Standard Errors Estimation in Mixtures of Partial Likelihood Models

Stéphane Bonhomme*
CEMFI, Madrid

March, 2006

Abstract

I consider the sequential Expectation-Maximization (EM) model of Arcidiacono and Jones (2003). Using a generalized version of the information matrix equality, I obtain an alternative estimate of the variance-covariance matrix of the parameters. In practice, using this formula can result in significant gains in computing time.

*CEMFI, Casado del Alisal, 5, 28014 Madrid, Spain. E-mail: bonhomme@cemfi.es

1 The model and the estimator

The model. Consider a dataset of N individuals and T time periods. Observations $y_i = (y_{i1}, ..., y_{iT})'$ are iid across individuals. The unconditional individual likelihood is given by:

$$g(y_i; \Theta; \pi) = \sum_{k=1}^{K} \pi_k f_k(y_i; \Theta). \tag{1}$$

In this expression, $k = 1...K$ indexes K unobserved groups, which are associated to probabilities $\pi = (\pi_1 ... \pi_K)'$. The group-specific likelihoods are denoted by $f_k(y_i; \Theta)$, $k = 1...K$.

Assumption 1 *There exists a partition $\Theta = (\Theta_1, \Theta_2)$ and functions f_{1k}, f_{2k}, $k = 1...K$ such that, for all sets of parameters Θ_1, Θ_2 and $\widetilde{\Theta}_1$, and all $k = 1...K$, the following equalities hold:*

$$f_k(y_i; \Theta) = f_{1k}(y_i; \Theta_1) f_{2k}(y_i; \Theta_1, \Theta_2), \tag{2}$$

$$\int f_{1k}(y; \Theta_1) f_{2k}(y; \widetilde{\Theta}_1, \Theta_2) dy = 1. \tag{3}$$

Equations (2)-(3) imply that the type-conditional likelihoods f_k are partial likelihoods in the sense of Cox (1975).

The estimator. The maximum likelihood estimator of Θ, π satisfies:

$$\left(\widehat{\Theta}, \widehat{\pi}\right) = \underset{\Theta, \pi, \sum_{k=1}^{K} \pi_k = 1}{\text{Argmax}} \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} \pi_k f_k(y_i; \Theta) \right).$$

Then it is well-known that $\left(\widehat{\Theta}, \widehat{\pi}\right)$ satisfies:

$$\widehat{\pi}_k = \frac{1}{N} \sum_{i=1}^{N} \Pr\left(k | y_i; \widehat{\Theta}, \widehat{\pi}\right), \quad k = 1...K, \tag{4}$$

$$\widehat{\Theta} = \underset{\Theta}{\text{Argmax}} \sum_{i=1}^{N} \sum_{k=1}^{K} \Pr\left(k | y_i; \widehat{\Theta}, \widehat{\pi}\right) \ln f_k(y_i; \Theta), \tag{5}$$

where the posterior probabilities of the groups are given by the Bayes formula:

$$\Pr\left(k | y_i; \Theta, \pi\right) = \frac{\pi_k f_k(y_i; \Theta)}{\sum_{k=1}^{K} \pi_k f_k(y_i; \Theta)}. \tag{6}$$

The EM algorithm of Dempster, Laird and Rubin (1977) is an iterative method to estimate $\left(\widehat{\Theta}, \widehat{\pi}\right)$ using equations(4)-(6). Let $\left(\widehat{\Theta}^{(s)}, \widehat{\pi}^{(s)}\right)$ be the estimate obtained after the (s)th iteration of the algorithm. Then $\left(\widehat{\Theta}^{(s+1)}, \widehat{\pi}^{(s+1)}\right)$ are estimated in two steps:

1. Expectation (E) step: compute the posterior probabilities: $\Pr\left(k|y_i;\widehat{\Theta}^{(s)},\widehat{\pi}^{(s)}\right)$ by (6).

2. Maximization (M) step: update the parameters

$$\begin{aligned}
\widehat{\pi}_k^{(s+1)} &= \frac{1}{N}\sum_{i=1}^{N}\Pr\left(k|y_i;\widehat{\Theta}^{(s)},\widehat{\pi}^{(s)}\right), \quad k=1...K, \\
\widehat{\Theta}^{(s+1)} &= \underset{\Theta}{\text{Argmax}}\sum_{i=1}^{N}\sum_{k=1}^{K}\Pr\left(k|y_i;\widehat{\Theta}^{(s)},\widehat{\pi}^{(s)}\right)\ln f_k(y_i;\Theta). \qquad (7)
\end{aligned}$$

These two step are iterated until numerical convergence.

In the case where type-conditional likelihoods are partial likelihoods, Arcidiacono and Jones (2003) propose to replace (7) by the sequential procedure:

$$\widehat{\Theta}_1^{(s+1)} = \underset{\Theta_1}{\text{Argmax}}\sum_{i=1}^{N}\sum_{k=1}^{K}\Pr\left(k|y_i;\widehat{\Theta}^{(s)},\widehat{\pi}^{(s)}\right)\ln f_{1k}(y_i;\Theta_1), \qquad (8)$$

$$\widehat{\Theta}_2^{(s+1)} = \underset{\Theta_2}{\text{Argmax}}\sum_{i=1}^{N}\sum_{k=1}^{K}\Pr\left(k|y_i;\widehat{\Theta}^{(s)},\widehat{\pi}^{(s)}\right)\ln f_{2k}(y_i;\Theta_1^{(s)},\Theta_2). \qquad (9)$$

They show that the resulting sequential EM algorithm provides consistent estimates of Θ and π under suitable regularity conditions.

2 Standard errors

Sandwich formula. Arcidiacono and Jones show that the sequential EM estimator can be seen as a GMM-type estimator based on the following population moment conditions, satisfied at true values of the parameters:

$$\mathbb{E}\left(\Pr\left(k|y_i;\Theta^0,\pi^0\right)-\pi_k^0\right) = 0, \quad k=1...K, \qquad (10)$$

$$\mathbb{E}\left(\sum_{k=1}^{K}\Pr\left(k|y_i;\Theta^0,\pi^0\right)\frac{\partial\ln f_{1k}(y_i;\Theta_1^0)}{\partial\Theta_1}\right) = 0, \qquad (11)$$

$$\mathbb{E}\left(\sum_{k=1}^{K}\Pr\left(k|y_i;\Theta^0,\pi^0\right)\frac{\partial\ln f_{2k}(y_i;\Theta_1^0,\Theta_2^0)}{\partial\Theta_2}\right) = 0. \qquad (12)$$

Hence the standard GMM theory can be applied to obtain the asymptotic variance-covariance matrix of the estimator Σ. Precisely, let $\beta = (\Theta, \pi)$ and:

$$\Psi(y_i;\beta) \quad = \quad \begin{pmatrix} \Pr(k=1|y_i;\Theta,\pi) - \pi_1 \\ \dots \quad \dots \quad \dots \\ \Pr(k=K|y_i;\Theta,\pi) - \pi_K \\ \sum_{k=1}^{K} \Pr(k|y_i;\Theta,\pi) \frac{\partial \ln f_{1k}(y_i;\Theta_1)}{\partial \Theta_1} \\ \sum_{k=1}^{K} \Pr(k|y_i;\Theta,\pi) \frac{\partial \ln f_{2k}(y_i;\Theta_1,\Theta_2)}{\partial \Theta_2} \end{pmatrix}.$$

Then:

$$\mathbb{E}\left(\Psi\left(y_i;\beta^0\right)\right) = 0.$$

It follows from classical arguments (e.g. Newey and McFadden, 1994) that

$$\Sigma = \mathbb{E}\left(\frac{\partial \psi\left(y_i;\beta^0\right)}{\partial \beta'}\right)^{-1} \mathbb{E}\left(\psi\left(y_i;\beta^0\right)\psi\left(y_i;\beta^0\right)'\right) \mathbb{E}\left(\frac{\partial \psi\left(y_i;\beta^0\right)}{\partial \beta'}\right)^{-1}. \tag{13}$$

This is the well-known "sandwich" formula. Note that, unlike in the case of maximum likelihood, their is no obvious simplification in this expression and estimating standard errors requires computing the moment conditions and their derivatives.

An alternative expression. We here show that a simplification analogous to the likelihood case holds for mixtures of partial likelihoods. We first prove the following lemma, which comes directly from Assumption 1.

Lemma 2 *For all parameter $\beta = (\Theta, \pi)$:*

$$\int \Psi(y;\beta)\, g(y;\beta) dy = 0, \tag{14}$$

where $g(y_i;\Theta,\pi)$ is the likelihood given by (1).

Proof. See Appendix A. ■

Equation (14) is clearly satisfied at true values of the parameters. Lemma 2 shows that it is satisfied at every set of parameters β. Newey and McFadden (1994) make the point that such a property has to be satisfied by all consistent GMM estimator.

Differentiating (14) with respect to β and evaluating at β^0 yields:

$$\int \frac{\partial \Psi\left(y;\beta^0\right)}{\partial \beta'} g(y;\beta^0) dy + \int \Psi\left(y;\beta^0\right) \frac{\partial g(y;\beta^0)}{\partial \beta'} dy = 0.$$

Hence:

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial\Psi\left(y;\beta^0\right)}{\partial\beta'}\right) &= \int \frac{\partial\Psi\left(y;\beta^0\right)}{\partial\beta'}g(y;\beta^0)dy, \\
&= -\int \Psi\left(y;\beta^0\right)\frac{\partial g(y;\beta^0)}{\partial\beta'}dy, \\
&= -\int \Psi\left(y;\beta^0\right)\frac{\partial \ln g(y;\beta^0)}{\partial\beta'}g(y;\beta^0)dy, \\
&= -\mathbb{E}\left(\Psi\left(y;\beta^0\right)\frac{\partial \ln g(y;\beta^0)}{\partial\beta'}\right). \qquad (15)
\end{aligned}$$

Equation (15) is a generalized information matrix equality. Moreover, it turns out that $\frac{\partial \ln g(y;\beta^0)}{\partial\beta'}$ has a simple expression in the case of mixtures of partial likelihoods. We prove the following lemma in the appendix.

Lemma 3 *The following equalities hold:*

$$\frac{\partial \ln g(y;\beta^0)}{\partial\pi_k} = \frac{\Pr\left(k=1|y_i;\Theta^0,\pi^0\right)}{\pi_k^0}, \quad k=1...K, \qquad (16)$$

$$\frac{\partial \ln g(y;\beta^0)}{\partial\Theta_1} = \sum_{k=1}^{K}\Pr\left(k=1|y_i;\Theta^0,\pi^0\right)\frac{\partial\left(\ln f_{1k}(y_i;\Theta_1^0)+\ln f_{2k}(y_i;\Theta_1^0,\Theta_2^0)\right)}{\partial\Theta_1}, \qquad (17)$$

$$\frac{\partial \ln g(y;\beta^0)}{\partial\Theta_2} = \sum_{k=1}^{K}\Pr\left(k=1|y_i;\Theta^0,\pi^0\right)\frac{\partial \ln f_{2k}(y_i;\Theta_1^0,\Theta_2^0)}{\partial\Theta_2}. \qquad (18)$$

Proof. See Appendix B. ■

Combining (13), (15) and the results in lemma 2 yield an expression for Σ which depends on first derivatives of the likelihood and posterior probabilities only.

3 Comments

The homogeneous case $K=1$ is illustrative. Then the previous result imply:

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial\Psi\left(y;\beta^0\right)}{\partial\beta'}\right) &= -\mathbb{E}\left(\Psi\left(y;\beta^0\right)\frac{\partial \ln g(y;\beta^0)}{\partial\beta'}\right), \\
&= -\mathbb{E}\left(\begin{pmatrix}\frac{\partial \ln f_1(y_i;\Theta_1^0)}{\partial\Theta_1} \\ \frac{\partial \ln f_2(y_i;\Theta_1^0,\Theta_2^0)}{\partial\Theta_2}\end{pmatrix}\begin{pmatrix}\frac{\partial\left(\ln f_1(y_i;\Theta_1^0)+\ln f_2(y_i;\Theta_1^0,\Theta_2^0)\right)}{\partial\Theta_1'} & \frac{\partial \ln f_2(y_i;\Theta_1^0,\Theta_2^0)}{\partial\Theta_2'}\end{pmatrix}\right) \\
&= -\mathbb{E}\left(\Psi\left(y;\beta^0\right)\Psi\left(y;\beta^0\right)'\right) - \mathbb{E}\begin{pmatrix}\frac{\partial \ln f_1(y_i;\Theta_1^0)}{\partial\Theta_1}\frac{\ln f_2(y_i;\Theta_1^0,\Theta_2^0)}{\partial\Theta_1'} & 0 \\ \frac{\partial \ln f_2(y_i;\Theta_1^0,\Theta_2^0)}{\partial\Theta_2'}\frac{\partial \ln f_2(y_i;\Theta_1^0,\Theta_2^0)}{\partial\Theta_1'} & 0\end{pmatrix}.
\end{aligned}$$

4 Conclusion

In problems where the number of parameters is large and the (sequential) likelihood is not straightforward to compute, this idea provides a fast alternative way to compute standard errors.

References

[1] Arcidiacono, P., and J. Jones (2003): "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, 71(3), 933-46.

[2] Cox, D. (1975): "Partial Likelihood," *Biometrika*, 62(2), 269-276.

[3] Dempster, A., N. Laird and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B 39(1), 1-38.

[4] Newey, W., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," *in* R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics* vol 4: 2111-245. Amsterdan: Elsevier Science.

APPENDIX

A Proof of lemma 2

In this section, we assume that some parameters Θ and π are given.

Let $k \in \{1...K\}$. Then

$$\begin{aligned}\int \left(\Pr\left(k|y;\Theta,\pi\right) - \pi_k\right) g(y;\Theta,\pi)dy &= \int \Pr\left(k|y;\Theta,\pi\right) g(y;\Theta,\pi)dy - \pi_k, \\ &= \int \pi_k f_{1k}(y;\Theta_1) f_{1k}(y;\Theta_1,\Theta_2)dy, \\ &= 0.\end{aligned}$$

Here, as in the rest of the Apendix, we make use of equation (6) that defines posterior probabilities by the Bayes formula.

Then

$$\begin{aligned}\int \left(\sum_{k=1}^{K} \Pr\left(k|y;\Theta,\pi\right) \frac{\partial \ln f_{1k}(y_i;\Theta_1)}{\partial \Theta_1}\right) g(y;\Theta,\pi)dy &= \sum_{k=1}^{K} \int \pi_k \frac{\partial \ln f_{1k}(y;\Theta_1)}{\partial \Theta_1} f_{1k}(y;\Theta_1) f_{2k}(y;\Theta_1,\Theta_2)dy, \\ &= \sum_{k=1}^{K} \pi_k \int \frac{\partial f_{1k}(y;\Theta_1)}{\partial \Theta_1} f_{2k}(y;\Theta_1,\Theta_2)dy.\end{aligned}$$

Now, for all $k \in \{1...K\}$, differentiating (3) with respect to Θ_1 and evaluating at (Θ_1, Θ_2) yields:

$$\int \frac{\partial f_{1k}(y;\Theta_1)}{\partial \Theta_1} f_{2k}(y;\Theta_1,\Theta_2)dy = 0.$$

Lastly:

$$\int \left(\sum_{k=1}^{K} \Pr\left(k|y;\Theta,\pi\right) \frac{\partial \ln f_{2k}(y_i;\Theta_1,\Theta_2)}{\partial \Theta_2}\right) g(y;\Theta,\pi)dy = \sum_{k=1}^{K} \pi_k \int f_{1k}(y;\Theta_1) \frac{\partial f_{2k}(y;\Theta_1,\Theta_2)}{\partial \Theta_2} dy.$$

Likewise, differentiating (3) with respect to Θ_2 and evaluating at (Θ_1, Θ_2) yields, for all k:

$$\int f_{1k}(y;\Theta_1) \frac{\partial f_{2k}(y;\Theta_1,\Theta_2)}{\partial \Theta_2} dy = 0.$$

B Proof of lemma 3

Let $k \in \{1...K\}$. Then

$$\begin{aligned}\frac{\partial \ln g(y;\beta^0)}{\partial \pi_k} &= \frac{1}{g(y;\beta^0)} \frac{\partial g(y;\beta^0)}{\partial \pi_k}, \\ &= \frac{1}{g(y;\beta^0)} \frac{\partial}{\partial \pi_k}\Big|_{\beta^0} \left(\sum_{k=1}^{K} \pi_k f_{1k}(y;\Theta_1) f_{2k}(y;\Theta_1,\Theta_2)\right), \\ &= \frac{f_{1k}(y;\Theta_1^0) f_{2k}(y;\Theta_1^0,\Theta_2^0)}{g(y;\beta^0)}, \\ &= \frac{\Pr\left(k=1|y;\Theta^0,\pi^0\right)}{\pi_k^0}, \quad k = 1...K.\end{aligned}$$

Next:

$$\begin{aligned}
\frac{\partial \ln g(y;\beta^0)}{\partial \Theta_1} &= \frac{1}{g(y;\beta^0)} \frac{\partial}{\partial \Theta_1}\Big|_{\beta^0} \left(\sum_{k=1}^{K} \pi_k f_{1k}(y;\Theta_1) f_{2k}(y_i;\Theta_1,\Theta_2) \right), \\
&= \frac{1}{g(y;\beta^0)} \sum_{k=1}^{K} \pi_k \frac{\partial \left(f_{1k}(y;\Theta_1) f_{2k}(y;\Theta_1,\Theta_2) \right)}{\partial \Theta_1}, \\
&= \frac{1}{g(y;\beta^0)} \sum_{k=1}^{K} \pi_k \frac{\partial \left(\ln f_{1k}(y;\Theta_1) + \ln f_{2k}(y;\Theta_1,\Theta_2) \right)}{\partial \Theta_1} f_{1k}(y;\Theta_1^0) f_{2k}(y;\Theta_1^0,\Theta_2^0), \\
&= \sum_{k=1}^{K} \Pr\left(k=1|y;\Theta^0,\pi^0\right) \frac{\partial \left(\ln f_{1k}(y;\Theta_1^0) + \ln f_{2k}(y;\Theta_1^0,\Theta_2^0) \right)}{\partial \Theta_1},
\end{aligned}$$

Lastly:

$$\begin{aligned}
\frac{\partial \ln g(y;\beta^0)}{\partial \Theta_2} &= \frac{1}{g(y;\beta^0)} \frac{\partial}{\partial \Theta_2}\Big|_{\beta^0} \left(\sum_{k=1}^{K} \pi_k f_{1k}(y;\Theta_1) f_{2k}(y_i;\Theta_1,\Theta_2) \right), \\
&= \frac{1}{g(y;\beta^0)} \sum_{k=1}^{K} \pi_k f_{1k}(y;\Theta_1^0) \frac{\partial f_{2k}(y;\Theta_1^0,\Theta_2^0)}{\partial \Theta_2}, \\
&= \sum_{k=1}^{K} \Pr\left(k=1|y_i;\Theta^0,\pi^0\right) \frac{\partial \ln f_{2k}(y_i;\Theta_1^0,\Theta_2^0)}{\partial \Theta_2}.
\end{aligned}$$