphson) the expected log-likelihood with respect to $\theta = (\tilde{\gamma}, \beta)$ after concentrating $\Sigma$ out as per §4.3.

An attempt to check for a global maximum[13] was made by starting the algorithm at different starting values. For the first "pass", the EM algorithm was started at $\tilde{\gamma} = \mathbf{0}$, $\beta = T_3^{-1}\mathbf{W}_3'\iota_{T_3} = \overline{\mathbf{W}}_3$ (the mean expenditure shares in the final subperiod), and $\Sigma = T_3^{-1} diag\left((\mathbf{W}_3 - \overline{\mathbf{W}}_3)'(\mathbf{W}_3 - \overline{\mathbf{W}}_3)\right)$; that is, the initial $\Sigma$ is a diagonal matrix, with elements set equal to the variance of the corresponding expenditure share in the final subperiod. Results are given in Table 2, with standard errors estimated as per §4.4.2.

For the second pass we started the EM algorithm at the MLE's obtained by maximizing the observed ("aggregated") log-likelihood assuming De Boer and Harkema's covariance matrix as per Grose and McLaren (1999). While the algorithm now took slightly longer to converge, and resulted in a slightly smaller maximized log-observed likelihood value, the parameter estimates obtained were very similar. By way of a final comparison, the algorithm was repeated, this time starting at the MLE's of $\theta$ obtained using the "complete" $T \times N$ expenditure set recently made available by

[13] While it can be shown that EM will converge to a maximum of the expected log-likelihood, this is not necessarily a global maximum.

the ABS. The parameter estimates obtained were once again very similar to those obtained from the mean expenditure share starters.

## Conclusion

The problem considered here is the ML estimation of a consumer demand system in the situation where not all expenditures are observed for all commodity categories in all time periods. The major problem with the estimation of such a system is that, while the likelihood function can be written down simply enough, it cannot be maximized directly because of the $\frac{1}{2}n(n+1)$ covariance parameters that must now also be included in the objective function. It may be worth noting that the complete log-likelihood cannot be satisfactorily maximized even with 95 observations available on all 18 expenditure categories, *unless* $\Sigma$ is concentrated out.

The strategy considered in this paper takes a "data augmentation" approach to our unobserved expenditure data. This leads to an EM method, in which we maximize the *expectation* of the complete log-likelihood, with the expectation taken with respect to the conditional density of the missing data. This has the advantage that it effectively reformulates the optimization problem such that $\Sigma$ *can* be concentrated out; and thus avoids the need to restrictively reparameterize the covariance matrix. We find that convergence, while slow, seems quite sure; in contrast to earlier attempts to maximize the observed log-likelihood directly. Furthermore, starting the algorithm at quite different parameter values did not result in substantially different estimates, giving us some confidence in concluding that the mode obtained is global, and hence the MLE.

The prime disadvantage of the EM approach is the slow rate of convergence, which would become even slower if the proportion of missing data were increased. This means that, while it may be feasible to estimate a model such as AIDS, the time taken for each M-step if we do may make the procedure impractical for routine application. Finally, EM naturally requires sufficient data to permit conventional maximization at each iteration; and so cannot be applied to very small (eg. annual) datasets any more than can ordinary ML. In other words, it is worth checking whether the chosen model can actually be estimated at the desired level of commodity disaggregation, before considering the additional complication of having some expenditure categories partially unobserved.

On the other hand, if *additional* data is available at a coarser level of commodity aggregation then the EM algorithm presented here may well represent a convenient means of estimating the model for the full commodity set that avoids the need to resort to a more restrictive respecification.

# References


De Boer, P.M.C. and R. Harkema (1986), “Maximum likelihood estimation of sum-constrained linear models with insufficient observations”. *Economics Letters*, 20, 325–329.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977), “Maximum likelihood from incomplete data via the EM algorithm”. *Journal of the Royal Statistical Society B*, 39, 1–38.

Fry, J.M., T.R.L. Fry and K.R. McLaren (1996), “The Stochastic Specification of Demand Share Equations: Restricting Budget Shares to the Unit Simplex”. *Journal of Econometrics*, 73(2), 377–385.

Graybill, F.A. (1983), *Matrices with Applications in Statistics, 2nd ed*. Wadsworth, Belmont California.

Grose, S.D. and K.R. McLaren (1999), “Estimating consumer demand with missing expenditure data”. *Econometric Society Australasian Meeting 1999, Proceedings*, UTS.

Grose, S.D. and K.R. McLaren (2000), “SUR estimation of AIDS”. *Working Paper, Econometrics and Business Statistics, Monash University*, Melbourne.

McLachlan, G.J. and T. Krishnan (1997), *The EM Algorithm and Extensions*. Wiley, New York.

Tanner, M.A. (1993), *Tools for Statistical Analysis*. Springer-Verlag, New York.

Table 1. Abbreviations for expenditure categories

| | |
|---|---|
| *Food* | FOD |
| *Cigarettes and Tobacco* | CGT |
| *Alcohol and spirits* | ALC |
| *Clothing and footwear* | CFF |
| *Dwelling rent* | RNT |
| *Purchase of motor vehicles* | MVP |
| *Household appliances* | HAP |
| *Other household durables* | HDU |
| *Postal and telecommunications* | TEL |
| *Gas, electricity and fuel* | GEF |
| *Fares* | FRS |
| *Operation of motor vehicles* | MVO |
| *Health* | MED |
| *Entertainment and recreation* | REC |
| *Financial services* | FIN |
| *Other goods* | OGD |
| *Other services* | OSV |
| *Net expenditure overseas* | NEO |

Table 2. EM estimation of the LES. Unrestricted covariance matrix; 3 subperiods; 18 expenditure categories; quarterly data

*Started from $\gamma = 0$, $\beta$ = mean expenditure share.*

| | β | | γ | | σ [‡] |
|---|---|---|---|---|---|
| Expenditure category | estimate | standard error | estimate [†] | standard error | estimate |
| FOD | 0.1191 | *0.0049* | 0.2636 | *0.0108* | 0.00472 |
| CGT | 0.0263 | *0.0013* | 0.0135 | *0.0018* | 0.00308 |
| ALC | 0.0580 | *0.0036* | 0.0455 | *0.0060* | 0.00736 |
| CFF | 0.0845 | *0.0052* | 0.0455 | *0.0073* | 0.01184 |
| RNT | 0.0408 | *0.0065* | 0.4765 | *0.0100* | 0.02344 |
| MVP | 0.0708 | *0.0085* | -0.0039 | *0.0163* | 0.00609 |
| HAP | 0.0624 | *0.0030* | -0.0089 | *0.0033* | 0.00835 |
| HDU | 0.0922 | *0.0046* | -0.0320 | *0.0079* | 0.01020 |
| TEL | 0.0245 | *0.0011* | 0.0047 | *0.0013* | 0.00228 |
| GEF | 0.0075 | *0.0037* | 0.0534 | *0.0065* | 0.00369 |
| FRS | 0.0336 | *0.0022* | 0.0257 | *0.0041* | 0.00164 |
| MVO | 0.0282 | *0.0021* | 0.1541 | *0.0037* | 0.00147 |
| MED | 0.0220 | *0.0053* | 0.1545 | *0.0082* | 0.01428 |
| REC | 0.0178 | *0.0043* | 0.1117 | *0.0072* | 0.00975 |
| FIN | -0.0070 | *0.0036* | 0.1009 | *0.0047* | 0.01297 |
| OGD | 0.1710 | *0.0045* | -0.0616 | *0.0051* | 0.01453 |
| OSV | 0.1879 | *0.0083* | -0.0134 | *0.0121* | 0.01842 |
| NEO | -0.0396 | *0.0032* | -0.0014 | *0.0008* | 0.01665 |

| | | | |
|---|---|---|---|
| Initial observed log-likelihood | 2487.26 | Final observed log-likelihood | 6438.54 |
| Initial expected log-likelihood | 7168.01 | Final expected log-likelihood | 8199.75 |
| Number of EM iterations | 769 | Final $\left\| \theta^{(i+1)} - \theta^{(i)} \right\|$ | 0.000105 |
| Time to convergence | 12.9 minutes | Number of observations | 95 |

[†] For estimation purposes the matrix of price ratios (that is, the ratio of price (an index, =1 in 1989/90) to total expenditure per capita (in Australian $)) has been scaled up by $10^3$. Estimates of γ are thus in units of thousands of 1989/90$.
[‡] σ is the square root of the corresponding diagonal element of the cross-commodity covariance matrix.

Table 3. EM estimation of the LES. Unrestricted covariance matrix; 3 subperiods; 18 expenditure categories; quarterly data. Summary statistics.

*Started from the MLE's obtained by maximizing the observed ("aggregated") log-likelihood assuming De Boer and Harkema's covariance matrix with quarterly data.*

| | | | |
|---|---|---|---|
| Initial observed log-likelihood | 5588.39 | Final observed log-likelihood | 6434.90 |
| Initial expected log-likelihood | 7414.93 | Final expected log-likelihood | 8189.93 |
| Number of EM iterations | 776 | Final $\left\| \theta^{(i+1)} - \theta^{(i)} \right\|$ | 0.0001284 |
| Time to convergence | 12.75 minutes | Number of observations | 95 |

Table 4. EM estimation of the LES. Unrestricted covariance matrix; 3 subperiods; 18 expenditure categories; quarterly data. Summary statistics.

*Started from the MLE's obtained by assuming complete expenditure information available in all subperiods.*

| | | | |
|---|---|---|---|
| Initial observed log-likelihood | 6306.13 | Final observed log-likelihood | 6438.12 |
| Initial expected log-likelihood | 7921.85 | Final expected log-likelihood | 8198.16 |
| Number of EM iterations | 520 | Final $\left\| \theta^{(i+1)} - \theta^{(i)} \right\|$ | $9.558 \times 10^{-5}$ |
| Time to convergence | 8.4 minutes | Number of observations | 95 |