Open Journal of Statistics, 2014, 4, 851-856
Published Online December 2014 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2014.410080

Estimation of Multivariate Sample Selection Models via a Parameter-Expanded Monte Carlo EM Algorithm

Phillip Li

Department of Economics, Office of the Comptroller of the Currency, Washington, DC, USA
Email: Phillip.Li@occ.treas.gov

Received 6 September 2014; revised 5 October 2014; accepted 2 November 2014

Abstract

This paper develops a parameter-expanded Monte Carlo EM (PX-MCEM) algorithm to perform maximum likelihood estimation in a multivariate sample selection model. In contrast to the current methods of estimation, the proposed algorithm does not directly depend on the observed-data likelihood, the evaluation of which requires intractable multivariate integrations over normal densities. Moreover, the algorithm is simple to implement and involves only quantities that are easy to simulate or have closed form expressions.

Keywords

Multivariate Sample Selection, Heckman Correction, Incidental Truncation, Expectation Maximization

1. Introduction

Sample selection models, pioneered in [1]-[3], are indispensable to researchers who use observational data for statistical inference. Among the many variants of these types of models, there is a growing interest in multivariate sample selection models. These are used to model a system of two or more seemingly unrelated equations, where the outcome variable for each equation may be non-randomly missing or censored according to its own stochastic selection variable. Applications range from modeling systems of demand equations [4] [5] to household vehicle usage [6]-[8]. A common specification is to assume a correlated multivariate normal distribution underlying both the outcomes of interest and the latent variables in the system.

There are two dominant approaches in the current literature to estimate these models. One approach is to use

maximum likelihood (ML) estimation. However, as noted in the literature, a major hurdle in evaluating the likelihood is that it requires computations of multivariate integrals over normal densities, which do not generally have closed form solutions. [9] discusses the ML estimation of these models and proposes to use the popular Geweke, Hajivassiliou, and Keane (GHK) algorithm to approximate these integrals in a simulated ML framework. While this strategy works reasonably well, the GHK algorithm can be difficult to implement. Another popular approach is to use two-step estimation (see [10] for a survey). In general, there is a tradeoff in the statistical properties and the computational simplicity for these estimators. If efficiency and consistency are of primary concern, then ML estimation should be preferred over two-step estimation.

The objective of this paper is to develop a simple ML estimation algorithm for a commonly used multivariate sample selection model. In particular, this paper develops a parameter-expanded Monte Carlo expectation maximization (PX-MCEM) algorithm that differs from [9] in a few important ways. First, the PX-MCEM algorithm does not use the observed-data likelihood directly, so it avoids the aforementioned integrations. Second, the proposed iterative algorithm does not require the evaluations of gradients or Hessians, which become increasingly difficult to evaluate with more parameters and equations. Third, the algorithm is straightforward to implement. It only depends on quantities that are either easy to simulate or have closed form expressions. This last point is especially appealing when estimating the covariance matrix parameter since there are non-standard restrictions imposed onto it for identification.

This paper is organized as follows. The multivariate sample selection model (MSSM) is formulated in Section 2. Section 3 begins with a brief overview of the EM algorithm for the MSSM and continues with the development of the PX-MCEM algorithm. Methods to obtain the standard errors are discussed. Section 4 offers some concluding remarks.

2. Multivariate Sample Selection Model

The MSSM is

$$y_{i,j}^* = x_{i,j}'\beta_j + \epsilon_{i,j} \tag{1}$$

$$s_{i,j}^* = w_{i,j}'\gamma_j + \nu_{i,j} \tag{2}$$

$$s_{i,j} = \mathbb{I}\left(s_{i,j}^* > 0\right) \tag{3}$$

$$y_{i,j} = \begin{cases} y_{i,j}^* & \text{if } s_{i,j} = 1 \\ \text{missing} & \text{if } s_{i,j} = 0 \end{cases} \tag{4}$$

for observations $i = 1, \cdots, N$, and equations $j = 1, \cdots, J$. In the previous expressions, $y_{i,j}^*$ is the continuous outcome of interest for observation i and equation j. Using similar indexing notation, $s_{i,j}^*$ is the latent variable underlying the binary selection variable $s_{i,j} = \mathbb{I}\left(s_{i,j}^* > 0\right)$, where $\mathbb{I}(A)$ denotes an indicator function that equals 1 if event A is true and 0 otherwise. Sample selection is incorporated by assuming that $y_{i,j}^*$ is missing when $s_{i,j} = 0$. Otherwise, $y_{i,j}^*$ is observed and equal to $y_{i,j}$. For later use, define $s_i = \left(s_{i,1}, \cdots, s_{i,J}\right)'$ and $s_i^* = \left(s_{i,1}^*, \cdots, s_{i,J}^*\right)'$, where the prime symbol in s_i, s_i^*, and in the rest of this paper is used to denote matrix transpose.

Furthermore, $x_{i,j}$ and $w_{i,j}$ are column vectors of exogenous covariates, and β_j and γ_j are conforming vectors of parameters. Define $\beta = \left(\beta_1', \beta_2', \cdots, \beta_J'\right)'$ and $\gamma = \left(\gamma_1', \gamma_2', \cdots, \gamma_J'\right)'$. For identification, $w_{i,j}$ must contain at least one exogenous covariate that does not overlap with $x_{i,j}$ (refer to [11] for these exclusion restrictions). The unobserved errors $\epsilon_i = \left(\epsilon_{i,1}, \epsilon_{i,2}, \cdots, \epsilon_{i,J}\right)'$ and $\nu_i = \left(\nu_{i,1}, \nu_{i,2}, \cdots, \nu_{i,J}\right)'$ are jointly distributed as a $2J$-dimensional multivariate normal with a mean vector of zeros and an unknown covariance matrix of Ω. Formally, $\left(\epsilon_i', \nu_i'\right)' \overset{iid}{\sim} \mathcal{N}_{2J}(0, \Omega)$ with

$$\Omega = \begin{pmatrix} \underset{(J\times J)}{\Omega_{\epsilon\epsilon}} & \Omega_{\epsilon\nu} \\ \Omega'_{\epsilon\nu} & \underset{(J\times J)}{\Omega_{\nu\nu}} \end{pmatrix}. \tag{5}$$

The submatrix $\Omega_{\nu\nu}$ is restricted to be in correlation form to identify the parameters corresponding to the latent variables [9]. The other elements of Ω are restricted such that the matrix is symmetric and positive definite.

The covariates and binary selection variables are always observed. Without loss of generality, assume that the outcomes for any observation i are only missing for the first m_i equations, where $0 \le m_i \le J$. Define $y_{i,\text{obs}} = \left(y_{i,m_i+1}, \cdots, y_{i,J}\right)'$, and let $y_{\text{obs}} = \left\{y_{i,\text{obs}}, s_i\right\}_{i=1}^{N}$ denote the observed data. The observed-data likelihood derived from (1) through (5) is denoted as $f\left(y_{\text{obs}} \middle| \beta, \gamma, \Omega\right)$. See [9] for an exact expression of this likelihood.

3. Estimation

3.1. Overview of the EM Algorithm

The PX-MCEM algorithm is based on the EM algorithm of [12]. The basic idea behind the EM algorithm is to first augment y_{obs} with a set of "missing data" y_{mis} such that the observed-data likelihood is preserved when the missing data are integrated out of the complete-data likelihood. Formally, the missing data must satisfy

$$f\left(y_{\text{obs}} \middle| \beta, \gamma, \Omega\right) = \mathbb{E}\left[f\left(y_{\text{mis}}, y_{\text{obs}} \middle| \beta, \gamma, \Omega\right)\right], \tag{6}$$

where $f\left(y_{\text{mis}}, y_{\text{obs}} \middle| \beta, \gamma, \Omega\right)$ is the complete-data likelihood to be defined later.

The EM algorithm then proceeds iteratively between an expectation step (E-step) and a maximization step (M-step) as follows. In iteration $(t+1)$ of the algorithm, compute in the E-step

$$Q\left(\beta, \gamma, \Omega \middle| \beta^{(t)}, \gamma^{(t)}, \Omega^{(t)}\right) = \mathbb{E}\left[\log\left(f\left(y_{\text{mis}}, y_{\text{obs}} \middle| \beta, \gamma, \Omega\right)\right)\right], \tag{7}$$

where the expectation is taken with respect to the conditional predictive distribution for the missing data, $\pi\left(y_{\text{mis}}, y_{\text{obs}} \middle| \beta^{(t)}, \gamma^{(t)}, \Omega^{(t)}\right)$, and in the M-step, find

$$\underset{\beta,\gamma,\Omega}{\arg\max}\, Q\left(\beta, \gamma, \Omega \middle| \beta^{(t)}, \gamma^{(t)}, \Omega^{(t)}\right). \tag{8}$$

Denote the maximal values as $\beta^{(t+1)}$, $\gamma^{(t+1)}$, and $\Omega^{(t+1)}$, and continue on with the algorithm until convergence. The final maximal values are at least local maxima of the observed-data likelihood function.

For the MSSM, y_{mis} consists of all the missing outcomes and latent variables. Specifically, $y_{\text{mis}} = \left\{y_{i,\text{mis}}, s_i^*\right\}_{i=1}^{N}$, where $y_{i,\text{mis}} = \left(y_{i,1}^*, \cdots, y_{i,m_i}^*\right)'$. Furthermore, denote $y_{i,\text{com}} = \left(y'_{i,\text{mis}}, y'_{i,\text{obs}}, s_i^{*\prime}\right)'$ as the vector of complete data, X_i as a block-diagonal matrix with the rows of covariates corresponding to the elements of $y_{i,\text{com}}$ on its block diagonals, and $\theta = \left(\beta', \gamma'\right)'$. The complete-data likelihood for the MSSM is given by

$$f\left(y_{\text{mis}}, y_{\text{obs}} \middle| \beta, \gamma, \Omega\right) = \prod_{i=1}^{N} f\left(y_{i,\text{com}} \middle| \beta, \gamma, \Omega\right) p\left(s_i \middle| y_{i,\text{com}}\right) \tag{9}$$

with $f\left(y_{i,\text{com}} \middle| \beta, \gamma, \Omega\right) = \phi_{2J}\left(y_{i,\text{com}} \middle| X_i\theta, \Omega\right)$ which is a density function for a $2J$-dimensional multivariate normal with mean $X_i\theta$ and covariance Ω, and

$$p\left(s_i \middle| y_{i,\text{com}}\right) = \prod_{j=1}^{J}\left\{\mathbb{I}\left(s_{i,j} = 1\right)\mathbb{I}\left(s_{i,j}^* > 0\right) + \mathbb{I}\left(s_{i,j} = 0\right)\mathbb{I}\left(s_{i,j}^* \le 0\right)\right\}. \tag{10}$$

Equation (10) is a degenerate density since conditioning on s_i^* in $y_{i,\text{com}}$ determines s_i from (3). Note that the observed-data likelihood from [9] is obtained when y_{mis} is integrated out of (9), hence the condition in (6) holds.

3.2. PX-MCEM Algorithm

The standard EM algorithm using (7) and (8) is difficult to implement for the MSSM as the E-step and M-step are intractable. The PX-MCEM algorithm addresses this issue by modifying the E-step in two ways and leads to an M-step that can be evaluated with closed form quantities. Stated succinctly, the PX-MCEM algorithm is as follows.

1. Initialize $\beta^{(0)}$, $\gamma^{(0)}$, $\Omega^{(0)}$, and the number of Gibbs sampling draws G.

At iteration $t+1$:

2. Draw G sets of missing data, denoted by $y_{\text{mis}}^{(1)},\cdots,y_{\text{mis}}^{(G)}$, from $\pi\left(y_{\text{mis}}\middle|y_{\text{obs}},\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)$ using Gibbs sampling.

3. PX-MC E-step: Estimate $Q\left(\alpha,\delta,\Sigma\middle|\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)$ as

$$Q_G\left(\alpha,\delta,\Sigma\middle|\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)=\frac{1}{G}\sum_{g=1}^{G}\log\left(f\left(y_{\text{mis}}^{(g)},y_{\text{obs}}\middle|\alpha,\delta,\Sigma\right)\right). \tag{11}$$

4. PX-MC M-step: Maximize $Q_G\left(\alpha,\delta,\Sigma\middle|\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)$ with iterative generalized least squares (IGLS) to obtain the maximizing parameters $\alpha^{(t+1)}$, $\delta^{(t+1)}$, and $\Sigma^{(t+1)}$.

5. Reduction step: Apply reduction functions to $\alpha^{(t+1)}$, $\delta^{(t+1)}$, and $\Sigma^{(t+1)}$ to obtain $\beta^{(t+1)}$, $\gamma^{(t+1)}$, and $\Omega^{(t+1)}$.

6. Repeat Steps 2 through 5 until convergence. The converged values are the ML estimates $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\Omega}$.

Each step is described in more detail in the subsequent sections.

3.2.1. PX-MC E-Step

Following [13], the first modification is to expand the parameter space of the complete-data likelihood function from (β,γ,Ω) to (α,δ,Σ). The expanded parameters play similar roles as the original parameters, however Σ is expanded into a standard covariance matrix without the correlation restrictions. The parameter-expanded complete-data likelihood function is

$$f\left(y_{\text{mis}},y_{\text{obs}}\middle|\alpha,\delta,\Sigma\right)=\prod_{i=1}^{N}f\left(y_{i,\text{com}}\middle|\alpha,\delta,\Sigma\right)p\left(s_i\middle|y_{i,\text{com}}\right) \tag{12}$$

with $f\left(y_{i,\text{com}}\middle|\alpha,\delta,\Sigma\right)=\phi_{2J}\left(y_{i,\text{com}}\middle|X_i\Theta,\Sigma\right)$, where $\Theta=\left(\alpha',\delta'\right)'$, and $\alpha=\left(\alpha_1',\cdots,\alpha_J'\right)'$ and $\delta=\left(\delta_1',\cdots,\delta_J'\right)'$ are defined analogously to β and γ. The advantage of using (12) instead of (9) is that Σ is easier to work with in the PX-MC M-step.

Second, instead of computing $Q\left(\alpha,\delta,\Sigma\middle|\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)=\mathbb{E}\left[\log\left(f\left(y_{\text{mis}},y_{\text{obs}}\middle|\alpha,\delta,\Sigma\right)\right)\right]$ analytically, it is approximated as (11) with Monte Carlo methods and Gibbs sampling. To draw from $\pi\left(y_{\text{mis}}\middle|y_{\text{obs}},\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)$, simply draw $y_{i,\text{mis}}$ and s_i^* from the conditional distribution $\pi\left(y_{i,\text{mis}},s_i^*\middle|y_{i,\text{obs}},s_i,\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)$ for $i=1,\cdots,N$. From (9), we have that

$$\pi\left(y_{i,\text{mis}},s_i^*\middle|y_{i,\text{obs}},s_i,\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\right)\propto\phi_{2J}\left(y_{i,\text{com}}\middle|X_i\theta^{(t)},\Omega^{(t)}\right)p\left(s_i\middle|y_{i,\text{com}}\right), \tag{13}$$

where $\theta^{(t)}=\left(\beta^{(t)\prime},\gamma^{(t)\prime}\right)'$. For the missing outcomes, it is easy to see from (13) that

$$y_{i,j}^*\middle|y_{i,\text{mis}(-j)},s_i^*,y_{i,\text{obs}},s_i,\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\sim\mathcal{N}_1\left(\mu_{i,j|(-j)},\sigma_{i,j|(-j)}^2\right) \tag{14}$$

for $j=1,\cdots,m_i$, where $y_{i,\text{mis}(-j)}$ is equivalent to $y_{i,\text{mis}}$ with $y_{i,j}^*$ removed, and $\mu_{i,j|(-j)}$ and $\sigma_{i,j|(-j)}^2$ are respectively the conditional mean and variance of $y_{i,j}^*$ given all other elements in $y_{i,\text{com}}$ from

$$\phi_{2J}\left(y_{i,\text{com}}\middle|X_i\theta^{(t)},\Omega^{(t)}\right).$$

Similarly, for the latent variables,

$$s_{i,j}^{*}\left|s_{i,(-j)}^{*},y_{i,\text{mis}},y_{i,\text{obs}},s_{i},\beta^{(t)},\gamma^{(t)},\Omega^{(t)}\sim\mathcal{TN}_{B_{i,j}}\left(\lambda_{i,j|(-j)},\omega_{i,j|(-j)}^{2}\right)\right. \tag{15}$$

for $j=1,\cdots,J$, where $\mathcal{TN}_{A}\left(a,b^{2}\right)$ denotes a univariate normal distribution with mean a and variance b^{2} truncated to the region A. In (15), $s_{i,(-j)}^{*}$ is s_{i}^{*} with $s_{i,j}^{*}$ removed, $B_{i,j}$ is the interval $\left(-\infty,0\right]$ if $s_{i,j}=0$ and $\left(0,+\infty\right)$ otherwise, and $\lambda_{i,j|(-j)}$ and $\omega_{i,j|(-j)}^{2}$ are respectively the conditional mean and variance of $s_{i,j}^{*}$ given all other elements of $y_{i,\text{com}}$ from $\phi_{2J}\left(y_{i,\text{com}}\left|X_{i}\theta^{(t)},\Omega^{(t)}\right.\right)$.

The Gibbs sampler recursively samples from the full conditional distributions in (14) and (15) in the usual way. After a sufficient burn-in period, the last G draws are used in (11).

3.2.2. PX-MC M-Step and Reduction Step

By recognizing that (11) is proportional to the log-likelihood function of a seemingly unrelated regression model with NG observations and $2J$ equations, the maximization can be performed with IGLS. IGLS utilizes the quantities

$$\tilde{\Theta}=\left(\sum_{g=1}^{G}\sum_{i=1}^{N}X_{i}'\tilde{\Sigma}^{-1}X_{i}\right)^{-1}\left(\sum_{g=1}^{G}\sum_{i=1}^{N}X_{i}'\tilde{\Sigma}^{-1}y_{i,\text{com}}^{(g)}\right) \tag{16}$$

and

$$\tilde{\Sigma}=\frac{1}{NG}\sum_{g=1}^{G}\sum_{i=1}^{N}\left(y_{i,\text{com}}^{(g)}-X_{i}\tilde{\Theta}\right)\left(y_{i,\text{com}}^{(g)}-X_{i}\tilde{\Theta}\right)', \tag{17}$$

where $y_{i,\text{com}}^{(g)}$ is equivalent to $y_{i,\text{com}}$ with $y_{i,\text{mis}}^{(g)}$ and $s_{i}^{*(g)}$. First evaluate (16) with $\tilde{\Sigma}^{-1}$ removed, which amounts to estimating Θ equation by equation, and then evaluate (17) based on $\tilde{\Theta}$. Proceed by iterating (16) and (17) recursively until convergence. Denote the converged values as $\alpha^{(t+1)}$, $\delta^{(t+1)}$, and $\Sigma^{(t+1)}$.

In the reduction step, set $\beta^{(t+1)}=\alpha^{(t+1)}$, $\gamma_{j}^{(t+1)}=\delta_{j}^{(t+1)}\big/d_{J+j}$ $\left(1\le j\le J\right)$, and $\Omega^{(t+1)}=D^{-1}\Sigma^{(t+1)}D^{-1}$, where $D=\text{diag}\left(1,\cdots,1,d_{J+1},\cdots,d_{2J}\right)$ is a $2J\times 2J$ diagonal matrix with the first J diagonals equal to 1 and the remaining J diagonals equal to the square root of the last J diagonals of $\Sigma^{(t+1)}$. The previous transformations are referred to as the reduction functions, and they are needed because (12) is used instead of (9) in the algorithm [13].

3.3. Standard Errors

The observed information matrix is

$$-\mathbb{E}\left\{\frac{\partial^{2}\log\left(f\left(y_{\text{mis}},y_{\text{obs}}\left|\beta,\gamma,\Omega\right.\right)\right)}{\partial\Psi\partial\Psi'}\right\}-\mathbb{V}\left\{\frac{\partial\log\left(f\left(y_{\text{mis}},y_{\text{obs}}\left|\beta,\gamma,\Omega\right.\right)\right)}{\partial\Psi}\right\}, \tag{18}$$

where $\Psi=\left(\beta',\gamma',\Xi'\right)'$, and Ξ is a column vector denoting the unique elements in Ω. Evaluate (18) at the ML estimates, and take the expectation and variance with respect to $\pi\left(y_{\text{mis}}\left|y_{\text{obs}},\hat{\beta},\hat{\gamma},\hat{\Omega}\right.\right)$. These moments are estimated by taking additional draws from the Gibbs sampler and constructing their Monte Carlo analogs. The standard errors are the square roots of the diagonals of the inverse estimated quantity in (18).

4. Concluding Remarks

A new and simple ML estimation algorithm is developed for multivariate sample selection models. Roughly speaking, the implementation of this algorithm only involves iteratively drawing sets of missing data from well-known distributions and using IGLS on the complete data, both of which are inexpensive to perform. By using parameter expansion and Monte Carlo methods, the algorithm only depends on quantities with closed form

expressions, even when estimating the covariance matrix parameter with correlation restrictions. This algorithm is readily extendable to other types of selection models, including extensions to various types of outcome and selection variables with an underlying normal structure, and modifications to time-series or panel data.

Acknowledgements

I would like to thank the referee, Alicia Lloro, Andrew Chang, Jonathan Cook, and Sibel Sirakaya for their helpful comments.

References

[1] Heckman, J. (1974) Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, **42**, 679-694. http://dx.doi.org/10.2307/1913937

[2] Heckman, J. (1976) The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, **5**, 475-492.

[3] Heckman, J. (1979) Sample Selection Bias as a Specification Error. *Econometrica*, **47**, 153-161. http://dx.doi.org/10.2307/1912352

[4] Su, S.J. and Yen, S.T. (2000) A Censored System of Cigarette and Alcohol Consumption. *Applied Economics*, **32**, 729-737. http://dx.doi.org/10.1080/000368400322354

[5] Yen, S.T., Kan, K. and Su, S.J. (2002) Household Demand for Fats and Oils: Two-Step Estimation of a Censored Demand System. *Applied Economics*, **34**, 1799-1806. http://dx.doi.org/10.1080/00036840210125008

[6] Hao, A.F. (2008) A Discrete-Continuous Model of Households' Vehicle Choice and Usage, with an Application to the Effects of Residential Density. *Transportation Research Part B*: *Methodological*, **42**, 736-758. http://dx.doi.org/10.1016/j.trb.2008.01.004

[7] Li, P. (2011) Estimation of Sample Selection Models with Two Selection Mechanisms. *Computational Statistics & Data Analysis*, **55**, 1099-1108. http://dx.doi.org/10.1016/j.csda.2010.09.006

[8] Li, P. and Rahman, M.A. (2011) Bayesian Analysis of Multivariate Sample Selection Models Using Gaussian Copulas. *Advances in Econometrics*, **27**, 269-288. http://dx.doi.org/10.1108/S0731-9053(2011)000027A013

[9] Yen, S.T. (2005) A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations. *American Journal of Agricultural Economics*, **87**, 453-466. http://dx.doi.org/10.1111/j.1467-8276.2005.00734.x

[10] Tauchmann, H. (2010) Consistency of Heckman-Type Two-Step Estimators for the Multivariate Sample-Selection Model. *Applied Economics*, **42**, 3895-3902. http://dx.doi.org/10.1080/00036840802360179

[11] Puhani, P.A. (2000) The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys*, **14**, 53-68. http://dx.doi.org/10.1111/1467-6419.00104

[12] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, **39**, 1-38. http://dx.doi.org/10.2307/2984875

[13] Liu, C., Rubin, D.B. and Wu, Y.N. (1998) Parameter Expansion to Accelerate EM: The PX-EM Algorithm. *Biometrika*, **85**, 755-770. http://dx.doi.org/10.1093/biomet/85.4.755