MEMORANDUM

No 19/2004

Repeated surveys and the Kalman filter

Jo Thori Lind

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.oekonomi.uio.no/
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no/
e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 18	Kari Eika When quality today affects service needs Tomorrow. 30 pp.
No 17	Rolf Golombek and Michael Hoel Unilateral emission reductions when there are cross -country technology spillovers. 23 pp.
No 16	Kjetil Bjorvatn and Alexander W. Cappelen Globalisation, inequality and redistribution. 17 pp.
No 15	Alexander W. Cappelen and Bertil Tungodden Rewarding effort. 31 pp.
No 14	Alexander W. Cappelen and Bertil Tungodden The liberal egalitarian paradox. 15 pp.
No 13	John K. Dagsvik and Steinar Strøm Sectoral labor supply, choice restrictions and functional form**.** 15 pp.
No 12	Tiziano Razzolini The Norwegian market for pharmaceuticals and the non-mandatory substitution reform of 2001: the case of enalapril. 58 pp.
No 11	Rolf Golombek and Michael Hoel Climate Agreements and Technology Policy. 38 pp.
No 10	Isilda Shima The shadow economy in Norway: Demand for currency approach. 23 pp.
No 09	Steinar Holden. Wage formation under low inflation. 22 pp.

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/

Repeated surveys and the Kalman filter*

Jo Thori Lind†

July 2, 2004

Abstract

The time series nature of repeated surveys is seldom taken into account. The few studies that take this into account usually smooth the period-wise estimates without using the cross sectional information. This leads to inefficient estimation. I present a statistical model of repeated surveys and construct a computationally simple estimator based on the Kalman filter which efficiently uses the whole underlying data set, but which is computationally very simple as we only need the first and second empirical moments of the data.

Keywords: Surveys, Kalman filter, time series

JEL Classification: C22, C53, C81

1 Introduction

A number of statistical series are estimated on the basis regularly repeated surveys. The most common approach is to publish parameter estimates at regular intervals, say each year, pooling surveys collected throughout the year but ignoring previous years. As it is natural to assume that most parameters of interest evolve slowly and smoothly, this is an inefficient use of the data.

*I wish to thank Jørgen Aasness, Eivind Bernhardsen, and Knut R. Wangen with whom I have benefited from comments and discussions.

†Department of Economics, University of Oslo, PB 1095 Blindern, 0317 Oslo, Norway. Email: j.t.lind@econ.uio.no

Jessen (1942) was the first to suggest to use times series techniques to improve the results from repeated surveys. It was studied in more detail by Gurney and Daly (1965), and the methodology was further improved by Blight and Scott (1973) and Scott and Smith (1974) who suggest using statistical signal extraction methods to filter the time specific estimates of the parameters of interest. See e.g. the survey by Binder and Hidiroglou (1988) for further details on subsequent developments within this tradition. A more general theory of signal extraction using the Kalman filter was suggested by Tam (1987) and further developed by e.g. Binder and Dick (1989), Harvey and Chung (2000), and Pfeffermann (1991).

Their approach is to estimate a parameter, such as the mean, on each individual survey and then apply the Kalman filter on the estimates. However, there is an important loss of efficiency as a lot of information contained in each cross section may be lost by this two step procedure. A more satisfactory approach is to integrate the time series model and the modelling of the individual observations at each period.

An important step in this direction is obtained by sequential processing of each element of the observation vector as first suggested by Anderson and Moore (1979, Section 6.4) and further developed by Durbin and Koopman (2001, Section 6.4). However, if each survey is relatively large, this leads to unnecessary long recursions, and may particularly render estimation of the hyper parameters more burdensome than necessary.

I suggest a model where the parameters of interest evolve smoothly over time, and where each observed data point is a noisy observation of the parameter of interest. If we use the ordinary Kalman filter algorithm, this will lead to extremely large matrices that has to be inverted, hence causing severe computational problems unless each survey is extremely small. In the present work I first show how repeated surveys may be written on state-space form and then how the Kalman filter algorithm may be transformed to make estimation feasible without running into computational problems. It turns out that to estimate the mean of the population, we only need the empirical first and second moments in each period, so both the computational burden and the data requirements are small. I also extend the estimator to allow for heterogeneity between groups of individuals.

2 A simple model

I first present a simplified version of the model an show how the Kalman filter may be applied to this model. Then I explore a more general model in Section 4. At a survey date $t \in (1, \ldots, T)$ we observe N_t individuals. Let y_{it} denote the observations on individual i at time t. At the time being I treat y_t as a scalar; this is generalized below. I focus on estimating averages of the y_{it}'s. We may write

$$y_{it} = \mu_t + \varepsilon_{it} \tag{1}$$

where $\varepsilon_{ij} \sim N(\mathbf{0}, \sigma_t^2)$ contain individual unobserved characteristics and possible sampling errors. I assume that the ε_{ij}s are independent both within and between surveys. The variable of interest is then μ_t.

Assume that there is a n-vector α_t following a linear Markov process, i.e.

$$\alpha_t = F\alpha_{t-1} + \xi_t, \tag{2}$$

such that $\mu_t = Z\alpha_t$, where $\xi_t \sim N(\mathbf{0}_{n\times 1}, Q)$, F is a $n \times n$ transition matrix, and Z a $1 \times n$ observation matrix. $\mathbf{0}_{n\times 1}$ denotes the n-dimensional zero vector. Defining the stacked matrices $\tilde{y}_t = (y_{1t}, \ldots, y_{N_t t})'$ and $\tilde{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{N_t t})'$ we can write the complete model as

$$\begin{aligned}
\tilde{y}_t &= \iota_{N_t} Z\alpha_t + \tilde{\varepsilon}_t && \text{(3a)}\\
\alpha_t &= F\alpha_{t-1} + \xi_t && \text{(3b)}\\
\tilde{\varepsilon}_t &\sim N\left(\mathbf{0}_{N_t\times 1}, \sigma_t^2 I_{N_t}\right) && \text{(3c)}\\
\xi_t &\sim N(\mathbf{0}_n, Q) && \text{(3d)}\\
\alpha_0 &\sim N(a_0, Q_0), && \text{(3e)}
\end{aligned}$$

where I also added assumptions about the distribution of the initial state α_0. ι_{N_t} denotes a N_t-dimensional unit vector. If we treat $\iota_{N_t} Z$ as a single matrix transforming the state vector into the expectation of the observed data, we see that this is a model on state space form[1].

3 The Kalman filter

If we know the parameters of the model, an optimal estimate of the α's and the μ's may be

[1] The dimension of $\tilde{y}_t$ usually varies with time, but it is easily shown that this does not cause any difficulties (Durbin and Koopman 2001: Section 4.10)

calculated by means of the Kalman filter (see e.g. Harvey (1989) or Hamilton (1994: Ch. 13)). At date t, the information set is $\mathcal{Y}_t = (\tilde{y}_1, \ldots, \tilde{y}_t)'$. Let us denote the expectation and the covariance matrix of the vector α_{t_1} given the information set at date t_2 as

$$\begin{aligned} a_{t_1|t_2} &\equiv E\left(\alpha_{t_1} | \mathcal{Y}_{t_2}\right) \\ V_{t_|t_2} &= E\left[\left(\alpha_{t_1} - a_{t_1|t_2}\right)\left(\alpha_{t_1} - a_{t_1|t_2}\right)' | \mathcal{Y}_{t_2}\right]. \end{aligned}$$

The Kalman filter is calculated by the following recursion:

$$a_{t|t-1} = Fa_{t-1|t-1} \tag{4a}$$

$$V_{t|t-1} = FV_{t-1|t-1}F' + Q \tag{4b}$$

$$a_{t|t} = a_{t|t-1} + K_t\left(\tilde{y}_t - \iota_{N_t} Z a_{t|t-1}\right) \tag{4c}$$

$$V_{t|t} = V_{t|t-1} - K_t \iota_{N_t} Z V_{t|t-1} \tag{4d}$$

$$K_t = V_{t|t-1}Z'\iota'_{N_t}\left(\iota_{N_t} Z V_{t|t-1} Z' \iota'_{N_t} + \sigma_t^2 I_{N_t}\right)^{-1}. \tag{4e}$$

The two first equations are straightforward to calculate. However, in their current form, (4c-4e) include the matrix $\left(\iota_{N_t} Z V_{t|t-1} Z' \iota'_{N_t} + \sigma_t^2 I_{N_t}\right)^{-1}$. Unless N_t is small, this matrix is of high dimension, and hence inversion requires large amounts of calculation. However, due to the data structure assumed above, (4c-4e) may be written as

$$\begin{aligned} V_{t|t} &= \left(V_{t|t-1}^{-1} + \sigma_t^{-2} N_t Z'Z\right)^{-1} \\ a_{t|t} &= a_{t|t-1} + \sigma_t^{-2} N_t V_{t|t} Z'\left(\bar{y}_t^G - Z a_{t|t-1}\right) \end{aligned} \tag{5}$$

where $\bar{y}_t$ denotes the average of y_{it}. The proof, which is based on the matrix inversion lemma, is given for the more general structure in Section 4.

Using the recursion (5), we calculate estimates of α_t given the information set $\mathcal{Y}_t$. To obtain efficient estimates of the states μ_{it} we should employ the full information set $\mathcal{Y}_T$. To achieve this, we use the ordinary Kalman smoother (Hamilton 1994), which only use matrices of low dimensionality.

4 A general model

Before I discuss estimation of the parameters of the model, I will set up a more general model that allows for vectors of observed variables and more importantly group-wise heterogeneity between observed individuals.

Now y_{it} denote the m-vector of observations on individual i at time t. Again, we can write

$$y_{it} = \mu_{it} + \varepsilon_{it} \tag{6}$$

where $\varepsilon_{ij} \sim N(\mathbf{0}_{m\times 1}, \Sigma_t)$ and μ_{it} are vectors. It is normally not particularly interesting to estimate a separate μ for every individual. Above I assumed that the μ_{it}'s were the same for all individuals at a particular date. But it is often fruitful to group individuals into e.g. geographical regions or household types, and allow the groups to have different μs. This is the approach we will pursue now. Assume that there are G such groups, and an associated μ_{gt} for all $g \in (1, \ldots, G)$ at every date.[2] It will be useful to consider the stacked vector of all the means at date t

$$\mu_t = (\mu_{1t}', \ldots, \mu_{Gt}')'. \tag{7}$$

Expression (6) may now be written as

$$y_{it} = J_{g(i)t}\mu_t + \varepsilon_{it} \tag{8}$$

where g is the function that associates to each individual i the group that it belongs to, and J_{gt} the selection matrix

$$J_{gt} = \left(\begin{array}{ccc} \mathbf{0}_{(g-1)m\times m} & \vdots & I_m \quad \vdots \quad \mathbf{0}_{(G-g)m\times m} \end{array} \right), \tag{9}$$

which selects the appropriate elements from the vector μ_t for individuals in group g.

As above we have a process

$$\alpha_t = F\alpha_{t-1} + \xi_t, \tag{10}$$

such that $\mu_t = Z\alpha_t$, so Z translates α_t into each group's vector of means. This structure allows for some components, e.g. seasonals, to be identical across groups and others to be group specific.

Defining the stacked matrices $J_t = \left(J_{g(1)t}', \ldots, J_{g(N_t)t}'\right)'$, $\tilde{\varepsilon}_t = \left(\varepsilon_{1t}', \ldots, \varepsilon_{N_t t}'\right)'$, and $\tilde{y}_t =$

[2]The covariance matrix Σ_t is assumed the be identical for every group, but this assumption is easily relaxed.

$\left(y_{1t}', \ldots, y_{N_t t}'\right)'$, we can write the complete model as

$$\begin{aligned}
\tilde{y}_t &= J_t Z \alpha_t + \tilde{\varepsilon}_t && \text{(11a)} \\
\alpha_t &= F\alpha_{t-1} + \xi_t && \text{(11b)} \\
\tilde{\varepsilon}_t &\sim N\left(\mathbf{0}_{N_t m \times 1}, I_{N_t} \otimes \Sigma_t\right) && \text{(11c)} \\
\xi_t &\sim N\left(\mathbf{0}_n, Q\right) && \text{(11d)} \\
\alpha_0 &\sim N\left(a_0, Q_0\right), && \text{(11e)}
\end{aligned}$$

In this model, the Kalman filter is calculated by the recursion

$$\begin{aligned}
a_{t|t-1} &= F a_{t-1|t-1} && \text{(12a)} \\
V_{t|t-1} &= F V_{t-1|t-1} F' + Q && \text{(12b)} \\
a_{t|t} &= a_{t|t-1} + K_t\left(\tilde{y}_t - J_t Z a_{t|t-1}\right) && \text{(12c)} \\
V_{t|t} &= V_{t|t-1} - K_t J_t Z V_{t|t-1} && \text{(12d)} \\
K_t &= V_{t|t-1} Z' J_t' \left(J_t Z V_{t|t-1} Z' J_t' + I_{N_t} \otimes \Sigma_t\right)^{-1}. && \text{(12e)}
\end{aligned}$$

Again, this implies inverting a high dimensional matrix, but the data structure allows us to write (12c-12e) as

$$\begin{aligned}
V_{t|t} &= \left[V_{t|t-1}^{-1} + Z'\left(\mathcal{N}_t^G \otimes \Sigma_t^{-1}\right) Z\right]^{-1} && \text{(13)} \\
a_{t|t} &= a_{t|t-1} + V_{t|t} Z'\left(\mathcal{N}_t^G \otimes \Sigma_t^{-1}\right)\left(\bar{y}_t^G - Z a_{t|t-1}\right).
\end{aligned}$$

In these expressions, $\bar{y}_t^G$ denotes the within group averages defined as

$$\bar{y}_t^G \equiv \left(\begin{array}{ccc} \frac{1}{N_1^g} \sum_{g(i)=1} y_{it}' & \cdots & \frac{1}{N_G^g} \sum_{g(i)=G} y_{it}' \end{array} \right)'. \qquad \text{(14)}$$

The matrix $\mathcal{N}_t^G$ is the $G \times G$ matrix with the number of members of each group at date t along the diagonal. The proof is found in Appendix A.

5 Estimation

The algorithm described above was based upon the knowledge of the parameters of the model. Since they are normally not known, they will have to be estimated. Below I derive estimators for the parameters, but assume that F and Z are known matrices. It is straightforward to extend the framework to allow for estimating selected parameters in these matrices.

The usual approach to estimating parameters in Kalman filter models is maximum likelihood. An alternative approach based on the EM algorithm is explored in Appendix B. The likelihood of the data given a set of parameter values is

$$f\left(\mathcal{Y}_T;\Theta\right)=f\left(\tilde{y}_1\right)f\left(\tilde{y}_2|\mathcal{Y}_1\right)\cdots f\left(\tilde{y}_T|\mathcal{Y}_{T-1}\right). \tag{15}$$

Furthermore, it follows from (11) that

$$\tilde{y}_t|\mathcal{Y}_{t-1}\sim N\left(J_tZa_{t|t-1},\Omega_t\right) \tag{16}$$

where

$$\begin{aligned}\Omega_t &= E\left[\left(J_tZ\left(\alpha_t-a_{t|t-1}\right)+\tilde{\varepsilon}_t\right)\left(J_tZ\left(\alpha_t-a_{t|t-1}\right)+\tilde{\varepsilon}_t\right)'\right]\\ &= J_tZV_{t|t-1}Z'J_t'+I_{N_t}\otimes\Sigma_t.\end{aligned}$$

The log likelihood of the observed sample is

$$\ln L=-\frac{\sum_{t=1}^{T}N_t}{2}\ln\left(2\pi\right)-\frac{1}{2}\sum_{t=1}^{T}\left[\ln|\Omega_t|+\left(\tilde{y}_t-J_tZa_{t|t-1}\right)'\Omega_t^{-1}\left(\tilde{y}_t-J_tZa_{t|t-1}\right)\right]. \tag{17}$$

Due to the high dimension of Ω_t, calculation of $|\Omega_t|$ by direct calculations is extremely time consuming, and will not work on most computer systems. However, Appendix A shows that we may rewrite the expression. First,

$$|\Omega_t|=|\Sigma_t|^{N_t-G}\prod_{h=1}^{G}|\Lambda_h| \tag{18}$$

where

$$\Lambda_h:=\begin{cases}N_1^gJ_1ZV_{t|t-1}Z'J_1'+\Sigma_t & \text{if } h=1\\ N_h^gJ_hZ\left[V_{t|t-1}^{-1}+\sum_{i=1}^{h-1}N_i^gZ'J_i'\Sigma_t^{-1}J_iZ\right]^{-1}Z'J_h'+\Sigma_t & \text{if } h>1.\end{cases}$$

Second,

$$\begin{aligned}\Psi_t &: = \left(\tilde{y}_t-J_tZa_{t|t-1}\right)'\Omega_t\left(\tilde{y}_t-J_tZa_{t|t-1}\right)\\ &= \sum_{h=1}^{G}\mathrm{tr}\left[N_{ht}^g\Sigma^{-1}\,\mathrm{Cov}_{ht}\,y_{it}\right]\\ &\quad+\left(\bar{y}_t^G-Za_{t|t-1}\right)'\Xi_t\left\{I_{Gm}-Z\left[V_{t|t-1}^{-1}+Z'\Xi_tZ\right]^{-1}Z'\Xi_t\right\}\left(\bar{y}_t^G-Za_{t|t-1}\right).\end{aligned} \tag{19}$$

where N_{ht}^{g} is the number of members of group h at data t, $Cov_{ht}\left(y_{it}\right)$ denotes the intra-group empirical variance-covariance matrix of the y_{it}s at date t (without degrees of freedom-adjustment), and $\Xi_t = \mathcal{N}^G \otimes \Sigma_t^{-1}$. From equations (18) and (19) we can then calculate the likelihood value

$$\ln L = -\frac{\sum_{t=1}^{T} N_t}{2} \ln\left(2\pi\right) - \frac{1}{2} \sum_{t=1}^{T} \left[\ln\left|\Omega_t\right| + \Psi_t\right]. \tag{20}$$

6 Conclusion

I have presented a modified Kalman filtering algorithm to perform calculations on repeated samples. The procedure makes it possible to obtain efficient estimates of underlying estimates of the laws of motion of the parameters of interest. By using the Kalman filter to smooth the estimates from each sample, we get more precise estimates in each period. Hence even if each survey is small, we get reliable estimates, so we can produce estimates with higher frequency than what has been possible so far. By defining each group as a geographical area, the procedure is also applicable for small area estimation and can be extended to improve upon the techniques described in e.g. Pfeffermann (2002). Finally, forecasting is simple to perform and have well-known properties when using techniques based on the Kalman filter. At the present stage, the method only admits estimation of sample means. An interesting extension would be to allow for estimation of repeated regression coefficients integrating the estimation of the regressions with the Kalman filter.

References

Anderson, B. D. O., and J. B. Moore (1979): *Optimal Filtering.* Englewood Cliffs: Prentice-Hall.

Binder, D. A. and J. P. Dick (1989): "Modelling and estimation for repeated surveys." *Survey Methodology* 15: 29-45.

Blight, B. J. N. and A. J. Scott (1973): "A stochastic model for repeated surveys." *Journal of the Royal Statistical Society* Series B, 35: 61-66.

Dempster, A. P., N. Laird and D. B. Rubin (1977): Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*

B39, 1-38.

Durbin, J., and S. J. Koopman (2001): *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Gurney, M. and J. F. Daly (1965): "A multivariate approach to estimation in periodic sample surveys." *American Statistical Association Proceedings of the Social Statistics Section* 8: 242-57.

Hamilton, J. D. (1994): *Time series analysis*. Princeton: Princeton University Press.

Harvey, A. C. (1989): *Forecasting, structural time series and the Kalman filter*. Cambridge: Cambridge University Press.

Harvey, A. C. and C.-H. Chung (1999): "Estimating the underlying change in UK unemployment." *Journal of the Royal Statistical Society Series A* 163: 303-39.

Jessen, R. J. (1942): "Statistical investigation of a farm survey for obtaining farm facts." *Iowa Agricultural Experimental Station Research Bulletin* 304: 54-59.

Lütkepohl, H. (1996): *Handbook of Matrices*. Chichester: Wiley.

Pfeffermann, D. (1991): "Estimation and seasonal adjustment of population means using data from repeated surveys." *Journal of Business and Economic Statistics 9*, 163-77.

Pfeffermann, D. (2002): "Small area estimation - New developments and directions." *International Statistical Review* 70: 125-43.

Scott, A. J. and T. M. F. Smith (1974): "Analysis of repeated surveys using time series methods." *Journal of the American Statistical Association* 69: 647-78.

Shumway, R. H. and D. S. Stoffer (1982): An approach to time series smoothing and forecasting from the EM algorithm. *Journal of Time Series Analysis 3*, 253-64.

Tam, S. M. (1987): "Analysis of repeated surveys using a dynamic linear model." *International Statistical Review* 55: 63-73.

Watson, M. W. and R. F. Engle (1983): Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *Journal of Econometrics 23*, 385-400.

Appendix

A Proofs

A.1 Proof of equation (13)

From the matrix inversion lemma (Lütkepohl 1996: 29), we have

$$\begin{aligned} & \left(J_t Z V_{t|t-1} Z' J_t' + I_{N_t} \otimes \Sigma_t\right)^{-1} \\ = \; & I_{N_t} \otimes \Sigma_t^{-1} - I_{N_t} \otimes \Sigma_t^{-1} J_t Z \left(V_{t|t-1}^{-1} + Z' J_t' \left(I_{N_t} \otimes \Sigma_t^{-1}\right) J_t Z\right)^{-1} Z' J_t' \left(I_{N_t} \otimes \Sigma_t^{-1}\right). \end{aligned} \tag{21}$$

Furthermore,

$$\begin{aligned} J_t' \left(I_{N_t} \otimes \Sigma_t^{-1}\right) J_t &= \begin{pmatrix} J_{g(1)t}' & \cdots & J_{g(N_t)t}' \end{pmatrix} \begin{pmatrix} \Sigma^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma^{-1} \end{pmatrix} \begin{pmatrix} J_{g(1)t} \\ \vdots \\ J_{g(N_t)t} \end{pmatrix} \\ &= \sum_{i=1}^{N_t} J_{g(i)t}' \Sigma^{-1} J_{g(i)t}, \end{aligned}$$

and

$$\begin{aligned} J_{g(i)t}' \Sigma^{-1} J_{g(i)t} &= \begin{pmatrix} \mathbf{0}_{m\times m} \\ \vdots \\ I_m \\ \vdots \\ \mathbf{0}_{m\times m} \end{pmatrix} \Sigma^{-1} \begin{pmatrix} \mathbf{0}_{m\times m} & \cdots & I_m & \cdots & \mathbf{0}_{m\times m} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m\times m} & \cdots & \Sigma^{-1} & \cdots & \mathbf{0}_{m\times m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} \end{pmatrix} \end{aligned}$$

where the Σ^{-1} is in the $g(i) \times g(i)$'th position. Let N_h^g denote the number of members in group h, and let $N^G = \operatorname{diag}(N_1^g, \ldots, N_G^g)$. Then

$$J_t' \left(I_{N_t} \otimes \Sigma_t^{-1}\right) J_t = \mathcal{N}^G \otimes \Sigma^{-1}.$$

Hence the Kalman gain may be written as

$$\begin{aligned} K_t &= V_{t|t-1}Z'J_t'\left[I_{N_t}\otimes\Sigma^{-1}-\left(I_{N_t}\otimes\Sigma^{-1}\right)J_tZ\left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}Z'J_t'\left(I_{N_t}\otimes\Sigma^{-1}\right)\right] \\ &= V_{t|t-1}\left[I_n-Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}\right]Z'J_t'\left(I_{N_t}\otimes\Sigma^{-1}\right) \\ &= \left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}Z'J_t'\left(I_{N_t}\otimes\Sigma^{-1}\right), \end{aligned}$$

and then

$$a_{t|t}-a_{t|t-1}=\left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}Z'J_t'\left(I_{N_t}\otimes\Sigma^{-1}\right)\left(\tilde{y}_t-J_tZa_{t|t-1}\right).$$

Since

$$J_{g(i)t}'\Sigma^{-1}\left(y_{it}-J_{g(i)t}Za_{t|t-1}\right)=\begin{pmatrix}\mathbf{0}_{m\times1}\\ \vdots\\ y_{it}-J_{g(i)t}Za_{t|t-1}\\ \vdots\\ \mathbf{0}_{m\times1}\end{pmatrix},$$

where the $y_{it}-J_{g(i)t}Za_{t|t-1}$ is in the $g\left(i\right)$'th position, we have

$$\begin{aligned} J_t'\left(I_{N_t}\otimes\Sigma^{-1}\right)\left(\tilde{y}_t-J_tZa_{t|t-1}\right) &= \sum_{i=1}^{N_t}J_{g(i)t}'\Sigma^{-1}\left(y_{it}-J_{g(i)t}Za_{t|t-1}\right) \\ &= \left(\mathcal{N}^G\otimes\Sigma^{-1}\right)\left(\bar{y}_t^G-Za_{t|t-1}\right) \end{aligned} \tag{22}$$

where

$$\bar{y}_t^G\equiv\begin{pmatrix}\frac{1}{N_1^g}\sum_{g(i)=1}y_{it}\\ \vdots\\ \frac{1}{N_G^g}\sum_{g(i)=G}y_{it}\end{pmatrix}$$

is the vector of stacked averages and we used the fact that $\left(J_{1t}',\ldots J_{Gt}'\right)'=I_{Gm}$. Consequently, the Kalman updating becomes

$$a_{t|t}=a_{t|t-1}+\left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)\left(\bar{y}_t^G-Za_{t|t-1}\right), \tag{23}$$

which is only a function of group averages, and where the matrix to be inverted is of dimension $n\times n$. The expression for updating the covariance simplifies to

$$\begin{aligned} V_{t|t} &= V_{t|t-1}-\left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}Z'J_t'\left(I_{N_t}\otimes\Sigma^{-1}\right)J_tZV_{t|t-1} \\ &= \left[I_n-\left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right]V_{t|t-1} \\ &= \left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}\left[V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z-Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right]V_{t|t-1} \\ &= \left(V_{t|t-1}^{-1}+Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)Z\right)^{-1}. \end{aligned} \tag{24}$$

It is seen that (23) may now be rewritten as

$$a_{t|t}=a_{t|t-1}+V_{t|t}Z'\left(\mathcal{N}^G\otimes\Sigma^{-1}\right)\left(\bar{y}_t^G-Za_{t|t-1}\right). \tag{25}$$

A.2 Proof of expressions (18)

Assume that $\tilde{y}_t$ is constructed such that the first $N_1^g m$ elements belong to group 1, the following $N_2^g m$ elements to group 2 and so on. Define for each group $h \in (1, \ldots, G)$

$$J_h^g = \mathbf{1}_{N_h^g \times 1} \otimes J_h,$$

so that

$$E\tilde{y}_t | \mathcal{Y}_{t-1} = \begin{pmatrix} J_1^g \\ \vdots \\ J_G^g \end{pmatrix} Z a_{t|t-1}.$$

Then the upper left $N_1^g m \times N_1^g m$-block of Ω_t contains the covariance of the elements from group 1; call this sub-matrix Ω_t^1. The upper left $(N_1^g + N_2^g)\, m \times (N_1^g + N_2^g)\, m$-block contains the covariance between the elements from group 1 and 2; call this sub-matrix $\Omega_t^{1:2}$. Generally, the covariance matrix of the elements belonging to group 1 to h is

$$\Omega_t^{1:h} = J_{1:h}^g Z V_{t|t-1} Z' J_{1:h}^{g\prime} + I_{\left(N_1^g + \ldots + N_h^g\right)} \otimes \Sigma_t$$

where

$$J_{1:h}^g = \begin{pmatrix} J_1^g \\ \vdots \\ J_h^g \end{pmatrix}.$$

Hence for each $h \geq 1$

$$\Omega_t^{1:h+1} = \begin{pmatrix} \Omega_t^{1:h} & J_{1:h}^g Z V_{t|t-1} Z' J_{h+1}^{g\prime} \\ J_{h+1}^g Z V_{t|t-1} Z J_{1:h}^{g\prime} & J_{h+1}^g Z V_{t|t-1} Z' J_{h+}^{g\prime} + I_{N_{h+1}^g} \otimes \Sigma_t \end{pmatrix},$$

which means that

$$\left|\Omega_t^{1:h+1}\right| = \left|\Omega_g^{1:h}\right| \left| J_{h+1}^g Z V_{t|t-1} Z' J_{h+}^{g\prime} + I_{N_{h+1}^g} \otimes \Sigma_t - J_{1:h}^g Z V_{t|t-1} Z' J_{h+1}^{g\prime} \left(\Omega_t^{1:h}\right)^{-1} J_{h+1}^g Z V_{t|t-1} Z J_{1:h}^{g\prime} \right|. \quad (26)$$

Furthermore, the matrix inversion lemma yields

$$\begin{aligned} \left(\Omega_t^{1:h}\right)^{-1} &= I_{\left(N_1^g + \ldots + N_h^g\right)} \otimes \Sigma_t^{-1} - \\ &\quad \left(I_{\left(N_1^g + \ldots + N_h^g\right)} \otimes \Sigma_t^{-1}\right) J_{1:h}^g Z \left[V^{-1} + Z' J_{1:h}^{g\prime} \left(I_{\left(N_1^g + \ldots + N_h^g\right)} \otimes \Sigma_t^{-1}\right) J_{1:h}^g Z\right] \\ &\quad \times Z' J_{1:h}^{g\prime} \left(I_{\left(N_1^g + \ldots + N_h^g\right)} \otimes \Sigma_t^{-1}\right). \end{aligned}$$

Hence

$$
\begin{aligned}
& J^g_{1:h} Z V_{t|t-1} Z' J^{g\prime}_{h+1} \Omega_p^{-1} J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{1:h} \\
= \; & J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z V_{t|t-1} Z' J^{g\prime}_{h+1} \\
& - J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} \\
& \times Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z V_{t|t-1} Z' J^{g\prime}_{h+1} \\
= \; & J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z \\
& \left\{ I_n - \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z \right\} V_{t|t-1} Z' J^{g\prime}_{h+1} \\
= \; & J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J^{g\prime}_{h+1}.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
& J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{h+1} + I_{N^g_{h+1}} \otimes \Sigma_t - J^g_{1:h} Z V_{t|t-1} Z' J^{g\prime}_{h+1} \Omega_p^{-1} J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{1:h} \\
= \; & J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{h+1} \\
& + I_{N^g_{h+1}} \otimes \Sigma_t - J^g_{h+1} Z V_{t|t-1} Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J^{g\prime}_{h+1} \\
= \; & J^g_{h+1} Z V_{t|t-1} \left\{ I_n - Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} \right\} Z' J^{g\prime}_{h+1} + I_{N^g_{h+1}} \otimes \Sigma_t \\
= \; & J^g_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J^{g\prime}_{h+1} + I_{N^g_{h+1}} \otimes \Sigma_t.
\end{aligned}
$$

It is difficult to calculate the determinant of this expression directly, but a Gauss-Jordan transformation yields

$$
\begin{aligned}
& \left| J^g_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J^{g\prime}_{h+1} + I_{N^g_{h+1}} \otimes \Sigma_t \right| \\
= \; & \left| \mathbf{1}_{N^g_{h+1} \times N^g_{h+1}} \otimes J_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J'_{h+1} + I_{N^g_{h+1}} \otimes \Sigma_t \right| \\
= \; & \left| \begin{array}{ccc}
J_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J'_{h+1} + \Sigma_t & \vdots & \mathbf{1}_{1 \times N^g_{h+1}-1} \otimes J_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J'_{h+1} \\
\cdots\cdots & \cdots\cdots & \cdots\cdots \\
\mathbf{1}_{N^g_{h+1}-1 \times 1} \otimes (-\Sigma_t) & \vdots & I_{N^g_{h+1}-1} \otimes \Sigma_t
\end{array} \right| \\
= \; & \left| I_{N^g_{h+1}-1} \otimes \Sigma_t \right| \left| J_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J'_{h+1} + \Sigma_t \right. \\
& \left. - \left(N^g_{h+1} - 1\right) J_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J'_{h+1} \left(-\Sigma_t\right) \Sigma_t^{-1} \right| \\
= \; & |\Sigma_t|^{N^g_{h+1}-1} \left| N^g_{h+1} J_{h+1} Z \left[V^{-1}_{t|t-1} + Z' J^{g\prime}_{1:h} \left(I_p \otimes \Sigma_t^{-1}\right) J^g_{1:h} Z\right]^{-1} Z' J'_{h+1} + \Sigma_t \right|.
\end{aligned}
$$

Substituting into (26), we get

$$
\left|\Omega^{1:h+1}_t\right| = \left|\Omega^{1:h}_t\right| |\Sigma_t|^{N^g_{h+1}-1} \left| N^g_{h+1} J_{h+1} Z \left[V^{-1}_{t|t-1} + \sum_{i=1}^{h} N^g_i Z' J'_i \Sigma_t^{-1} J_i Z\right]^{-1} Z' J'_{h+1} + \Sigma_t \right|. \tag{27}
$$

Furthermore,

$$\begin{aligned}\left|\Omega_t^{11}\right| &= \left|\mathbf{1}_{N_1^g\times N_1^g}\otimes J_1 Z V_{t|t-1} Z' J_1' + I_{N_1^g}\otimes\Sigma_t\right| \\ &= \left|\begin{array}{cc} J_1 Z V_{t|t-1} Z' J_1' + \Sigma_t & \mathbf{1}_{1\times N_1^g-1}\otimes J_1 Z V_{t|t-1} Z' J_1' \\ \mathbf{1}_{N_1^g-1\times 1}\otimes(-\Sigma_t) & I_{N_1^g-1}\otimes\Sigma_t \end{array}\right| \\ &= \left|\Sigma_t\right|^{N_1^g-1}\left|N_1^g J_1 Z V_{t|t-1} Z' J_1' + \Sigma_t\right|.\end{aligned} \tag{28}$$

Consequently, we may rewrite $|\Omega_t|$ as

$$|\Omega_t| = |\Sigma_t|^{N_t-G}\left|N_1^g J_1 Z V_{t|t-1} Z' J_1' + \Sigma_t\right| \prod_{h=2}^{G}\left|N_{h+1}^g J_h Z\left[V_{t|t-1}^{-1} + \sum_{i=1}^{h-1} N_i^g Z' J_i' \Sigma_t^{-1} J_i Z\right]^{-1} Z' J_h' + \Sigma_t\right|, \tag{29}$$

which is clearly a tractable expression.

A.3 Proof of expressions (19)

Next, we want to simplify the expression for Ψ_t. Using the result from (21), we get

$$\begin{aligned}\Psi_t &= \left(\tilde{y}_t - J_t Z a_{t|t-1}\right)'\left(I_{N_t}\otimes\Sigma_t^{-1}\right)\left(\tilde{y}_t - J_t Z a_{t|t-1}\right) \\ &\quad - \left(\tilde{y}_t - J_t Z a_{t|t-1}\right)'\left(I_{N_t}\otimes\Sigma_t^{-1}\right) J_t Z\left[V_{t|t-1}^{-1} + Z' J_t'\left(I_{N_t}\otimes\Sigma_t^{-1}\right) J_t Z\right]^{-1} \\ &\quad \times Z' J_t'\left(I_{N_t}\otimes\Sigma_t^{-1}\right)\left(\tilde{y}_t - J_t Z a_{t|t-1}\right).\end{aligned}$$

Furthermore, $y_{it} - J_{g(i)} Z a_{t|t-1} = \left(y_{it} - \bar{y}_{g(i)t}\right) + \left(\bar{y}_{g(i)t} - J_{g(i)} Z a_{t|t-1}\right)$ where $\bar{y}_{gt}$ is the average value of y in group g at date t. Hence

$$\begin{aligned}&\left(\tilde{y}_t - J_t Z a_{t|t-1}\right)'\left(I_{N_t}\otimes\Sigma_t^{-1}\right)\left(\tilde{y}_t - J_t Z a_{t|t-1}\right) \\ =\ & \sum_{i=1}^{N_t}\left[\left(y_{it} - \bar{y}_{g(i)t}\right)'\Sigma_t^{-1}\left(yt - \bar{y}_{g(i)t}\right) + \left(\bar{y}_{g(i)t} - J_{g(i)} Z a_{t|t-1}\right)'\Sigma^{-1}\left(\bar{y}_{g(i)t} - J_{g(i)} Z a_{t|t-1}\right)\right] \\ =\ & \sum_{g=1}^{G}\operatorname{tr}\left[N_g^g \Sigma^{-1}\operatorname*{Cov}_{gt} y_{it}\right] + \left(\bar{y}_t^G - Z a_{t|t-1}\right)'\left(\mathcal{N}^G\otimes\Sigma_t^{-1}\right)\left(\bar{y}_t^G - Z a_{t|t-1}\right)\end{aligned}$$

where the last line uses the fact that the trace of a scalar is the scalar. >From (22) it follows that

$$\begin{aligned}&\left(\tilde{y}_t - J_t Z a_{t|t-1}\right)'\left(I_{N_t}\otimes\Sigma_t^{-1}\right) J_t Z\left[V_{t|t-1}^{-1} + Z' J_t'\left(I_{N_t}\otimes\Sigma_t^{-1}\right) J_t Z\right]^{-1} \\ &\times Z' J_t'\left(I_{N_t}\otimes\Sigma_t^{-1}\right)\left(\tilde{y}_t - J_t Z a_{t|t-1}\right) \\ =\ & \left(\bar{y}_t^G - Z a_{t|t-1}\right)'\left(\mathcal{N}^G\otimes\Sigma_t^{-1}\right) Z\left[V_{t|t-1}^{-1} + Z' J_t'\left(I_{N_t}\otimes\Sigma_t^{-1}\right) J_t Z\right]^{-1} \\ &\times Z'\left(\mathcal{N}^G\otimes\Sigma_t^{-1}\right)\left(\bar{y}_t^G - Z a_{t|t-1}\right).\end{aligned}$$

Consequently,

$$\begin{aligned}
\Psi_t &= \sum_{g=1}^{G} \operatorname{tr}\left[N_g^g \Sigma^{-1} \underset{gt}{\operatorname{Cov}}\, y_{it}\right] \\
&+ \left(\bar{y}_t^G - Z a_{t|t-1}\right)' \left(\mathcal{N}^G \otimes \Sigma_t^{-1}\right) \left\{ I_{Gm} - Z\left[V_{t|t-1}^{-1} + Z' J_t' \left(I_{N_t} \otimes \Sigma_t^{-1}\right) J_t Z\right]^{-1} \right. \qquad (30) \\
&\left. \times\, Z' \left(\mathcal{N}^G \otimes \Sigma_t^{-1}\right)\right\} \left(\bar{y}_t^G - Z a_{t|t-1}\right) \\
= &\sum_{g=1}^{G} \operatorname{tr}\left[N_g^g \Sigma^{-1} \underset{gt}{\operatorname{Cov}}\, y_{it}\right] \\
&+ \left(\bar{y}_t^G - Z a_{t|t-1}\right)' \left(\mathcal{N}^G \otimes \Sigma_t^{-1}\right) \left\{ I_{Gm} - Z\left[V_{t|t-1}^{-1} + Z' \left(\mathcal{N}^G \otimes \Sigma_t^{-1}\right) Z\right]^{-1} \right. \\
&\left. \times\, Z' \left(\mathcal{N}^G \otimes \Sigma_t^{-1}\right)\right\} \left(\bar{y}_t^G - Z a_{t|t-1}\right).
\end{aligned}$$

B Estimation by the EM-algorithm

An alternative approach to standard maximum likelihood estimation, which is very robust although somewhat slow, is the EM-algorithm developed by Dempster et al. (1977), introduced to the estimation of state space models by Engle and Watson (1983) and Shumway and Stoffer (1982). In some cases, this algorithm is superior to Simplex initially, but it should be supplemented with a more efficient algorithm when it starts converging. The idea of the EM-algorithm is to treat $\mathcal{A}_T \equiv (\alpha_1', \ldots . \alpha_T')$ as missing data. From an initial estimate Θ^0 of the hyper-parameters, we can use the Kalman smoother to obtain estimates of the latent $\mathcal{A}_T$. Instead of considering the ordinary likelihood function, the EM-algorithm employs the joint likelihood function, which for model (11) is

$$\begin{aligned}
L\left(\mathcal{Y}_T, \mathcal{A}_T; \Theta\right) = &-\frac{\sum_{t=1}^{T} N_t}{2} \ln\left(2\pi\right) - \frac{\sum_t N_t}{2} \ln\left|\Sigma\right| \\
&- \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left(y_{it} - J_{g(i)t} Z \alpha_t\right)' \Sigma_t^{-1} \left(y_{it} - J_{g(i)t} Z \alpha_t\right) \\
&- \frac{\sum_t N_t}{2} \ln\left|Q\right| - \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left(\alpha_t - F\alpha_{t-1}\right)' Q^{-1} \left(\alpha_t - F\alpha_{t-1}\right) \\
&- \frac{1}{2} \ln\left|Q_0\right| - \frac{1}{2} \left(\alpha_0 - a_0\right)' Q^{-1} \left(\alpha_0 - a_0\right).
\end{aligned}$$

Having obtained estimates of $\mathcal{A}_t$ from an estimate Θ^j,the next step in the algorithm is to maximize the expected joint likelihood function with regard to Θ. In this case, we get

$$
\begin{aligned}
&E\left[L\left(\mathcal{Y}_T,\mathcal{A}_T;\Theta\right)\left|\Theta^i\right.\right] \propto \qquad (31)\\
&-\frac{\sum_t N_t}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\operatorname{tr}\left\{\Sigma_t^{-1}\left[\left(y_{it}-J_{g(i)t}Za_{t|T}^j\right)\left(y_{it}-J_{g(i)t}Za_{t|T}^j\right)' + J_{g(i)t}ZV_{t|T}^jZ'J_{g(i)t}'\right]\right\}\\
&-\frac{\sum_t N_t}{2}\ln|Q| - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\operatorname{tr}\left\{Q^{-1}\left[\left(a_{t|T}^j-Fa_{t-1|T}^j\right)\left(a_{t|T}^j-Fa_{t-1|T}^j\right)'\right]\right\}\\
&+V_{t|T}^j + FV_{t-1|T}^jF' - FB_t^jV_{t|T}^j - V_{t|T}^jB_t^jF'\Big]\\
&-\frac{1}{2}\ln|Q_0| - \frac{1}{2}\operatorname{tr}\left\{Q^{-1}\left[\left(a_0-a_{0|T}^j\right)\left(a_0-a_{0|T}^j\right)' + V_{0|T}^j\right]\right\}
\end{aligned}
$$

where $B_t^j = V_{t-1|t-1}^jF'V_{t|t-1}^{j-1}$ and the parameters with superscript j are estimates from the Kalman smoother conditional on Θ^j, the hyper-parameters from the j'th iteration of the EM-algorithm. Calculating the first order conditions and simplifying, we obtain a new set of parameters Θ^{j+1}:

$$
\begin{aligned}
\Sigma_t^{j+1} &= \frac{1}{N_t}\sum_{i=1}^{N_t}\left[\left(y_{it}-J_{g(i)t}Za_{t|T}^j\right)\left(y_{it}-J_{g(i)t}Za_{t|T}^j\right)' + J_{g(i)t}ZV_{t|T}^jZ'J_{g(i)t}'\right] \qquad (32)\\
&= \sum_{g=1}^{G}\frac{N_g^g}{N_t}\left[\underset{gt}{\operatorname{Cov}}\left(y_{it}\right) + \left(\bar{y}_{gt}-J_{g(i)t}Za_{t|T}^j\right)\left(\bar{y}_t-J_{g(i)t}Za_{t|T}^j\right)' + J_{gt}ZV_{t|T}^jZ'J_{gt}'\right]
\end{aligned}
$$

$$
\begin{aligned}
Q^{j+1} &= \frac{1}{\sum_t N_t}\sum_{t=1}^{T}N_t\left[\left(a_{t|T}^j-Fa_{t-1|T}^j\right)\left(a_{t|T}^j-Fa_{t-1|T}^j\right)' \qquad (33)\right.\\
&\quad \left. +V_{t|T}^j + FV_{t-1|T}^jF' - FB_t^jV_{t|T}^j - V_{t|T}^jB_t^jF'\right]\\
a_0^{j+1} &= a_{0|T}^j \qquad Q_0^{j+1} = V_{0|T}^j \qquad (34)
\end{aligned}
$$

If Σ is time-invariant, an obvious estimator is

$$\Sigma^{j+1} = \frac{1}{\sum_{t=1}^{T}N_t}\sum_{t=1}^{T}N_t\Sigma_t^{j+1}.$$

We can then go on to calculate a new estimate of $\mathcal{A}_t$, a new expression for the expected joint likelihood value from (31), and then calculate new estimates of the hyper-parameters from (32-34). As shown by Dempster at al. (1977), each step in this iteration will increase the likelihood value, and the estimated hyper-parameters will converge towards a local maximum of the likelihood function.

It is clear that consistent estimates of a_0 and Q_0 are not available since we do not gain further information on these parameters from a longer time series. Also, it seems that Q_0 is not well identified since it tends towards zero in most applications of the algorithm. Following Shumway and Stoffer (1982: 257), it is then probably advisable to choose a reasonable value for Q_0 rather than trying to estimate it.