Tauchmann, Harald

Working Paper

A Note on Consistency of Heckman-type two-step Estimators for the Multivariate Sample-Selection Model

RWI Discussion Papers, No. 40

Provided in Cooperation with:
RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Tauchmann, Harald (2006) : A Note on Consistency of Heckman-type two-step Estimators for the Multivariate Sample-Selection Model, RWI Discussion Papers, No. 40, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Essen

This Version is available at:
https://hdl.handle.net/10419/18591

RWI : Discussion Papers

Harald Tauchmann

A Note on Consistency of Heckman-type two-step Estimators for the Multivariate Sample-Selection Model

No. 40

RWI
ESSEN

Rheinisch-Westfälisches Institut für Wirtschaftsforschung

RWI : Discussion Papers
No. 40

Published by Rheinisch-Westfälisches Institut für Wirtschaftsforschung,
Hohenzollernstrasse 1/3, D-45128 Essen, Phone +49 (0) 201/81 49-0

Editor: Prof. Dr. Christoph M. Schmidt, Ph.D.
ISSN 1612-3565 – ISBN 3-936454-64-7

RWI : Discussion Papers
No. 40

Harald Tauchmann

A Note on Consistency of Heckman-type two-step Estimators for the Multivariate Sample-Selection Model

RWI
ESSEN

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

ISSN 1612-3565
ISBN 3-936454-64-7

Harald Tauchmann*

A Note on Consistency of Heckman-type two-step Estimators for the Multivariate Sample-Selection Model

Abstract
This analysis shows that multivariate generalizations to the classical Heckman (1976 and 1979) two-step estimator that account for cross-equation correlation and use the inverse Mills ratio as a correction-term are consistent only if certain restrictions apply to the true error-covariance structure. We derive an alternative class of generalizations to the classical Heckman two-step approach that conditions on the entire selection pattern rather than the selection of particular equations and, therefore, uses modified correction-terms. This class of estimators is shown to be consistent. In addition, Monte-Carlo results illustrate that these estimators display a smaller mean square prediction error.

JEL Classification: C15, C34, C51

Keywords: Multivariate sample-selection model, censored system of equations, Heckman-correction

April 2006

*All correspondence to Harald Tauchmann, RWI Essen, Hohenzollernstraße 1-3, 45128 Essen, Germany, Fax: +49–201–8149200, Email: harald.tauchmann@rwi-essen.de. – The author is grateful to Ch.M. Schmidt for many helpful comments. Any remaining errors are my own.

1 Introduction

Using non-aggregated micro-data for estimating systems of seemingly unrelated equations – the most prominent among them being demand systems – often encounters the problem of numerous zero-observations in the dependent variables. These cannot be appropriately explained by conventional continuous *SUR*[1] models. Instead, zero-observations may be modelled as determined by an upstream multivariate binary choice problem. Under the assumption of normally distributed errors, the resulting joint model represents a multivariate generalization to the classical univariate sample-selection model, cf. Heckman (1976 and 1979). In the literature, this model is often referred to as a "censored system of equations", yet censoring in the narrow sense just represents a special case of the general model.[2]

The question of how to estimate the parameters of this model is subject to an ongoing debate. Clearly, under parametric distributional assumptions full information maximum likelihood (*FIML*) is the efficient estimation technique. In fact, *FIML* has recently been applied to this problem by Yen (2005). However, the *FIML* estimator is computationally extremely demanding, rendering much simpler two-step approaches worth considering for many applications.

Among two-step estimators the one proposed by Heien & Wessels (1990) has been particularly popular. Besides numerous other authors, it has been applied by Heien & Durham (1991), Gao et al. (1995), and Nayga et al. (1999). However, Shonkwiler & Yen (1999) as well as Vermeulen (2001) show that this estimator lacks a decent basis in statistical theory and cannot be interpreted in terms of conditional means. The Heien & Wessels (1990) estimator, therefore, is inconsistent despite its popularity. Chen & Yen (2005) further investigate the nature of its inconsistency and show that even a modified variant of this estimator fails to correct properly for sample-selection bias. Shonkwiler & Yen (1999) propose an alternative simple two-step estimator that

[1] See Zellner (1963) for the seemingly unrelated regression equations (*SUR*) model.

[2] We stick to the relevant literature und use the term "censored" as a synonym for "not selected".

– in contrast to Heien & Wessels (1990) – is theoretically well founded. This estimator is based on the mean of dependent variables that is unconditional on the outcome of the upstream discrete choice model. Su & Yen (2000), Yen et al. (2002) and Goodwin et al. (2004) may serve as examples for applications of this procedure.

Tauchmann (2005) compares the performance of the Shonkwiler & Yen (1999) estimator and two-step estimators that – analogously to the classical Heckman (1976 and 1979) two-step approach, yet in contrast to Shonkwiler & Yen (1999) – condition on the outcome the upstream discrete choice model. In terms of the mean square prediction error, the unconditional Shonkwiler & Yen (1999) estimator is shown to perform poorly if the conditional mean of the dependent variables is large compared to its conditional variance. Tauchmann (2005), however, exclusively focuses on the mean square error yet does not check for unbiasedness and consistency of the conditional estimators. Though one may argue that it is of no relevance in applied work whether an error originates from an estimator's bias or from its variance, many researches do avoid inconsistent estimators, even if their mean square error is small. For this reason, addressing unbiasedness and consistency of Heckman-type two-step estimators for censored systems of equations is a relevant task.

The analysis presented in this article shows that some of the estimators proposed by Tauchmann (2005) are consistent only for restrictive error-covariance structures. It also shows that a modified two-step Heckman-type estimator is generally consistent and performs well in terms of the mean square prediction error. In order to yield these results, the remainder of this paper is organized as follows: Section 2 introduces the model to be analyzed in more detail and analyzes the properties of straightforward multivariate generalizations to the Heckman (1976 and 1979) two-step estimator. In Section 3 an alternative class of generalized two-step Heckman-type estimators is derived. Section 4 presents results from Monte-Carlo simulations that illustrate the theoretical results and extends the analysis to the estimators' mean square error. Section 5 concludes.

2 An analysis of sample-selection models

2.1 A multivariate sample-selection model

Recall the m-variate sample-selection model, which is analyzed by Heinen & Wessels (1990), Shonkwiler & Yen (1999), Tauchmann (2005), Yen (2005), and Chen & Yen (2005). The equations

$$y_{it}^{*} = x_{it}'\beta_i + \varepsilon_{it} \tag{1}$$

$$d_{it}^{*} = z_{it}'\alpha_i + \upsilon_{it}, \tag{2}$$

characterize the latent model, that is y_{it}^{*} and d_{it}^{*} are unobserved. Their observed counterparts y_{it} and d_{it} are determined by

$$d_{it} = \begin{cases} 1 & \text{if} \quad d_{it}^{*} > 0 \\ 0 & \text{if} \quad d_{it}^{*} \leq 0 \end{cases} \tag{3}$$

$$y_{it} = d_{it} y_{it}^{*}. \tag{4}$$

Here, $i = 1, \ldots, m$ indexes the m equations of the system, and $t = 1, \ldots, T$ indexes the individuals. x_{it} and z_{it} are vectors of observed exogenous variables. The vector $d_t = [d_{1t} \ldots d_{mt}]'$ describes the entire individual selection pattern. Finally, $\varepsilon_t = [\varepsilon_{1t} \ldots \varepsilon_{mt}]'$ and $\upsilon_t = [\upsilon_{1t} \ldots \upsilon_{mt}]'$ are normally distributed, zero-mean error vectors with the covariance matrix

$$\operatorname{Var}(\varepsilon_t, \upsilon_t) = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\upsilon}' \\ \Sigma_{\varepsilon\upsilon} & \Sigma_{\upsilon\upsilon} \end{bmatrix}. \tag{5}$$

The diagonal-elements of $\Sigma_{\upsilon\upsilon}$ are subject to the normalization $\sigma_{ii}^{\upsilon\upsilon} = 1$, $i = 1 \ldots \ m$.

2.2 Inconsistency of Heckman-type estimators

For the model (1) through (4) Tauchmann (2005) suggests a class of system two-step estimators that – analogously to the original Heckman two-step approach – con-

ditions on d_{it} equation-by-equation. That is, after first-step estimation of the vectors α_i by univariate or multivariate probit, the second-step regressions yielding estimates for the vectors β_i are based on the conditional expectations[3] $\mathrm{E}(y_{it}|x_{it}, d_{it}) = d_{it}x'_{it}\beta_i + d_{it}\sigma_{ii}^{\varepsilon\upsilon}\lambda(z'_{it}\alpha_i)$. Each regression equation, therefore, includes the inverse Mills ratio $\lambda(z'_{it}\widehat{\alpha}_i)$ as an auxiliary regressor and the parameters $\sigma_{ii}^{\varepsilon\upsilon}$ are estimated as regression coefficients. Note that d_{it} serves as a weighting variable, i.e. censored observations are weighted by zero and are therefore effectively excluded from the regression.[4]

Tauchmann (2005) distinguishes three variants of this estimator: The first one uses ordinary least squares (*OLS*) and ignores cross-equation correlation of ε_{it}, another variant accounts for it in a simplified *SUR* fashion, and a third accounts for cross-equation correlation and heteroscedasticity using a proper generalized least squares (*GLS*) approach.[5]

In order to analyze these estimators' properties, we consider α as known and focus on the second-step regression. Let X denote the stacked, $mT \times mk$ regressor-matrix[6] consisting of rows $\left[\mathbf{0}_{1\times k(i-1)}\ x'_{it}\ \lambda(z'_{it}\alpha_i)\ \mathbf{0}_{1\times k(m-i)}\right]$. Note that inverse Mills ratios are included to the list of regressors. Let D denote a $mT \times mT$ matrix with diagonal-elements d_{it} and zero off-diagonal elements. This matrix allocates zero weight to censored units. Ω denotes the $mT \times mT$ block-diagonal weighting-matrix with elements

[3]To simplify notation, $\mathrm{E}(y_{it}|x_{it}, d_{it} = 1)$ is used as short term for $\mathrm{E}(y_{it}|x_{it}, \upsilon_{it} > -z'_{it}\alpha_i)$ throughout this paper. Yet, it does *not* denote $\mathrm{E}_z[\mathrm{E}(y_{it}|x_{it}, \upsilon_{it} > -z'_{it}\alpha_i)]$, although z_{it} is not explicitly mentioned in list of the conditioning variables. This analogously applies to any moment that is conditional on either $d_{it} = 1$, $d_{it}d_{jt} = 1$, d_{it}, $d_{it}d_{jt}$, or d_t.

[4]Because of (4), which implies $\mathrm{E}(y_{it}|x_{it}, d_{it} = 0) = 0$, the original Heckman (1976 and 1979) estimator can well be interpreted as a procedure that conditions on d_{it} in the full sample and, therefore, uses d_{it} as a weighting variable rather than an estimation procedure that conditions on $d_{it} = 1$ and uses the sub-sample of selected units; see Tauchmann (2005) for details.

[5]Because of $\mathrm{var}(\varepsilon_{it}|d_{it} = 1) = \sigma_{ii}^{\varepsilon\varepsilon}\left(\left(1 - \sigma_{ii}^{\varepsilon\upsilon 2}\sigma_{ii}^{\varepsilon\varepsilon -1}\right) + \sigma_{ii}^{\varepsilon\upsilon 2}\sigma_{ii}^{\varepsilon\varepsilon -1}\left(1 - z'_{it}\alpha_i\lambda(z'_{it}\alpha_i) - \lambda(z'_{it}\alpha_i)^2\right)\right)$, cf. Heckman (1976), the errors are heteroscedastic and *SUR* is not a proper *GLS* estimator.

[6]k_i denotes the number of coefficients in the ith equation. In order to simplify notation, yet with no loss of generality, we assume $k_i = k$ for $i = 1, \ldots, m$. The matrix X is arranged as such that all m rows belonging to an individual t adjoin each other.

ω_{ijt}. It coincides with the identity-matrix if the model is estimated using the classical Heckman approach equation-by-equation, i.e. *OLS*. In the case of *SUR* estimation, the individual $m \times m$ sub-matrices Ω_t are uniform across all t. In the case of *GLS* estimation, these weighting matrices are individually derived through matrix-inversion from estimates for $\text{var}(\varepsilon_{it}|d_{it}=1)$ and $\text{cov}(\varepsilon_{it},\varepsilon_{jt}|d_{it}d_{jt}=1)$. Finally, let Y denote the stacked $mT \times 1$ vector of dependent variables y_{it} and $\widetilde{\varepsilon}$ denote the corresponding $mT \times 1$ error-vector. Because of the inclusion of $\lambda(z_{it}'\alpha_i)$ to the list of regressors and $\text{E}(\varepsilon_{it}|d_{it}=1) = \sigma_{ii}^{\varepsilon v}\lambda(z_{it}'\alpha_i)$, the error vector $\widetilde{\varepsilon}$ consists of elements $\varepsilon_{it} - \text{E}(\varepsilon_{it}|d_{it}=1)$ rather than ε_{it}. Now the generalized Heckman-estimators for $\widehat{\beta}$ proposed by Tauchmann (2005) can be written

$$\widehat{\beta} = (X'D\Omega DX)^{-1}X'D\Omega Y. \tag{6}$$

Because of $Y = D(X\beta + \widetilde{\varepsilon})$ equation (6) is equivalent to

$$\widehat{\beta} = \beta + (X'D\Omega DX)^{-1}X'\Xi, \qquad \text{with} \quad \Xi \equiv D\Omega D\widetilde{\varepsilon}. \tag{7}$$

Here, the condition $\text{E}(\Xi|X) = \mathbf{0}$ implies $\text{plim}\, T^{-1}(X'\Xi) = \mathbf{0}$ and, therefore, implies consistency of $\widehat{\beta}$ under standard regularity conditions. To check whether $\text{E}(\Xi|X) = \mathbf{0}$ holds, consider an arbitrary element from Ξ:

$$\begin{aligned} \xi_{it} &= \omega_{iit}d_{it}\widetilde{\varepsilon}_{it} + \sum_{j\neq i}\omega_{ijt}d_{it}d_{jt}\widetilde{\varepsilon}_{jt} \\ &= \omega_{iit}d_{it}[\varepsilon_{it} - \text{E}(\varepsilon_{it}|d_{it}=1)] + \sum_{j\neq i}\omega_{ijt}d_{it}d_{jt}[\varepsilon_{jt} - \text{E}(\varepsilon_{jt}|d_{jt}=1)]. \end{aligned} \tag{8}$$

We apply the law of iterated expectations to (8). First, we take the expectation of ξ_{it} conditional on x_t as well as on the individual selection pattern d_t.

$$\text{E}(\xi_{it}|x_t,d_t) = \omega_{iit}d_{it}[\text{E}(\varepsilon_{it}|d_t) - \text{E}(\varepsilon_{it}|d_{it}=1)] + \sum_{j\neq i}\omega_{ijt}d_{it}d_{jt}[\text{E}(\varepsilon_{jt}|d_t) - \text{E}(\varepsilon_{jt}|d_{jt}=1)] \tag{9}$$

Second, we take the expectation with respect to d_t, yielding

$$\text{E}(\xi_{it}|x_t) = \sum_{j\neq i}\omega_{ijt}\text{Pr}(d_{it}d_{jt}=1)[\text{E}(\varepsilon_{jt}|d_{it}d_{jt}=1) - \text{E}(\varepsilon_{jt}|d_{jt}=1)]. \tag{10}$$

From (10) it becomes obvious that the estimator $\widehat{\beta}$ is biased and inconsistent unless either (i) $\mathrm{E}(\varepsilon_{jt}|d_{it}d_{jt} = 1)$ equals $\mathrm{E}(\varepsilon_{jt}|d_{jt} = 1)$ for any pair $i \neq j$ and any t; that is $\mathrm{E}(\varepsilon_{it}|d_t)$ exclusively depends on d_{it}, yet does not depend on any $d_{jt},\ j \neq i$. This requires that Σ_{vv} as well as $\Sigma_{\varepsilon v}$ are diagonal matrices. The estimator $\widehat{\beta}$ does also not suffer from inconsistency if (ii) $\omega_{ijt} = 0$ holds for all $i \neq j$ and t. Condition (ii) implies that equation-by-equation Heckman is consistent, since cross-equation correlations are not taken into account. Yet, in contrast, any system estimator that involves non-zero weights ω_{ijt} is inconsistent, unless Σ_{vv} as well as $\Sigma_{\varepsilon v}$ are diagonal matrices. Clearly, the inconsistency of $\widehat{\beta}$ originates from conditioning on d_{it} equation-by-equation.

3 A consistent generalized Heckman estimator

The above discussion clearly suggests, how to construct a consistent system-estimator as generalization to the original Heckman-approach. From (9) follows that if $\widetilde{\varepsilon}_{it}$ were defined as $\varepsilon_{it} - \mathrm{E}(\varepsilon_{it}|d_t)$ rather than $\varepsilon_{it} - \mathrm{E}(\varepsilon_{it}|d_{it} = 1)$, the condition $\mathrm{E}(\xi_{it}|x_t, d_t) = 0$ and subsequently $\mathrm{E}(\xi_{it}|x_t) = 0$ would be satisfied for any weighting matrix Ω, rendering the entire class of estimators consistent. Uniformly conditioning on d_t, i.e. conditioning on the entire selection pattern, in all equations rather than conditioning on d_{it} equation-by-equation and, correspondingly, including $\mathrm{E}(\varepsilon_{it}|d_t)$ rather than the inverse Mills ratio as correction-term would lead to errors defined as $\varepsilon_{it} - \mathrm{E}(\varepsilon_{it}|d_t)$. That is, the regression must be based on the conditional mean $\mathrm{E}(y_{it}|x_{it}, d_t)$ rather than $\mathrm{E}(y_{it}|x_{it}, d_{it})$.

In order to implement this estimator, an expression for $\mathrm{E}(\varepsilon_{it}|d_t)$ is required. It is easily shown that

$$\mathrm{E}(\varepsilon_t|d_t) = \mathrm{E}(\varepsilon_t) + \Sigma_{\varepsilon v}(\Sigma_{vv})^{-1}[\mathrm{E}(\upsilon_t|d_t) - \mathrm{E}(\upsilon_t)] \tag{11}$$

holds. Since the unconditional expectations of ε_t and υ_t equal zero, the expression reduces to a linear-combination of truncated first moments $\mathrm{E}(\upsilon_t|d_t)$ from the multivariate normal distribution. Therefore, in each regression equation m truncated means

from the multivariate normal distribution have to be included to correct for sample-selection bias. Results for these truncated means are provided by Tallis (1961) as well as for the special case $m = 2$ – albeit in more detail – by Shah & Parikh (1964). Including these expressions and rearranging terms leads to the system of regression equations

$$y_{it} = d_{it}x'_{it}\beta_i + d_{it}\sum_{j=1}^{m}\delta_{ij}\psi_{jt}\phi(z'_{jt}\alpha_j)\frac{\Phi^{m-1}(\widetilde{A}_{jt},\widetilde{R}_{jt})}{\Phi^m(\bullet)} + d_{it}\widetilde{\varepsilon}_{it}, \qquad i = 1,...,m. \tag{12}$$

As in the original Heckman-model, the coefficients δ_{ij} attached to the correction-terms $\psi_{jt}\phi(z'_{jt}\alpha_j)\frac{\Phi^{m-1}(\widetilde{A}_{jt},\widetilde{R}_{jt})}{\Phi^m(\bullet)}$ are subject to estimation. Here, ϕ denotes the probability density function of the univariate standard normal distribution, while Φ^m denotes the cumulative density function of the m-variate standard normal distribution. ψ_{jt} is defined as $2d_{jt}-1$ and distinguishes truncation from either below or above. $\widetilde{A}_{jt}$ represents a vector which consists of $m-1$ elements $\frac{\psi_{lt}(z'_{lt}\alpha_l - \sigma^{vv}_{lj}z'_{jt}\alpha_j)}{(1-(\sigma^{vv}_{lj})^2)^{1/2}}$; $l = 1\ \ldots m$, $l \neq j$. Correspondingly, $\widetilde{R}_{jt}$ is defined as $\Psi_{jt}R_{jt}\Psi_{jt}$, where R_{jt} denotes the partial conditional correlation-matrix $\mathrm{Cor}(\upsilon_t|\upsilon_{jt})$ and Ψ_{jt} denotes a diagonal-matrix with diagonal elements ψ_{lt}, $l \neq j$. Finally, $\Phi^m(\bullet)$ denotes the joint probability of the observed pattern d_t. Note that the regression equations are still weighted by d_{it}.[7]

In applied work α and Σ_{vv} are likely to be unknown. In order to calculate the auxiliary regressors $\psi_{jt}\phi(z'_{jt}\alpha_j)\frac{\Phi^{m-1}(\widetilde{A}_{jt},\widetilde{R}_{jt})}{\Phi^m(\bullet)}$, one has to replace the true parameters with estimates obtained from first-step multivariate probit estimation. In the special case $m = 2$ the regression equations are equivalent to the one used by Poirier (1980), except for the fact that Poirier (1980) conditions on $d_{1t}d_{2t} = 1$ rather than d_{1t} and d_{2t}, i.e. ψ_{jt} equals one for all j and t.[8] For $m = 2$, $\delta_{ij} = \sigma^{\varepsilon v}_{ij}$ holds for the auxiliary regression coefficients.

One may estimate the system (12) equation-by-equation using *OLS*. Yet, the simple equation-by-equation Heckman-estimator is consistent as well in this case. So, condi-

[7]Since $\mathrm{E}(y_{it}|x_{it}, d_t, d_{it} = 0) = 0$ holds, the ith equation of the tth observation still receives zero weight if y_{it} equals zero because of censoring.

[8]See Vella (1997) for other related models.

tioning on d_t makes sense only in the context of simultaneous estimation. As a simple variant, one can construct such a system-estimator in the standard *SUR* fashion. However, this ignores the heteroscedasticity of the individual conditional error-variances. In order to be able to construct a proper *GLS* estimator, expressions for $\mathrm{Var}(\varepsilon_t|d_t)$ are required from which one can calculate an appropriate weighting matrix Ω. Through the use of the normality assumption and the decomposition rule for variances in a joint distribution such an expression can easily be derived as

$$\mathrm{Var}(\varepsilon_t|d_t) = \Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon v}(\Sigma_{vv})^{-1}\Sigma'_{\varepsilon v} + \Sigma_{\varepsilon v}(\Sigma_{vv})^{-1}\mathrm{Var}(\upsilon_t|d_t)(\Sigma_{vv})^{-1}\Sigma'_{\varepsilon v}. \qquad (13)$$

Obviously, any element of $\mathrm{Var}(\varepsilon_t|d_t)$ is a linear function of all elements of the truncated m-variate normal variance-covariance matrix $\mathrm{Var}(\upsilon_t|d_t)$. Therefore, estimates for the elements of $\mathrm{Var}(\varepsilon_t|d_t)$ can be obtained as fitted values from regressing squared residuals and residual cross-products – which, in turn, have been obtained from initial *OLS* regressions – on a constant and on estimates for all elements of $\mathrm{Var}(\upsilon_t|d_t)$.[9] Results for the latter ones are provided by Tallis (1961). Therefore, with estimates for α and Σ_{vv} in hand, one can calculate these auxiliary regressors.

4 Monte-Carlo analysis

In addition to the theoretical analysis we carry out Monte-Carlo simulations. On the one hand, we want to illustrate the theoretical results derived in Section 2. Test results on the joint unbiasedness of the second-step coefficients are provided for this purpose.[10]

[9] Because of $\mathrm{var}(d_{it}\varepsilon_{it}|d_{it}=0)=0$, the variance-covariance matrix $\mathrm{Var}(d_{1t}\varepsilon_{1t}\ldots d_{mt}\varepsilon_{mt}|d_t)$ that is effectively required for the construction of the *GLS* estimator in general is short-ranked and cannot ordinarily be inverted in order to obtain individual weighting-matrices Ω_t. Yet, using a generalized Moore-Penrose inverse is appropriate for this purpose.

[10] Tables of raw coefficients' estimates are provided in the appendix. The *LIMDEP* command file used for carrying out the *MC*-simulations is available from the author upon request.

On the other, we also want to address the estimators' performance beyond the issue of consistency. Therefore, we present estimates for the *CP*-conditional mean square error prediction criterion

$$\mathrm{CP}(\widehat{\beta}) = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m}(\beta_i - \widehat{\beta}_i)' x_{it} x_{it}' (\beta_i - \widehat{\beta}_i)\middle| X\right], \tag{14}$$

cf. Judge et al. (1980). $\mathrm{CP}(\widehat{\beta})$ measures the mean squared deviation of the estimated conditional mean from its true counterpart $\mathrm{E}(y_{it}^*|x_{it})$ and, therefore, translates an estimator's *MSE*-matrix to a scalar performance measure that takes into account its variance as well as a potential bias.

Unknown values for α and Σ_{vv} rather than known ones appear to be the relevant case from the viewpoint of applied econometrics. In our Monte-Carlo simulations, therefore, these parameters are estimated by first-step probit models. We consider six different estimators. In particular, conditioning on either d_{it} or d_t is combined with *OLS*, *SUR* and, finally *GLS* estimation. Conditioning on d_{it} combined with *OLS* or *SUR* allows for estimating the first step using univariate probit models. All other estimators require simultaneous estimation of all vectors α_i along with Σ_{vv}.

4.1 The experimental setup

The design of the Monte-Carlo experiment is equivalent to the one used by Tauchmann (2005).[11] We consider the case $m = 2$.[12] The sample size is 4000. The size of the Monte Carlo experiment is 1000 iterations. The vectors of exogenous variables each consist of three elements:

$$z_{it} = [1\ z_{1,it}\ z_{2,it}]', \quad x_{it} = [1\ x_{1,it}\ x_{2,it}]', \quad i = 1, 2.$$

[11] In contrast to the analysis presented here, Tauchmann (2005) imposes restrictions on the coefficients' estimates $\widehat{\beta}_i$. This does not allow for directly comparing estimated *CP*-measures.

[12] For $m \geq 3$, simulated *ML* were required for estimation the first-step multivariate probit models. This would increase computing time for the Monte-Carlo experiments enormously.

Here $z_{1,1t}$, $z_{2,1t}$, $z_{1,2t}$, and $x_{2,1t}$ are independently drawn from the standard normal distribution, while $z_{2,2t} = z_{2,1t}$, $x_{1,1t} = z_{1,1t}$, $x_{1,2t} = z_{1,2t}$ and $x_{2,2t} = x_{2,1t}$. These variables are drawn only once and then kept fixed. For the coefficient vectors $\beta_i = [1\ 1\ 1]'$, $i = 1, 2$ holds.[13]

The value $\sqrt{0.5}$ is assigned to all coefficients α attached to $z_{1,it}$ and $z_{2,it}$. In order to allow for different unconditional censoring probabilities $\Pr(d_{it}^* \leq 0)$, the constants $\alpha_{0,i}$ are varied. We run two simulations with unconditional censoring probabilities that are uniform across equations, in particular 0.25 and 0.5, which corresponds to constants 0.9539 and 0, respectively. Another simulation is carried out for mixed unconditional censoring probabilities, i.e. 0.25 for equation one and 0.75 for equation two, which corresponds to constants 0.9539 and -0.9539, respectively. The error-covariance structure is specified as

$$\Sigma_{\varepsilon\varepsilon} = \begin{bmatrix} 1.5 & \\ -1 & 2 \end{bmatrix},\ \Sigma_{vv} = \begin{bmatrix} 1 & \\ -0.5 & 1 \end{bmatrix},\ \Sigma_{\varepsilon v} = \begin{bmatrix} 0.75 & -0.25 \\ -0.25 & 0.75 \end{bmatrix}.$$

As an alternative specification, the value zero is assigned to all off-diagonal elements of Σ_{vv} and $\Sigma_{\varepsilon v}$ everything else remaining unchanged, i.e.

$$\Sigma_{\varepsilon\varepsilon} = \begin{bmatrix} 1.5 & \\ -1 & 2 \end{bmatrix},\ \Sigma_{vv} = \begin{bmatrix} 1 & \\ 0 & 1 \end{bmatrix},\ \Sigma_{\varepsilon v} = \begin{bmatrix} 0.75 & 0 \\ 0 & 0.75 \end{bmatrix}.$$

This defines the four-variate $N(0, \Sigma)$ distribution, from where the random components are drawn separately for each model. After drawing the error vector, the dependent variables are calculated as defined by model (1) through (4). Subsequently, the generated data serves as input to the estimators.

4.2 Simulation results

Results for Wald-tests on the unbiasedness of the six estimators are displayed in Table 1. These simulation results are consistent with the theoretical ones, obtained

[13] We do not vary these parameters, since – in contrast to the estimator proposed by Shonkwiler & Yen (1999) – the performance of generalized Heckman estimators does not depend on the true value of β, c.f. Tauchmann (2005).

Table 1: **Tests on joint unbiasedness of regression coefficients**

	OLS	*SUR*	*GLS*	*OLS*	*SUR*	*GLS*
	conditional on d_{it}			**conditional on** d_t		
dense error variance-covariance matrix						
censoring prob. 0.25	0.800	0.000	0.000	0.449	0.484	0.155
censoring prob. 0.5	0.545	0.000	0.000	0.964	0.070	0.642
censoring prob. 0.25 and 0.75	0.446	0.000	0.000	0.117	0.929	0.259
Σ_{vv} and $\Sigma_{\varepsilon v}$ with zero off-diagonal elements						
censoring prob. 0.25	0.320	0.415	0.052	0.805	0.208	0.082
censoring prob. 0.5	0.595	0.659	0.610	0.900	0.760	0.807
censoring prob. 0.25 and 0.75	0.832	0.620	0.604	0.963	0.634	0.215

Note: P-values for Wald-tests reported.

in Section 2. If Σ_{vv} and $\Sigma_{\varepsilon v}$ are dense matrices, unbiasedness is clearly rejected for those estimators that condition on d_{it} equation-by-equation and use *SUR* or *GLS*. In contrast, the classical Heckman estimator employed equation-by-equation does not display a significant bias. The estimators that condition on the entire selection pattern do not display a significant bias either. If, instead, Σ_{vv} and $\Sigma_{\varepsilon v}$ are diagonal-matrices, neither of the estimators display a bias that is significant at the 0.05-level. Therefore, the Monte-Carlo simulation confirms that system-estimators that condition on d_t are consistent, while system-estimators that condition on d_{it} equation-by-equation are biased, unless certain restrictions apply to the true error-covariance matrix.

In order to analyze the estimators' performance beyond the issue of unbiasedness, estimates for the *CP*-conditional mean square error prediction criterion are displayed in Table 2. Comparing the *SUR* estimator that conditions on d_t with its counterpart that conditions on d_{it} yields the following plausible result: If the true covariance-matrix is dense, the consistent estimator that conditions on d_t yields smaller *CP*-measures than the inconsistent one that conditions on d_{it}. If Σ_{vv} and $\Sigma_{\varepsilon v}$ are diagonal-matrices – i.e. both estimators are consistent – the more parsimoniously parameterized one that

Table 2: **Estimated conditional mean square prediction errors**

	OLS	*SUR*	*GLS*	*OLS*	*SUR*	*GLS*
	conditional on d_{it}			**conditional on** d_t		
dense error variance-covariance matrix						
censoring prob. 0.25	6.499	6.038	6.455	5.826	5.468	5.357
	(0.154)	(0.149)	(0.169)	(0.130)	(0.137)	(0.140)
censoring prob. 0.5	12.810	13.444	15.816	12.275	11.776	11.178
	(0.342)	(0.377)	(0.483)	(0.309)	(0.335)	(0.330)
cens. prob. 0.25 & 0.75	23.747	23.129	35.492	23.095	21.877	19.387
	(0.806)	(0.755)	(1.391)	(0.784)	(0.762)	(0.786)
Σ_{vv} and $\Sigma_{\varepsilon v}$ with zero off-diagonal elements						
censoring prob. 0.25	6.420	5.520	5.872	6.304	5.452	5.269
	(0.156)	(0.135)	(0.212)	(0.147)	(0.138)	(0.140)
censoring prob. 0.5	13.451	11.776	15.880	13.533	11.932	11.702
	(0.386)	(0.346)	(1.601)	(0.364)	(0.324)	(0.352)
cens. prob. 0.25 & 0.75	23.219	20.627	28.690	24.156	21.091	17.748
	(0.768)	(0.744)	(2.733)	(0.853)	(0.710)	(0.655)

Notes: Standard errors in parenthesis.

Displayed *CP*-measures are scaled by the factor 1000.

conditions on d_{it} performs better except for one simulation. Yet, the latter differences in estimated *CP*-measures are statistically insignificant at the 0.05-level.

The comparison of *OLS* estimators that either condition on d_t or d_{it} yields similar results. If the error-covariance matrix is dense, the first estimator seems to perform better, though both are consistent. If, instead, Σ_{vv} and $\Sigma_{\varepsilon v}$ are diagonal-matrices the latter displays smaller *CP*-measures. However, these differences never are statistically significant, except for one simulation.

Finally, we examine the performance of *GLS* estimators. Here, we observe substantial deviations in estimated *CP*-measures. While *GLS* conditional on d_t yields the

smallest mean square prediction error among all considered estimators in any simulation, *GLS* conditional on d_{it}, except for two simulations, displays the largest one. Moreover, the deviations in *CP*-measures between both *GLS* estimators always are significant. In fact, if the error covariance-matrix is dense, *GLS* conditional on d_t significantly outperforms any other estimator in any simulation. As the only exception to this result, in some cases *SUR* conditional on d_t displays *CP*-measures which are not significantly lager.

Our key simulation result – that *GLS* conditional on d_t displays the best performance in terms of the mean square prediction error – fits theory. Among the considered estimators, *GLS* conditional on d_t is the only one that not only is consistent, but also as efficiently accounts for cross-equation correlation and heteroscedasticity.

5 Conclusions

This analysis of estimation procedures for the multivariate sample-selection model shows that multivariate generalizations to the classical Heckman (1976 and 1979) two-step approach that account for cross-equation correlation and use the inverse Mills ratio as a correction-term are consistent only if certain restrictions apply to the true error-covariance structure. However, generalizations to the classical Heckman two-step estimator that condition on the entire selection pattern rather than the selection of particular single equations – and, therefore, use generalized correction-terms – are shown to be generally consistent. Moreover, these estimators display a smaller mean square prediction error. These new estimators are computationally more demanding since they generally require simultaneous estimation of a multivariate probit model. Nowadays, however, hard-coded procedures for this estimation problem are provided by econometric software packages, rendering computational complexity a minor obstacle to the practical application of the suggested estimation procedure.

Finally, we discuss how our results fit into the general debate on which estimator

is the best choice for estimating the multivariate sample selection model. If efficiency is the major concern and numerical complexity and computing time do not matter, then two-step approaches – including those suggested in this analysis – are generally to be avoided, and full information maximum likelihood as proposed by Yen (2005) is the best choice. If, in contrast, computational simplicity and consistency is the major concern, then equation-by-equation Heckman appears to be the best choice. If a small mean square error and computational simplicity are a researcher's main criteria, while consistency is of secondary relevance, one might even argue in favor of the inconsistent *SUR* estimator that conditions equation-by-equation on the outcome of the upstream choice problem. Finally, if both consistency and a small mean square error are desired, and the computational burden of full information maximum likelihood is to be avoided, then the *GLS* estimator that conditions on the entire selection pattern appears to be the best choice.

References

Chen Z. and S.T. Yen (2005): On bias correction in the multivariate sample-selection model, *Applied Economics* 37, 2459-2468.

Gao X.M. E.J. Wailes and G.L. Cramer (1995): A Microeconometric Model Analysis of US Consumer Demand for Alcoholic Beverages, *Applied Economics* 27, 59-69.

Goodwin B.K. M.L. Vandeveer and J.L. Deal (2004): An Empirical Analysis of Acreage Effects of Participation in the Federal Crop Insurance Program, *American Journal of Agricultural Economics* 86, 1058-1077.

Heckman J.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 Economics and Social Measurement* 5, 475-492.

Heckman J.J. (1979): Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-161.

Heien D. and C. Durham (1991): A Test of the Habit Formation Hypothesis using Household Data, *Review of Economics and Statistics* 73, 189-199.

Heien D. and C.R. Wessells (1990): Demand Systems Estimation with Microdata: A Censored Regression Approach, *Journal of Business & Economic Statistics* 8, 365-371.

Judge G.G. et al. (1980): The Theory and Practice of Econometrics, 2nd ed., John Wiley, New York.

Nayga R.M. B.J. Tepper and L. Rosenzweig (1999): Assessing the Importance of Health and Nutrition related Factors on Food Demand: a Variable Preference Investigation, *Applied Economics* 31, 1541-1549.

Poirier D.J. (1980): Partial Observability in Bivariate Probit Models, *Journal of Econometrics* 12, 209-217.

Shah S.M. and N.T. Parikh (1964): Moments of Singly and Doubly Truncated Standard Bivariate Normal Distribution, *Vidya* 7, 51-91.

Shonkwiler J.S. and S.T. Yen (1999): Two-Step Estimation of a Censored System of Equations, *American Journal of Agricultural Economics* 81, 972-982.

Su S.-J. and S.T. Yen (2000): A Censored System of Cigarette and Alcohol Consumption, *Applied Economics* 32, 729-737.

Tallis G.M. (1961): The Moment Generating Function of the Truncated Multi-Normal Distribution, *Journal of the Royal Statistical Society* (Series B) 23, 223-229.

Tauchmann H. (2005): Efficiency of two-step Estimators for Censored Systems of Equations: Shonkwiler and Yen reconsidered, *Applied Economics* 37, 367-374.

Vella F. (1997): Estimating Models with Sample Selection Bias: A Survey, *Journal of Human Resources* 33, 127-169.

Vermeulen F. (2001): A Note on Heckman-type Corrections in Models for Zero Expenditures, *Applied Economics* 33, 1089-1092.

Yen S.T. K. Kah and S.-J. Su (2002): Household Demand for Fat and Oils: Two-step Estimation of a Censored Demand System, *Applied Economics* 34, 1799-1806.

Yen S.T. (2005): A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol demand with Zero Observations, *American Journal of Agricultural Economics* 87, 453-466.

Zellner A. (1963): An Efficient Method of Estimating Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results, *Journal of the American Statistical Association* 58, 977-992.

Appendix: Estimated coefficients from Monte-Carlo simulations

Table 3: **Estimated coefficients: censoring prob.** 0.25**, dense error variance-covariance matrix**

		conditional on d_{it}						**conditional on** d_t					
	true value	*OLS*		*SUR*		*GLS*		*OLS*		*SUR*		*GLS*	
		mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
$\beta_{0,1}$	1	1.0002	0.0389	0.9881	0.0373	0.9859	0.0373	0.9994	0.0363	1.0009	0.0361	1.0014	0.0363
$\beta_{1,1}$	1	1.0000	0.0285	1.0041	0.0253	1.0041	0.0254	0.9989	0.0280	1.0000	0.0248	0.9991	0.0249
$\beta_{2,1}$	1	0.9993	0.0209	1.0007	0.0203	1.0005	0.0202	0.9997	0.0206	0.9993	0.0202	0.9986	0.0205
$\sigma_{11}^{\varepsilon v}$	0.75	0.7488	0.0900	0.7459	0.0873	0.7464	0.0868	0.7489	0.0848	0.7464	0.0813	0.7459	0.0830
$\sigma_{12}^{\varepsilon v}$	-0.25	–	–	–	–	–	–	-0.2505	0.0369	-0.2496	0.0375	-0.2499	0.0362
$\beta_{0,2}$	1	1.0015	0.0458	0.9850	0.0427	0.9756	0.0424	1.0020	0.0413	1.0001	0.0409	0.9993	0.0419
$\beta_{1,2}$	1	1.0003	0.0325	1.0071	0.0285	1.0092	0.0283	1.0004	0.0317	1.0016	0.0303	1.0005	0.0290
$\beta_{2,2}$	1	1.0002	0.0246	0.9993	0.0238	0.9998	0.0234	1.0011	0.0250	1.0003	0.0240	0.9992	0.0241
$\sigma_{22}^{\varepsilon v}$	0.75	0.7488	0.1079	0.7475	0.0953	0.7548	0.0979	0.7466	0.0979	0.7513	0.0930	0.7509	0.0956
$\sigma_{21}^{\varepsilon v}$	-0.25	–	–	–	–	–	–	-0.2495	0.0433	-0.2512	0.0426	-0.2490	0.0430

Table 4: **Estimated coefficients: censoring prob.** 0.5**, dense error variance-covariance matrix**

		conditional on d_{it}						**conditional on** d_t					
	true value	*OLS*		*SUR*		*GLS*		*OLS*		*SUR*		*GLS*	
		mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
$\beta_{0,1}$	1	0.9963	0.0596	0.9886	0.0586	0.9670	0.0641	0.9993	0.0587	1.0014	0.0604	0.9994	0.0586
$\beta_{1,1}$	1	1.0016	0.0351	1.0036	0.0339	1.0079	0.0346	1.0006	0.0361	0.9992	0.0334	1.0008	0.0327
$\beta_{2,1}$	1	1.0006	0.0250	1.0001	0.0249	1.0000	0.0234	0.9995	0.0256	1.0013	0.0245	0.9999	0.0239
$\sigma_{11}^{\varepsilon v}$	0.75	0.7543	0.0783	0.7579	0.0763	0.7793	0.0846	0.7507	0.0759	0.7455	0.0765	0.7507	0.0736
$\sigma_{12}^{\varepsilon v}$	-0.25	–	–	–	–	–	–	-0.2492	0.0411	-0.2466	0.0410	-0.2513	0.0402
$\beta_{0,2}$	1	1.0031	0.0699	0.9803	0.0734	0.9628	0.0707	1.0003	0.0681	0.9988	0.0650	1.0037	0.0650
$\beta_{1,2}$	1	0.9984	0.0413	1.0079	0.0398	1.0087	0.0418	1.0008	0.0380	1.0005	0.0382	0.9987	0.0369
$\beta_{2,2}$	1	1.0008	0.0295	0.9999	0.0278	0.9990	0.0269	0.9998	0.0294	0.9989	0.0282	0.9998	0.0278
$\sigma_{22}^{\varepsilon v}$	0.75	0.7472	0.0911	0.7660	0.0977	0.7824	0.0917	0.7490	0.0879	0.7511	0.0867	0.7482	0.0829
$\sigma_{21}^{\varepsilon v}$	-0.25	–	–	–	–	–	–	-0.2510	0.0476	-0.2476	0.0477	-0.2489	0.0449

Table 5: **Estimated coefficients: censoring prob.** 0.25 **and** 0.75**, dense error variance-covariance matrix**

		conditional on d_{it}						**conditional on** d_t					
	true value	*OLS*		*SUR*		*GLS*		*OLS*		*SUR*		*GLS*	
		mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
$\beta_{0,1}$	1	0.9984	0.0396	0.9975	0.0386	0.9963	0.0385	1.0001	0.0373	1.0005	0.0363	1.0005	0.0378
$\beta_{1,1}$	1	1.0011	0.0283	1.0010	0.0278	0.9964	0.0281	1.0009	0.0272	0.9995	0.0274	0.9990	0.0278
$\beta_{2,1}$	1	0.9991	0.0207	1.0000	0.0205	0.9999	0.0213	1.0004	0.0211	1.0008	0.0215	0.9993	0.0209
$\sigma_{11}^{\varepsilon v}$	0.75	0.7550	0.0937	0.7537	0.0915	0.7592	0.0948	0.7489	0.0850	0.7490	0.0848	0.7498	0.0852
$\sigma_{12}^{\varepsilon v}$	-0.25	–	–	–	–	–	–	-0.2485	0.0354	-0.2484	0.0351	-0.2505	0.0351
$\beta_{0,2}$	1	1.0060	0.1275	0.9725	0.1240	0.8927	0.1262	1.0057	0.1259	0.9992	0.1235	0.9983	0.1171
$\beta_{1,2}$	1	0.9979	0.0566	1.0090	0.0542	1.0311	0.0560	1.0014	0.0572	1.0009	0.0527	1.0022	0.0529
$\beta_{2,2}$	1	1.0005	0.0410	0.9998	0.0391	0.9983	0.0394	0.9992	0.0401	0.9996	0.0391	0.9984	0.0374
$\sigma_{22}^{\varepsilon v}$	0.75	0.7440	0.1058	0.7564	0.1014	0.7775	0.1113	0.7444	0.1056	0.7495	0.1011	0.7533	0.0981
$\sigma_{21}^{\varepsilon v}$	-0.25	–	–	–	–	–	–	-0.2468	0.0745	-0.2493	0.0720	-0.2496	0.075

Table 6: **Estimated coefficients: censoring prob.** 0.25**, diagonal-matrices** $\Sigma_{\upsilon\upsilon}$ **and** $\Sigma_{\varepsilon\upsilon}$

		conditional on d_{it}						**conditional on** d_t					
	true value	*OLS*		*SUR*		*GLS*		*OLS*		*SUR*		*GLS*	
		mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
$\beta_{0,1}$	1	0.9985	0.0626	1.0025	0.0602	0.9978	0.0704	1.0005	0.0608	0.9997	0.0585	0.9978	0.0592
$\beta_{1,1}$	1	1.0012	0.0355	0.9984	0.0326	1.0014	0.0395	0.9995	0.0356	1.0009	0.0323	1.0008	0.0327
$\beta_{2,1}$	1	1.0010	0.0249	1.0002	0.0231	0.9997	0.0241	1.0005	0.0258	1.0000	0.0242	0.9988	0.0238
$\sigma_{11}^{\varepsilon\upsilon}$	0.75	0.7528	0.0787	0.7464	0.0774	0.7543	0.1000	0.7511	0.0797	0.7494	0.0728	0.7522	0.0742
$\sigma_{12}^{\varepsilon\upsilon}$	0	–	–	–	–	–	–	0.0012	0.0372	0.0007	0.0369	0.0000	0.0367
$\beta_{0,2}$	1	1.0015	0.0721	1.0003	0.0668	0.9958	0.0771	0.9974	0.0730	1.0004	0.0693	1.0029	0.0685
$\beta_{1,2}$	1	0.9991	0.0410	0.9993	0.0373	1.0025	0.0460	1.0007	0.0422	0.9990	0.0379	0.9997	0.0368
$\beta_{2,2}$	1	0.9997	0.0295	0.9992	0.0277	0.9995	0.0288	1.0001	0.0297	0.9989	0.0271	1.0008	0.0283
$\sigma_{22}^{\varepsilon\upsilon}$	0.75	0.7498	0.0946	0.7487	0.0854	0.7566	0.1076	0.7524	0.0959	0.7505	0.0891	0.7474	0.0871
$\sigma_{21}^{\varepsilon\upsilon}$	0	–	–	–	–	–	–	-0.0009	0.0427	0.0009	0.0429	0.0010	0.0447

Table 7: **Estimated coefficients: censoring prob.** 0.5**, diagonal-matrices** Σ_{vv} **and** $\Sigma_{\varepsilon v}$

		conditional on d_{it}						**conditional on** d_t					
	true value	*OLS*		*SUR*		*GLS*		*OLS*		*SUR*		*GLS*	
		mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
$\beta_{0,1}$	1	0.9985	0.0626	1.0025	0.0602	0.9978	0.0704	1.0005	0.0608	0.9997	0.0585	0.9978	0.0592
$\beta_{1,1}$	1	1.0012	0.0355	0.9984	0.0326	1.0014	0.0395	0.9995	0.0356	1.0009	0.0323	1.0008	0.0327
$\beta_{2,1}$	1	1.0010	0.0249	1.0002	0.0231	0.9997	0.0241	1.0005	0.0258	1.0000	0.0242	0.9988	0.0238
$\sigma_{11}^{\varepsilon v}$	0.75	0.7528	0.0787	0.7464	0.0774	0.7543	0.1000	0.7511	0.0797	0.7494	0.0728	0.7522	0.0742
$\sigma_{12}^{\varepsilon v}$	0	–	–	–	–	–	–	0.0012	0.0372	0.0007	0.0369	0.0000	0.0367
$\beta_{0,2}$	1	1.0015	0.0721	1.0003	0.0668	0.9958	0.0771	0.9974	0.0730	1.0004	0.0693	1.0029	0.0685
$\beta_{1,2}$	1	0.9991	0.0410	0.9993	0.0373	1.0025	0.0460	1.0007	0.0422	0.9990	0.0379	0.9997	0.0368
$\beta_{2,2}$	1	0.9997	0.0295	0.9992	0.0277	0.9995	0.0288	1.0001	0.0297	0.9989	0.0271	1.0008	0.0283
$\sigma_{22}^{\varepsilon v}$	0.75	0.7498	0.0946	0.7487	0.0854	0.7566	0.1076	0.7524	0.0959	0.7505	0.0891	0.7474	0.0871
$\sigma_{21}^{\varepsilon v}$	0	–	–	–	–	–	–	-0.0009	0.0427	0.0009	0.0429	0.0010	0.0447

Table 8: **Estimated coefficients: censoring prob.** 0.25 **and** 0.75**, diagonal-matrices** Σ_{vv} **and** $\Sigma_{\varepsilon v}$

		conditional on d_{it}						**conditional on** d_t					
	true value	*OLS*		*SUR*		*GLS*		*OLS*		*SUR*		*GLS*	
		mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
$\beta_{0,1}$	1	0.9998	0.0395	0.9984	0.0385	1.0004	0.0443	1.0009	0.0383	1.0012	0.0373	1.0025	0.0385
$\beta_{1,1}$	1	1.0006	0.0281	0.9999	0.0272	1.0001	0.0338	1.0003	0.0281	0.9999	0.0274	0.9994	0.0269
$\beta_{2,1}$	1	1.0000	0.0212	0.9996	0.0206	1.0006	0.0215	0.9999	0.0211	1.0005	0.0206	0.9990	0.0208
$\sigma_{11}^{\varepsilon v}$	0.75	0.7502	0.0900	0.7524	0.0876	0.7502	0.1129	0.7467	0.0880	0.7480	0.0865	0.7454	0.0892
$\sigma_{12}^{\varepsilon v}$	0	–	–	–	–	–	–	-0.0001	0.0339	0.0001	0.0336	-0.0003	0.0335
$\beta_{0,2}$	1	0.9991	0.1267	0.9998	0.1178	0.9994	0.1393	1.0000	0.1288	0.9971	0.1205	1.0037	0.1114
$\beta_{1,2}$	1	1.0013	0.0557	0.9984	0.0530	0.9988	0.0652	1.0001	0.0586	0.9997	0.0531	0.9999	0.0491
$\beta_{2,2}$	1	0.9992	0.0386	1.0000	0.0386	0.9993	0.0419	1.0000	0.0411	0.9989	0.0374	1.0004	0.0354
$\sigma_{22}^{\varepsilon v}$	0.75	0.7523	0.1056	0.7506	0.0958	0.7531	0.1234	0.7502	0.1077	0.7507	0.0982	0.7454	0.0911
$\sigma_{21}^{\varepsilon v}$	0	–	–	–	–	–	–	0.0010	0.0743	0.0034	0.0717	0.0005	0.0756