Sample selection in random effects models

Leonardo Grilli and **Carla Rampichini**

grilli@ds.unifi.it - carla@ds.unifi.it

Dipartimento di Statistica "Giuseppe Parenti" - Firenze - Italia

Aims of the paper

- study the consequences of sample selection in random effects models
- assess the performances of ML estimators in correcting for selection bias in binary response random effects models

- **Outline of the paper**
  - the linear bivariate random effects (multilevel) model
  - the selection mechanism in random effects models
  - consequences of selection in the linear random effects model
  - the binary bivariate multilevel model with selection
  - simulation study: selection bias in the binary model and ML estimators
  - simulation study: misspecification of errors distribution

Random effects (multilevel) models

In many settings the observations are nested in hierarchical structures. For example:

- workers in firms
- repeated measures of the occupational status of a set of individuals

elementary units (workers, repeated measurements) are embedded in *clusters* (firms, individuals).

Such hierarchical structure is neither accidental nor ignorable.

This kind of structure often implies correlated responses at the elementary level, which can be taken into account by means of *random effects* models, also known as *multilevel* models (Goldstein, 2003).

Selection bias in random effects models

The phenomenon of selection in a multilevel model is much more complex than in a single-level model (without random effects):

- the selection process can act at different levels, giving rise to a wide variety of patterns;
- the model of interest is quite complex, as it is characterized not only by the regression coefficients, but also by the variance-covariance structure which is often of primary interest, so the effect of selection on the variance-covariance structure must be carefully assessed;
- the selection process modifies the hierarchical structure of the data (number of clusters and cluster sizes), a feature that is relevant in the estimation phase, as it influences the behavior of the estimation algorithms, the accuracy of the asymptotic approximations and the power of the estimators.

The bivariate linear random effects model

- ✓ Response variables: $\widetilde{Y}^S$ and $\widetilde{Y}^P$
- ✓ S= $Selection$; P = $Principal$ (variable of main interest)

$$\boxed{\begin{array}{rcl} \widetilde{Y}_{ij}^S & = & \mathbf{z}_{ij}^S \boldsymbol{\theta}^S + u_j^S + e_{ij}^S \\ \widetilde{Y}_{ij}^P & = & \mathbf{z}_{ij}^P \boldsymbol{\theta}^P + u_j^P + e_{ij}^P \end{array}}$$

$i = 1, 2, \ldots, n_j$ elementary index, $j = 1, 2, \ldots, J$ cluster index

- $\mathbf{z}_{ij}$ covariates at elementary or cluster level
- $\boldsymbol{\theta}$ regression coefficients
- each covariate may enter both equations, or may be equation-specific

The u_j's are cluster-level errors (random effects), e_{ij}'s elementary-level errors.

$$\begin{bmatrix} e_{ij}^S \\ e_{ij}^P \end{bmatrix} \overset{iid}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_S^2 & \\ \sigma_{SP} & \sigma_P^2 \end{bmatrix}$$

$$\begin{bmatrix} u_j^S \\ u_j^P \end{bmatrix} \overset{iid}{\sim} N\left(\mathbf{0}, \mathbf{T}\right), \qquad \mathbf{T} = \begin{bmatrix} \tau_S^2 & \\ \tau_{SP} & \tau_P^2 \end{bmatrix}.$$

Variance decomposition in the random effects model

- Given the hierarchical structure of the model, the marginal variances and covariance are the sum of the cluster-level and elementary-level variances and covariances, respectively:

$$\boxed{\begin{aligned}
\text{var}(\tilde{Y}_{ij}^S) &= \text{var}(u_j^S) + \text{var}(e_{ij}^S) = \tau_S^2 + \sigma_S^2 \\
\text{var}(\tilde{Y}_{ij}^P) &= \text{var}(u_j^P) + \text{var}(e_{ij}^P) = \tau_P^2 + \sigma_P^2 \\
\text{cov}(\tilde{Y}_{ij}^S, \tilde{Y}_{ij}^P) &= \text{cov}(u_j^S, u_j^P) + \text{cov}(e_{ij}^S, e_{ij}^P) = \tau_{SP} + \sigma_{SP} \quad .
\end{aligned}}$$

- the Intraclass Correlation Coefficient (ICC) is the proportion of variance due to clustering:

$$\boxed{\begin{array}{ll}
\tau_S^2/(\tau_S^2 + \sigma_S^2) & \text{for the } \textit{Selection} \text{ equation} \\
\tau_P^2/(\tau_P^2 + \sigma_P^2) & \text{for the } \textit{Principal} \text{ equation}
\end{array}}$$

- The marginal correlation among the responses is:

$$\boxed{\rho_{tot} = \text{corr}(\tilde{Y}_{ij}^S, \tilde{Y}_{ij}^P) = \frac{\tau_{SP} + \sigma_{SP}}{\sqrt{(\tau_S^2 + \sigma_S^2)(\tau_P^2 + \sigma_P^2)}}}$$

Selection mechanism

$\widetilde{Y}^P$ is observed depending on the value of $\widetilde{Y}^S$

$$\widetilde{Y}^P_{ij} = \begin{cases} \widetilde{Y}^P_{ij} & \text{if } \widetilde{Y}^S_{ij} > 0 \\ \textit{not observed} & \text{otherwise} \end{cases}$$

- The selection mechanism operates at elementary level, as it causes the missingness of single observations (even when σ_{SP} is null, as in many models for panel or longitudinal data)
- within a given cluster the pattern of missingness can be of any kind ("non-monotone missingness") while in many studies attention is restricted to the special case of drop-out or attrition, where missingness at a given time point implies missingness at all subsequent time points
- A selection mechanism that operates at elementary level modifies the hierarchical structure of the data in terms of the cluster sizes and possibly also in terms of the number of clusters

Some definitions: analysis in the linear case

$\rightsquigarrow$ The selection mechanism is **ignorable** when *both* the couples of error terms at cluster (u_j^S, u_j^P) and elementary level (e_{ij}^S, e_{ij}^P) are uncorrelated, i.e. *both* covariance parameters σ_{SP} and τ_{SP} are null: in this case the models for the *Selection* and *Principal* equations can be **fitted separately**, without any bias or loss of efficiency.

$\rightsquigarrow$ When the selection mechanism is not ignorable it is of interest to determine the bias which arises when fitting the *Principal* equation alone.

- ✓ w_{ij}^S is the composite error of the *Selection* equation: $w_{ij}^S = u_j^S + e_{ij}^S$
- ✓ $\widetilde{Y}_{ij}^P$ is observed if and only if $w_{ij}^S > -\mathbf{z}_{ij}^S \boldsymbol{\theta}^S$
- ✓ after selection $\Leftrightarrow$ conditional on truncation on w_{ij}^S

Some definitions cont'd

⇝ Set of truncation events of the whole cluster:

$$\begin{aligned} A_{.j} &= \left\{\bigcap_{i:\widetilde{Y}^S_{ij}>0}\left\{\widetilde{Y}^S_{ij}>0\right\}\right\}\bigcap\left\{\bigcap_{i:\widetilde{Y}^S_{ij}\le 0}\left\{\widetilde{Y}^S_{ij}\le 0\right\}\right\} \\ &= \left\{\bigcap_{i:\widetilde{Y}^S_{ij}>0}\left\{w^S_{ij}>-\mathbf{z}^S_{ij}\boldsymbol{\theta}^S\right\}\right\}\bigcap\left\{\bigcap_{i:\widetilde{Y}^S_{ij}\le 0}\left\{w^S_{ij}\le -\mathbf{z}^S_{ij}\boldsymbol{\theta}^S\right\}\right\}. \end{aligned}$$

⇝ Truncation event for the first elementary unit of the j-th cluster (observed): $A_{1j}=\{w^S_{1j}>-\mathbf{z}^S_{1j}\boldsymbol{\theta}^S\}$

Consider first elementary unit ($i=1$) of cluster j (observed). To derive the properties of the model $\widetilde{Y}^P_{1j}=\mathbf{z}^P_{1j}\boldsymbol{\theta}^P+u^P_j+e^P_{1j}$ after selection:

- the observations pertaining to other clusters are irrelevant, as it was assumed independence among clusters
- The relevant variables are the two errors in $\widetilde{Y}^P_{1j}$, namely u^P_j and e^P_{1j}, plus all the composite errors determining selection in the cluster under consideration, namely $(w^S_{1j},w^S_{2j},\ldots,w^S_{n_j j})$
- Truncation is below for the elementary units which are observed and above for the others.

Moments for the linear random effects model

To evaluate the consequences of selection on the model $\widetilde{Y}_{1j}^P = \mathbf{z}_{1j}^P \boldsymbol{\theta}^P + u_j^P + e_{1j}^P$, the key quantities are:

$$
\boxed{
\begin{aligned}
E\left(\widetilde{Y}_{1j}^P \mid u_j^P, A_{.j}\right) &= \mathbf{z}_{1j}^P \boldsymbol{\theta}^P + u_j^P + E\left(e_{1j}^P \mid u_j^P, A_{.j}\right) \\
E\left(\widetilde{Y}_{1j}^P \mid A_{.j}\right) &= \mathbf{z}_{1j}^P \boldsymbol{\theta}^P + E\left(u_j^P \mid A_{.j}\right) + E\left(e_{1j}^P \mid A_{.j}\right) \\
Var\left(\widetilde{Y}_{1j}^P \mid A_{.j}\right) &= Var\left(u_j^P \mid A_{.j}\right) + Var\left(e_{1j}^P \mid A_{.j}\right) + 2Cov\left(u_j^P, e_{1j}^P \mid A_{.j}\right)
\end{aligned}
}
$$

Selection modifies the relationships among the involved errors, leading to a complex configuration where the basic model assumptions may break down.
If the selection is not ignorable, fitting the *Principal* equation on available data involves the following potential problems:

- the regression coefficients are biased for the covariates that enter both equations
- the well-known equivalence between conditional and marginal regression coefficients in linear random effects models is corrupted
- the errors are no more homoscedastic, nor independent
- the ICC is biased

Means and variances of the errors after selection

For certain configurations of the model parameters not all the **potential problems** are in effect:

(a) when $\sigma_S^2 > 0$, $\sigma_P^2 > 0$, $\tau_S^2 > 0$ and $\tau_P^2 > 0$

	$0<\rho_\tau^2<1$	$\rho_\tau^2=0$	$\rho_\tau^2=0$
$0<\rho_\sigma^2<1$	$E(e_{1j}^P \mid u_j^P, A_{.j})$ $E(u_j^P \mid A_{.j}) + E(e_{1j}^P \mid A_{.j})$ $Var(u_j^P + e_{1j}^P \mid A_{.j})$	$E(e_{1j}^P \mid u_j^P, A_{.j}) = E(e_{1j}^P \mid A_{.j})$ $E(u_j^P) + E(e_{1j}^P \mid A_{.j})$ $Var(u_j^P) + Var(e_{1j}^P \mid A_{.j})$	$E(e_{1j}^P \mid u_j^P, A_{1j})$ $E(u_j^P \mid A_{.j}) + E(e_{1j}^P \mid A_{.j})$ $Var(u_j^P + e_{1j}^P \mid A_{.j})$
$\rho_\sigma^2=0$	0 $E(u_j^P \mid A_{.j})$ $Var(u_j^P \mid A_{.j}) + Var(e_{1j}^P)$	0 $E(u_j^P \mid A_{.j})$ $Var(u_j^P) + Var(e_{1j}^P)$	0 $E(u_j^P \mid A_{.j})$ $Var(u_j^P \mid A_{.j}) + Var(e_{1j}^P)$
$\rho_\sigma^2=1$	$E(e_{1j}^P \mid u_j^P, A_{.j})$ $E(u_j^P \mid A_{.j}) + E(e_{1j}^P \mid A_{.j})$ $Var(u_j^P + e_{1j}^P \mid A_{.j})$	$E(e_{1j}^P \mid A_{.j})$ $E(e_{1j}^P \mid A_{.j})$ $Var(u_j^P) + Var(e_{1j}^P \mid A_{.j})$	//

(b) when $\sigma_S^2 > 0$, $\sigma_P^2 > 0$ and $\rho_\sigma^2 > 0$

$\tau_S^2>0,\ \tau_P^2>0,\ \rho_\tau^2=0$	$\tau_S^2=0,\ \tau_P^2>0,\ \rho_\tau^2=0$	$\tau_S^2>0,\ \tau_P^2=0,\ \rho_\tau^2=0$
$E(e_{1j}^P \mid u_j^P, A_{.j}) = E(e_{1j}^P \mid A_{.j})$ $E(e_{1j}^P \mid A_{.j})$ $Var(u_j^P) + Var(e_{1j}^P \mid A_{.j})$	$E(e_{1j}^P \mid u_j^P, A_{1j}) = E(e_{1j}^P \mid A_{1j})$ $E(e_{1j}^P \mid A_{1j})$ $Var(u_j^P) + Var(e_{1j}^P \mid A_{1j})$	$E(e_{1j}^P \mid u_j^P, A_{.j}) = E(e_{1j}^P \mid A_{.j})$ $E(e_{1j}^P \mid A_{.j})$ $Var(e_{1j}^P \mid A_{.j})$

For each cell:

1st row $E(e_{1j}^P \mid u_j^P, A_{.j})$

2nd row $E(u_j^P \mid A_{.j}) + E(e_{1j}^P \mid A_{.j})$

3rd row $Var(u_j^P + e_{1j}^P \mid A_{.j})$

The bivariate model with selection

Two binary variables are generated by two corresponding latent continuous responses $\tilde{Y}^P$ and $\tilde{Y}^S$ in the following way:

$$\boxed{\{Y^S = 1\} \Leftrightarrow \{\tilde{Y}^S > 0\} \quad \text{and} \quad \{Y^P = 1\} \Leftrightarrow \{\tilde{Y}^P > 0\}}$$

The conditional likelihood function consists of two parts:

$$\boxed{L_{ij}(\boldsymbol{\psi}_{(-\mathbf{T})} \mid u_j^S, u_j^P) = \begin{cases} P(Y_{ij}^S = 0 \mid u_j^S) & \text{if} \quad Y_{ij}^S = 0 \\ \\ P(Y_{ij}^S = 1, Y_{ij}^P = y_{ij}^P \mid u_j^S, u_j^P) & \text{if} \quad Y_{ij}^S = 1 \end{cases}}$$

Marginal likelihood function:

$$\boxed{L_j(\boldsymbol{\psi}) = \int\int \prod_{i=1}^{n_j} L_{ij}(\boldsymbol{\psi}_{(-\mathbf{T})} \mid u_j^S, u_j^P) g(u_j^S, u_j^P; \mathbf{0}, \mathbf{T}) du_j^S du_j^P}$$

where $\boldsymbol{\psi}$ is the vector of all parameters and $g(., .; \mathbf{0}, \mathbf{T})$ is the bivariate normal density with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{T}$.

$\rightsquigarrow$ $\sigma_{SP} = \tau_{SP} = 0$ (i.e. ignorable selection) implies the factorization of the marginal likelihood: in this case the models for the *Selection* and *Principal* equations can be fitted separately, without any bias or loss of efficiency.

Simulation design: the model

A bivariate random effects linear model is assumed for the latent responses:

$$\boxed{\begin{aligned}\tilde{Y}^S_{ij} &= \alpha^S + \beta^S_1 x_{1ij} + \beta^S_2 x_{2ij} + \gamma^S v_j + e^S_{ij} + u^S_j \\ \tilde{Y}^P_{ij} &= \alpha^P + \beta^P_1 x_{1ij} + \beta^P_3 x_{3ij} + \gamma^P v_j + e^P_{ij} + u^P_j .\end{aligned}}$$

✓ Some of the covariates enter both equations, while other covariates are equation-specific to avoid identification problems.

✓ x's elementary-level covariates, v a cluster-level covariate

✓ distribution of the errors as before

Regression model for the joint distribution of two observed binary variables Y^P and Y^S:

$$\boxed{\{Y^S = 1\} \Leftrightarrow \{\tilde{Y}^S > 0\} \quad \text{and} \quad \{Y^P = 1\} \Leftrightarrow \{\tilde{Y}^P > 0\}}$$

♢ Variances of the elementary-level errors fixed to 1 for identification (one could be left free when model equations are jointly estimated).

Simulation design: covariates and parameters

✓ The three elementary-level covariates and the cluster-level covariate are generated independently from standard normal distributions.

✓ The values of *true* parameters used in the experiments are:

- *regression parameters*: $\alpha^S = 1$, $\beta_1^S = 0.5$, $\beta_2^S = 0.3$, $\gamma^S = 0.5$, $\alpha^P = 0$, $\beta_1^P = 0.5$, $\beta_3^P = 0.3$, $\gamma^P = 0.5$;
- *variance parameters*: $\tau_S^2 = 1$, $\tau_P^2 = 1$.

✓ Six different experiments are performed, varying the values of the errors' *covariances* σ_{SP} and τ_{SP}: the covariances are chosen to obtain three distinct values of the residual marginal correlation between the two considered latent responses, ρ_{tot}: 0, 0.5 and 0.9. All the other parameters used in the data generation process are held constant among the experiments.

✓ Since the cluster-level variances are both assumed to be unity, $\tau_S^2 = \tau_P^2 = 1$, the ICC is 0.5 for both equations. This value of the ICC means that the clustering of the units is quite relevant, though in a panel setting should be considered as moderate.

✓ The value of α^S is crucial in determining the strength of the selection mechanism. Fixing α^S to one leads to a selection that excludes about 27% of the data on the *Principal* equation, varying from 20% to 39% on the performed replications.

Simulation design: hierarchical structure of the data

Balanced design with a total of 500 observations, arranged in three different structures:

- a structure with many clusters (100) and few observations per cluster (5), like in a longitudinal or panel study;
- a structure with few clusters (25) and many observations per cluster (20), like in cross-sectional studies, e.g. in the educational setting;
- an intermediate structure with 50 clusters and 10 observations per cluster.

We compare these three different structures using a single set of parameters, with $\sigma_{SP} = 0.5$ and $\tau_{SP} = 0.5$.

$\diamondsuit$ Estimation: NLMIXED of SAS, Quasi-Newton with non-adaptive Gaussian quadrature, 8 pt of quadrature

Results: role of the hierarchical structure of the data

Estimated variance-covariance parameters (s.e. in parenthesis) from joint estimation of the model equations: data without and with selection for three different hierarchical structures (J=No. of clusters, $n = n_j$= No. of elementary units per cluster). Mean on 100 replications.

Parameter	J=100, n=5		J=50, n=10		J=25, n=20	
	Selection		Selection		Selection	
	no	yes	no	yes	no	yes
σ_{SP}	0.498	0.354	0.510	0.289	0.513	0.350
	(0.112)	(0.574)	(0.106)	(0.535)	(0.107)	(0.508)
$\tau_S^2 = 1$	1.018	1.036	1.110	1.118	1.160	1.164
	(0.314)	(0.314)	(0.458)	(0.486)	(0.537)	(0.534)
$\tau_P^2 = 1$	0.998	0.927	1.077	1.040	1.140	1.133
	(0.292)	(0.292)	(0.364)	(0.419)	(0.667)	(0.605)
τ_{SP}	0.501	0.456	0.552	0.450	0.483	0.457
	(0.180)	(0.261)	(0.270)	(0.325)	(0.360)	(0.392)

selection

- no = complete data
- yes = data without missing on $\widetilde{Y}^P$

✓ The variance-covariance parameters at cluster level are estimated with low bias regardless of selection, except for a small underestimation of the covariance under selection. On the other hand, the covariance at elementary level is well estimated in the case of no selection, but it is largely underestimated in the case of selection.

✓ Due to the reduction of the sample size for the *Principal* equation, the selection induces an increase in the standard errors of the estimators, especially for the covariance parameters. Obviously this is not the case for τ_S. Since the selection mechanism here simulated has a low probability to eliminate a whole cluster, the standard error of the estimator of τ_{SP} is less affected than that of σ_{SP}.

✓ The hierarchical structure of the data has an effect mainly on the standard errors of the estimators, which has the same direction regardless of selection: when increasing the size of clusters and decreasing the number of clusters, i.e. reading the table from left to right, the standard errors of the cluster-level variance-covariance estimators substantially increase, while the standard error of the elementary-level covariance estimator shows a modest reduction.

Results: power of the LRT, normal errors

LRT test of size 5% for the hypothesis of no sample selection (joint estimation of model equations vs separate estimation): percentage of rejection on 100 replications. Design with 50 clusters and 10 observations per cluster.

Parameters			data without selection		data with selection	
σ_{SP}	τ_{SP}	ρ_{tot}	one-level model	two-level model	one-level model	two-level model
0.5	0.5	0.5	100.0	100.0	14.2	34.8
0.1	0.9	0.5	100.0	100.0	12.3	100.0
0.9	0.1	0.5	100.0	100.0	18.6	44.9
0.9	0.9	0.9	100.0	100.0	44.3	95.1
0.5	-0.5	0.0	22.4	99.0	6.4	78.4
0.9	-0.9	0.0	36.0	100.0	12.1	98.8

✓ When the model is correctly specified as a two-level model, i.e. with the random effects, the null hypothesis is $H_0 : \sigma_{SP} = \tau_{SP} = 0$. In contrast, when the model is incorrectly specified as a one-level model, i.e. without random effects, the null hypothesis is $H_0 : \rho_{tot} = 0$.

✓ Considering the data set without selection, the multilevel model always rejects the null hypothesis in all the 6 considered configurations, while the one-level model often fails when the two correlations balance each other to give a null total correlation.

✓ When the data are affected by selection, the ability to detect the selection using the multilevel model depends crucially on the value of τ_{SP}, that is the covariance at cluster level: the test performance tends to be better when τ_{SP} is high in absolute value and quite different from σ_{SP}. If the one-level model is incorrectly used, the power of the LRT for $H_0 : \rho_{tot} = 0$ is very low, except when ρ_{tot} is 0.9.

Results: parameter estimates, normal errors

Table 3: Estimated parameters (standard errors in parenthesis) from joint and separate estimation of the model equations. Data with selection. Design with 50 clusters and 10 observations per cluster. Means on 100 replications.

Model	$\rho_{tot} = 0.5$					
	$\sigma_{SP} = 0.5, \tau_{SP} = 0.5$		$\sigma_{SP} = 0.1, \tau_{SP} = 0.9$		$\sigma_{SP} = 0.9, \tau_{SP} = 0.1$	
Parameter	joint	sep.	joint	sep.	joint	sep.
Regression						
$\alpha^S = 1$	1.027	1.034	0.966	1.015	1.015	1.015
	(0.202)	(0.213)	(0.250)	(0.232)	(0.206)	(0.197)
$\beta_1^S = 0.5$	0.512	0.512	0.496	0.500	0.516	0.514
	(0.071)	(0.072)	(0.085)	(0.086)	(0.092)	(0.091)
$\beta_2^S = 0.3$	0.322	0.323	0.287	0.286	0.313	0.304
	(0.091)	(0.088)	(0.079)	(0.079)	(0.092)	(0.090)
$\gamma^S = 0.5$	0.513	0.516	0.518	0.553	0.504	0.500
	(0.233)	(0.227)	(0.291)	(0.260)	(0.246)	(0.241)
$\alpha^P = 0$	0.089	0.249	0.022	0.142	0.128	0.394
	(0.292)	(0.218)	(0.275)	(0.175)	(0.276)	(0.238)
$\beta_1^P = 0.5$	0.476	0.443	0.467	0.451	0.508	0.448
	(0.114)	(0.102)	(0.116)	(0.100)	(0.089)	(0.097)
$\beta_3^P = 0.3$	0.288	0.308	0.295	0.305	0.321	0.359
	(0.082)	(0.087)	(0.083)	(0.084)	(0.091)	(0.098)
$\gamma^P = 0.5$	0.443	0.380	0.481	0.448	0.466	0.405
	(0.234)	(0.229)	(0.270)	(0.229)	(0.229)	(0.250)
Var-cov						
σ_{SP}	0.289	.	0.031	.	0.744	.
	(0.535)	(.)	(0.513)	(.)	(0.376)	(.)
$\tau_S^2 = 1$	1.118	1.077	0.978	1.027	1.037	1.060
	(0.486)	(0.432)	(0.324)	(0.351)	(0.377)	(0.456)
$\tau_P^2 = 1$	1.040	1.024	0.920	0.824	1.320	1.558
	(0.419)	(0.409)	(0.422)	(0.295)	(0.620)	(0.655)
τ_{SP}	0.450	.	0.812	.	-0.017	.
	(0.325)	(.)	(0.288)	(.)	(0.335)	(.)

- The bias caused by selection can be seen by comparing the estimates from the columns labelled *joint* and *sep*.
- In the probit model all the estimable parameters are scaled by the elementary-level standard deviation σ_P, whose underestimation depends on the value of σ_{SP}.
- Therefore even the regression coefficient β_3^P, which appear only in the *Principal* equation, may suffer from selection bias: the higher σ_{SP} the higher the bias on β_3^P.

Results: parameter estimates, normal errors cont'd

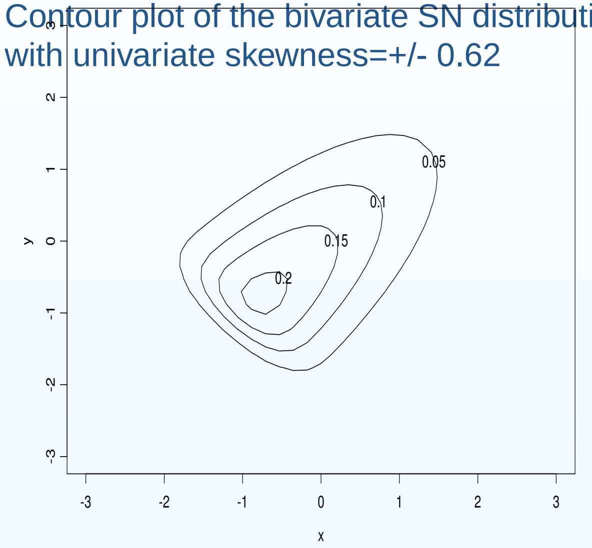
Table 4: (continued) Estimated parameters (standard errors in parenthesis) from joint and separate estimation of the model equations. Data with selection. Design with 50 clusters and 10 observations per cluster. Means on 100 replications.

Model	$\rho_{tot} = 0.9$		$\rho_{tot} = 0.0$			
	$\sigma_{SP} = 0.9, \tau_{SP} = 0.9$		$\sigma_{SP} = 0.5, \tau_{SP} = -0.5$		$\sigma_{SP} = 0.9, \tau_{SP} = -0.9$	
Parameter	joint	sep.	joint	sep.	joint	sep.
Regression						
$\alpha^S = 1$	0.979	1.003	1.018	1.022	1.070	1.053
	(0.217)	(0.187)	(0.220)	(0.224)	(0.282)	(0.235)
$\beta_1^S = 0.5$	0.510	0.508	0.515	0.512	0.508	0.507
	(0.096)	(0.086)	(0.083)	(0.085)	(0.092)	(0.088)
$\beta_2^S = 0.3$	0.309	0.300	0.302	0.301	0.309	0.302
	(0.092)	(0.086)	(0.083)	(0.079)	(0.087)	(0.086)
$\gamma^S = 0.5$	0.504	0.525	0.523	0.513	0.497	0.477
	(0.279)	(0.250)	(0.216)	(0.242)	(0.260)	(0.246)
$\alpha^P = 0$	0.106	0.436	0.077	0.139	0.081	0.282
	(0.212)	(0.155)	(0.335)	(0.220)	(0.294)	(0.259)
$\beta_1^P = 0.5$	0.464	0.381	0.471	0.470	0.485	0.458
	(0.096)	(0.087)	(0.119)	(0.105)	(0.100)	(0.107)
$\beta_3^P = 0.3$	0.319	0.362	0.292	0.307	0.321	0.355
	(0.085)	(0.083)	(0.091)	(0.094)	(0.082)	(0.100)
$\gamma^P = 0.5$	0.446	0.315	0.480	0.481	0.492	0.478
	(0.225)	(0.180)	(0.223)	(0.231)	(0.282)	(0.311)
Var-cov						
σ_{SP}	0.605	.	0.324	.	0.745	.
	(0.445)	(.)	(0.548)	(.)	(0.315)	(.)
$\tau_S^2 = 1$	1.006	1.018	0.976	0.967	1.053	1.093
	(0.379)	(0.414)	(0.353)	(0.355)	(0.483)	(0.504)
$\tau_P^2 = 1$	0.930	0.681	1.120	1.220	1.188	1.754
	(0.384)	(0.303)	(0.644)	(0.490)	(0.562)	(0.799)
τ_{SP}	0.815	.	-0.525	.	-0.972	.
	(0.313)	(.)	(0.337)	(.)	(0.308)	(.)

- The regression coefficients present in both equations, β_1^P and γ^P, and the cluster-level variance τ_P^2 are doubly affected by the selection: as for β_3^P there is the bias coming from the scaling, moreover these parameters are biased also w.r.t. the latent variable model.
- The full information ML estimator corrects quite well the selection bias, even if some bias still remains due to the systematic underestimation of σ_{SP}.

Results: power of the LRT, skew-normal errors

Contour plot of the bivariate SN distribution
with univariate skewness=+/- 0.62



LRT test of size 5% for the hypothesis of no sample selection (joint model vs separate models): percentage of rejection on 100 replications. Design with 50 clusters and 10 observations per cluster. $\sigma_{SP} = \tau_{SP} = 0.5$

Errors distribution		Selection	
Elementary	Cluster	No	Yes
Normal	Normal	100.0	34.8
SN +	Normal	99.0	38.3
Normal	SN +	100.0	28.6
SN -	Normal	100.0	15.0
Normal	SN -	100.0	30.1

✓The estimation under the assumption of normal distributed errors when the distribution of the errors is indeed skewed corresponds to a link misspecification

✓ The percentage of rejection on 100 replications for the LRT test of size 5% for the hypothesis of no sample selection is about 35% if the errors are both normal.

✓ This percentage changes with the errors distribution. Particularly it substantially decreases if the elementary-level errors are negatively skewed.

Results: parameter estimates, skew-normal errors

Table 6: Estimated parameters (standard errors in parenthesis) from joint estimation of the model equations under different distributional assumptions on the error terms. Data with selection. Design with 50 clusters and 10 observations per cluster. Means on 100 replications.

Parameter	Normal errors at both levels	Positive SN errors elementary	Positive SN errors cluster	Negative SN errors elementary	Negative SN errors cluster
Regression					
$\alpha^S = 1$	1.027	1.078	1.025	1.000	1.066
	(0.202)	(0.229)	(0.180)	(0.213)	(0.245)
$\beta_1^S = 0.5$	0.512	0.546	0.503	0.465	0.516
	(0.071)	(0.088)	(0.091)	(0.089)	(0.101)
$\beta_2^S = 0.3$	0.322	0.331	0.312	0.285	0.305
	(0.091)	(0.078)	(0.087)	(0.081)	(0.088)
$\gamma^S = 0.5$	0.513	0.538	0.518	0.458	0.469
	(0.233)	(0.232)	(0.205)	(0.221)	(0.229)
$\alpha^P = 0$	0.089	-0.035	0.043	0.230	0.149
	(0.292)	(0.233)	(0.241)	(0.360)	(0.261)
$\beta_1^P = 0.5$	0.476	0.482	0.489	0.474	0.471
	(0.114)	(0.099)	(0.109)	(0.151)	(0.112)
$\beta_3^P = 0.3$	0.288	0.281	0.305	0.323	0.298
	(0.082)	(0.079)	(0.086)	(0.079)	(0.091)
$\gamma^P = 0.5$	0.443	0.431	0.502	0.460	0.441
	(0.234)	(0.211)	(0.218)	(0.244)	(0.214)
Var-cov					
σ_{SP}	0.289	0.539	0.372	0.016	0.259
	(0.535)	(0.373)	(0.443)	(0.762)	(0.569)
$\tau_S^2 = 1$	1.118	1.298	0.969	0.896	1.185
	(0.486)	(0.450)	(0.405)	(0.380)	(0.432)
$\tau_P^2 = 1$	1.040	0.923	1.051	1.039	0.878
	(0.419)	(0.343)	(0.421)	(0.443)	(0.362)
τ_{SP}	0.450	0.542	0.423	0.367	0.398
	(0.325)	(0.312)	(0.319)	(0.306)	(0.337)

The estimation under the assumption of normal distributed errors when the distribution of the errors is indeed skewed:

- seems to have little effect on the regression coefficients and on the cluster-level variance
- affects the covariance parameters, especially at the elementary level (σ_{SP}).