INSTITUTE OF ACTUARIAL SCIENCE & ECONOMETRICS

REPORT AE 19/92

HETEROSCEDASTICITY IN SELECTIVITY MODELS

Hans van Ophem

University of Amsterdam

HETEROSCEDASTICITY IN SELECTIVITY MODELS.

Hans van Ophem.*)

October 1992.

*) Department of Actuarial Science and Econometrics, Faculty of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. I have benefitted from advice received from Risto Heymans and Coen Teulings.

Title: Heteroscedasticity in selectivity models

Author: Hans van Ophem

Address University of Amsterdam
Institute of Actuarial Science and Econometrics
Roetersstraat 11
1018 WB Amsterdam
The Netherlands

Date: 6 oktober 1992

Series and Number: Report AE 19/92

Pages: 32

Price: No Charge

JEL Subject Classification: C35

Keywords: heteroscedasticity, selectivity models, ML-estimation, consistency.

Abstract:

Selectivity models usually consist of two equations: a linear and a qualitative variables equation. This paper investigates the consistency of the ML-estimates of two selectivity models in which the heteroscedasticity of the error term of the linear equation is ignored. Homoscedastic estimation of a heteroscedastic model is proved to yield inconsistent estimates. Some Monte Carlo evidence is also provided.

1. Introduction.

The problem of estimating selectivity models has received considerable theoretical and empirical attention in the econometric literature. Maddala (1983, chapter 9) provides an overview. Consider the following selectivity model:[1]

$$y_i = \alpha_0 + X_i\alpha_1 + \theta I_i + \varepsilon_i \qquad (1)$$

$$I_i^* = \gamma_0 + Z_i\gamma_1 - \nu_i$$

where $I_i = 1$ iff $I_i^* \geq 0$ and $I_i = 0$ otherwise. The continuous dependent variable y_i is related to a constant, some explanatory variables contained in the vector X_i, the variable I_i equaling either 1 or 0 and an error term with zero expectation. The latent variable I_i^* depends upon a constant, a vector of explanatory variables Z_i and the error term ν_i with zero mean. This model has been utilized to measure the effect of social training programs on wages. The key parameter in that case is θ. It reflects the wage effect of the training program or, in other words, it is a measure of the effectiveness of the program. Obviously, model (1) has a very restrictive nature. An important generalization is the switching regression model (see e.g. Maddala, 1983, p. 261 or Willis and Rosen, 1979). It enhances the model by permitting the program to not only affect the constant term of the wage equation but also its slope α_1. Furthermore, it allows for a more elaborate error structure than model (1). In particular, the switching regression model does not assume equal variances across regimes. To clarify this point consider model (1) where an individual specific stochastic money return of participating in the training program θ_i replaces the constant money return θ. Assume that $\theta_i = \lambda_0 + V_i\lambda_1 + \kappa_i$, with $E\kappa_i = 0$. Rewriting model (1) yields:

$$y_i = \alpha_0 + \lambda_0 I_i + X_i\alpha_1 + V_i\lambda_1 I_i + (\varepsilon_i + \kappa_i I_i) \qquad (2)$$

$$I_i^* = \gamma_0 + Z_i\gamma_1 - \nu_i$$

If we choose $V_i = X_i$ we obtain the switching regression model. The difference between models

---

[1] This model is the starting point of the analysis in e.g. Björklund and Moffitt (1987) and Goddeeris (1988).

(1) and (2) lies not only in the addition of slope effects but also in allowing for a heteroscedastic error term. Indeed, this last generalization of model (1) is a very straightforward one. For instance, even if we relate the uncertain individual specific returns of the training program to a constant and an error term: $\theta_i = \theta +$ error term, $E\theta_i = \theta$, the resulting model would be heteroscedastic. Now, consider the estimation of model (2). We can estimate the heteroscedastic model by either full information maximum likelihood, this procedure yields consistent, asymptotically efficient parameter estimates and correct standard errors, or use a two step estimation method, which differs slightly from the one commonly encountered (see below), yielding consistent but not asymptotically efficient estimates and incorrect standard errors for the second step estimates. But what will happen if we ignore the heteroscedasticity and perform a homoscedastic maximum likelihood estimation?

With respect to ignoring heteroscedasticity the following results are established in the literature. Disregarding heteroscedasticity in a linear model will yield consistent, but inefficient, estimates. Homoscedastic estimation of a heteroscedastic probit or truncated model (cf. Yatchew and Griliches, 1984, and Hurd, 1979) will result in inconsistent estimates. For simultaneous models much less is known of the influence of heteroscedasticity on estimators of a homoscedastic specification. In this paper a more or less intermediate position is considered. The problem of homoscedastic estimation of a heteroscedastic simultaneous model will be addressed for the case that one of the endogenous variables of the model has a qualitative nature and the other endogenous variable is heteroscedastic. In Section 2 it will be proved that ignoring heteroscedasticity will frustrate the consistent estimation of model (1) in which the heteroscedasticity is assumed to have a very simple form. Section 3 will consider a variation of the well known selection model discussed in Heckman (1979) in which the linear equation is heteroscedastic. Section 4 describes some small scale Monte Carlo experiments to settle some loose ends of sections 2 and 3 and to get some insight in the magnitude of the bias of the homoscedatic ML-estimates and section 5 concludes.

2. Heteroscedasticity in a self-selection model.

Consider the following rewritten version of model (1):[2]

$$y_i = X_i\alpha + \varepsilon_i \tag{3a}$$

$$I_i^* = Z_i\gamma - \nu_i \tag{b}$$

The constants α_0 and γ_0 are now included in the vectors of explanatory variables X_i and Z_i. Furthermore, I_i may also be included in X_i. The errors ε_i and ν_i are jointly normally distributed with expectations 0, variances σ_i^2 and 1 and correlation ρ.[3] The heteroscedastic variance σ_i^2 of ε_i will be restricted to a very simple form $\sigma_i^2 = \sigma_1^2 I_i + \sigma_0^2(1-I_i)$. A rationalization for this specification is given in the introduction. If we define S_1 (S_0) as the set of individuals i with $I_i = 1$ (0), the loglikelihoodfunction of this model equals:

$$\log L = \sum_{i \epsilon S_1} \log(P_{i1}) + \sum_{i \epsilon S_0} \log(P_{i0}) \tag{4}$$

where:

$$P_{i1} = P(\varepsilon_i = y_i - X_i\alpha,\ \nu_i \le Z_i\gamma) = P(\nu_i \le Z_i\gamma \mid \varepsilon_i = y_i - X_i\alpha)\cdot P(\varepsilon_i = y_i - X_i\alpha)$$

$$= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma_1^2}(y_i - X_i\alpha)^2\right)\cdot\Phi\left[\frac{Z_i\gamma - \frac{\rho}{\sigma_1}(y_i - X_i\alpha)}{\sqrt{1-\rho^2}}\right]$$

and

$$P_{i0} = P(\varepsilon_i = y_i - X_i\alpha,\ \nu_i > Z_i\gamma) = P(\nu_i > Z_i\gamma \mid \varepsilon_i = y_i - X_i\alpha)\cdot P(\varepsilon_i = y_i - X_i\alpha)$$

---

[2] This model is discussed in Maddala (1983, p. 120).

[3] The variance of ν_i has to be restricted to 1 because we only observe the sign of I_i^*. The correlation is assumed to be homoscedastic and consequently the covariance of ε_i and ν_i is heteroscedastic.

$$= \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2\sigma_0^2}(y_i - X_i\alpha)^2)\cdot[1 - \Phi\left[\frac{Z_i\gamma - \frac{\rho}{\sigma_0}(y_i - X_i\alpha)}{\sqrt{1-\rho^2}}\right]]$$

To economize on the notation the loglikelihood function can be written as:

$$\log L = -\frac{N}{2}\log 2\pi - n_1\log\sigma_1 - n_0\log\sigma_0 - \frac{1}{2\sigma_1^2}\sum_{i\epsilon S_1}(y_i - X_i\alpha)^2 - \tag{5}$$

$$\frac{1}{2\sigma_0^2}\sum_{i\epsilon S_0}(y_i - X_i\alpha)^2 + \sum_{i\epsilon S_1}\log\Phi_i(\sigma_1) + \sum_{i\epsilon S_0}\log(1 - \Phi_i(\sigma_0)),$$

where:

$$\Phi_i(\sigma_*) = \Phi\left[\frac{Z_i\gamma-\rho\frac{1}{\sigma_*}(y_i - X_i\alpha)}{\sqrt{1-\rho^2}}\right] = \Phi\left[\frac{Z_i\gamma - \rho\varepsilon_{*i}}{\sqrt{1-\rho^2}}\right]$$

$$\varepsilon_{*i} = \frac{y_i - X_i\alpha}{\sigma_*}$$

N is the number of observations and n_1 (n_0) the number of elements of S_1 (S_0) ($N = n_0 + n_1$). The loglikelihoodfunction of the homoscedastic model ($\sigma = \sigma_1 = \sigma_2$) can be obtained by substituting σ for σ_1 and σ_2 in eq. (5). Maximization of (5) with respect to α, γ, σ_1, σ_0 and ρ will, under some general conditions (cf. Cramer, 1986, Chapter 2), yield consistent, asymptotically efficient estimates $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\sigma}_1$ $\hat{\sigma}_0$ and $\hat{\rho}$. The properties of the estimates obtained by ignoring the heteroscedasticity of the error term ($\tilde{\alpha}$, $\tilde{\gamma}$, $\tilde{\sigma}$ and $\tilde{\rho}$) are unknown and are the object of the present analysis.[4]

The ML-estimates of the model taking account of the heteroscedastic error terms can be calculated by solving the following system of equations:

---

[4] The analysis will concentrate on whether the most important parameters of the model ($\tilde{\alpha}$ and $\tilde{\gamma}$) are asymptotically biased or not.

$$\frac{\partial logL(\sigma_1,\sigma_0)}{\partial\gamma} = 0 = \frac{1}{\sqrt{1-\hat{\rho}^2}}\sum_{i\epsilon S_1}\hat{F}(\hat{\varepsilon}_{1i})Z_i - \frac{1}{\sqrt{1-\hat{\rho}^2}}\sum_{i\epsilon S_0}\hat{G}(\hat{\varepsilon}_{0i})Z_i \tag{6a}$$

$$\frac{\partial logL(\sigma_1,\sigma_0)}{\partial\alpha} = 0 = \frac{1}{\hat{\sigma}_1^2}\sum_{i\epsilon S_1}(y_i - X_i\hat{\alpha})X_i + \frac{1}{\hat{\sigma}_0^2}\sum_{i\epsilon S_0}(y_i - X_i\hat{\alpha})X_i + \frac{1}{\hat{\sigma}_1}\hat{\rho}^*\sum_{i\epsilon S_1}\hat{F}(\hat{\varepsilon}_{1i})X_i - \tag{b}$$

$$\frac{1}{\hat{\sigma}_0}\hat{\rho}^*\sum_{i\epsilon S_0}\hat{G}(\hat{\varepsilon}_{0i})X_i$$

$$\frac{\partial logL(\sigma_1,\sigma_0)}{\partial\sigma_1} = 0 = -\frac{n_1}{\hat{\sigma}_1} + \frac{1}{\hat{\sigma}_1^3}\sum_{i\epsilon S_1}(y_i - X_i\hat{\alpha})^2 + \frac{1}{\hat{\sigma}_1^2}\hat{\rho}^*\sum_{i\epsilon S_1}\hat{F}(\hat{\varepsilon}_{1i})(y_i - X_i\hat{\alpha}) \tag{c}$$

$$\frac{\partial logL(\sigma_1,\sigma_0)}{\partial\sigma_0} = 0 = -\frac{n_0}{\hat{\sigma}_0} + \frac{1}{\hat{\sigma}_0^3}\sum_{i\epsilon S_0}(y_i - X_i\hat{\alpha})^2 - \frac{1}{\hat{\sigma}_0^2}\hat{\rho}^*\sum_{i\epsilon S_0}\hat{G}(\hat{\varepsilon}_{0i})(y_i - X_i\hat{\alpha}) \tag{d}$$

$$\frac{\partial logL(\sigma_1,\sigma_0)}{\partial\rho} = 0 = \sum_{i\epsilon S_1}\hat{F}(\hat{\varepsilon}_{1i})\hat{\xi}_i(\hat{\sigma}_1) - \sum_{i\epsilon S_0}\hat{G}(\hat{\varepsilon}_{0i})\hat{\xi}_i(\hat{\sigma}_0) \tag{e}$$

where[5]

$$\hat{F}(\varepsilon_i) = \phi\left(\frac{Z_i\hat{\gamma} - \hat{\rho}\varepsilon_i}{\sqrt{1-\hat{\rho}^2}}\right)\cdot[\Phi\left(\frac{Z_i\hat{\gamma} - \hat{\rho}\varepsilon_i}{\sqrt{1-\hat{\rho}^2}}\right)]^{-1} = \frac{\phi(Z_i\hat{\gamma}^* - \hat{\rho}^*\varepsilon_i)}{\Phi(Z_i\hat{\gamma}^* - \hat{\rho}^*\varepsilon_i)}$$

$$\hat{G}(\varepsilon_i) = \phi\left(\frac{Z_i\hat{\gamma} - \hat{\rho}\varepsilon_i}{\sqrt{1-\hat{\rho}^2}}\right)\cdot[1 - \Phi\left(\frac{Z_i\hat{\gamma} - \hat{\rho}\varepsilon_i}{\sqrt{1-\hat{\rho}^2}}\right)]^{-1} = \frac{\phi(Z_i\hat{\gamma}^* - \hat{\rho}^*\varepsilon_i)}{1 - \Phi(Z_i\hat{\gamma}^* - \hat{\rho}^*\varepsilon_i)}$$

$$\hat{\xi}_i(\sigma_*) = \frac{1}{\sqrt{1-\hat{\rho}^2}}\left(\hat{\rho}\cdot\frac{Z_i\hat{\gamma} - \hat{\rho}\hat{\varepsilon}_{*i}}{\sqrt{1-\hat{\rho}^2}} - \hat{\varepsilon}_{*i}\right)$$

$$\hat{\rho}^* = \frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}}$$

---

[5] $\hat{\varepsilon}_{0i}$ ($\hat{\varepsilon}_{1i}$) denotes ε_{0i} (ε_{1i}) evaluated in $\hat{\alpha}$ and $\hat{\sigma}_0$ ($\hat{\sigma}_1$).

$$\hat{\gamma}^{*} = \frac{\hat{\gamma}}{\sqrt{1-\hat{\rho}^2}}$$

In general this system of equations will have to be solved numerically. Assuming a homoscedastic error term ε_i would require solving:[6]

$$\frac{\partial logL(\sigma)}{\partial\gamma} = 0 = \frac{1}{\sqrt{1-\tilde{\rho}^2}}\sum_{i\epsilon S_1}\tilde{F}(\tilde{\varepsilon}_{1i})Z_i - \frac{1}{\sqrt{1-\tilde{\rho}^2}}\sum_{i\epsilon S_0}\tilde{G}(\tilde{\varepsilon}_{0i})Z_i \tag{7a}$$

$$\frac{\partial logL(\sigma)}{\partial\alpha} = 0 = \frac{1}{\tilde{\sigma}^2}\sum_{i\epsilon S_1}(y_i - X_i\tilde{\alpha})X_i + \frac{1}{\tilde{\sigma}^2}\sum_{i\epsilon S_0}(y_i - X_i\tilde{\alpha})X_i + \frac{1}{\tilde{\sigma}}\tilde{\rho}^{*}\sum_{i\epsilon S_1}\tilde{F}(\tilde{\varepsilon}_{1i})X_i - \tag{b}$$

$$\frac{1}{\tilde{\sigma}}\tilde{\rho}^{*}\sum_{i\epsilon S_0}\tilde{G}(\tilde{\varepsilon}_{0i})X_i$$

$$\frac{\partial logL(\sigma)}{\partial\sigma} = 0 = -\frac{n_1}{\tilde{\sigma}} + \frac{1}{\tilde{\sigma}^3}\sum_{i\epsilon S_1}(y_i - X_i\tilde{\alpha})^2 + \frac{1}{\tilde{\sigma}^2}\tilde{\rho}^{*}\sum_{i\epsilon S_1}\tilde{F}(\tilde{\varepsilon}_{1i})(y_i - X_i\tilde{\alpha}) - \tag{c}$$

$$\frac{n_0}{\tilde{\sigma}} + \frac{1}{\tilde{\sigma}^3}\sum_{i\epsilon S_0}(y_i - X_i\tilde{\alpha})^2 - \frac{1}{\tilde{\sigma}^2}\tilde{\rho}^{*}\sum_{i\epsilon S_0}\tilde{G}(\tilde{\varepsilon}_{0i})(y_i - X_i\tilde{\alpha})$$

$$\frac{\partial logL(\sigma)}{\partial\rho} = 0 = \sum_{i\epsilon S_1}\tilde{F}(\tilde{\varepsilon}_{1i})\tilde{\xi}_i(\tilde{\sigma}) - \sum_{i\epsilon S_0}\tilde{G}(\tilde{\varepsilon}_{0i})\tilde{\xi}_i(\tilde{\sigma}) \tag{d}$$

From now on the analysis will concentrate on the estimation of the most important parameters of the model: α and γ. To prove that the probability limits of the estimators under the assumption of homoscedasticity ($\tilde{\alpha}$ and $\tilde{\gamma}$, these estimates result from solving (7)), equal α and γ, it suffices to demonstrate that the probability limits of (7a) and (7b) equal 0 under heteroscedasticity.

The following lemma is very useful in pursuing this objective.

---

[6] The tilde distinguishes the homoscedastic ML-estimation from the heteroscedastic ML-estimation. $\tilde{F}$, $\tilde{G}$ etc. are defined analogously to $\hat{F}$, $\hat{G}$ etc. except for substituting $\tilde{\alpha}$, $\tilde{\gamma}$ and $\tilde{\rho}$ for $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\rho}$ and $\tilde{\sigma}$ for $\hat{\sigma}_1$ and $\hat{\sigma}_0$.

Lemma 1.

For the function

$$G(\varepsilon) = \frac{\phi(Z\gamma^* - \rho^*\varepsilon)}{1 - \Phi(\gamma^* - \rho^*\varepsilon)}$$

the following holds:

(a) $\partial G(\varepsilon)/\partial\varepsilon < 0$ if $\rho > 0$

(b) $\partial G(\varepsilon)/\partial\varepsilon > 0$ if $\rho < 0$

(c) $\partial^2 G(\varepsilon)/\varepsilon^2 > 0$

The proof is given in the Appendix.

Starting with (7a), assume that for an infinite number of observations a proportion r_1 (= lim n_1/N) belong to the set S_1 and a proportion r_0 (= lim $n_0/N = 1 - r_1$) belong to S_0. Eq. (7a) can be written as:

$$\frac{n_1}{N}\frac{1}{n_1}\sum_{i\epsilon S_1}\tilde{F}(\tilde{\varepsilon}_{1i})Z_i - \frac{n_0}{N}\frac{1}{n_0}\sum_{i\epsilon S_0}\tilde{G}(\tilde{\varepsilon}_{0i})Z_i = 0 \tag{8}$$

For an infinite number of observations plim($\tilde{\gamma}$) solves:

$$\tilde{T} = r_1 \mathrm{plim}\left[\frac{1}{n_1}\sum_{i\epsilon S_1}\tilde{F}(\tilde{\varepsilon}_{1i})Z_i\right] - r_0 \mathrm{plim}\left[\frac{1}{n_0}\sum_{i\epsilon S_0}\tilde{G}(\tilde{\varepsilon}_{0i})Z_i\right] = 0 \tag{9}$$

where $\tilde{\varepsilon}_{1i} \sim N(0,(\sigma_1/\breve{\sigma})^2)$ and $\tilde{\varepsilon}_{0i} \sim N(0,(\sigma_0/\breve{\sigma})^2)$, $\sigma_1 \neq \sigma_0$. The essential feature of this specification is the deviation between the variance of both error terms. Therefore, we assume for simplicity that $\tilde{\varepsilon}_{1i} \sim N(0,1)$ and $\tilde{\varepsilon}_{0i} \sim N(0,\breve{\sigma}^2)$, $\breve{\sigma}^2 \neq 1$.[7] If the variance of the first and the second term (9) goes to 0, the probability limits can be replaced by expectations. The problem with eq. (9) is that it contains five different stochastic variables: $\tilde{\alpha}$, $\tilde{\gamma}$, $\tilde{\rho}$, $\breve{\sigma}$ and y_i. Assume that $\tilde{\alpha}$ and $\tilde{\rho}$ are consistent estimates of α and ρ. Furthermore, assume that plim $\breve{\sigma} = \sigma$ and plim $\tilde{\gamma} = \gamma(\sigma)$. We know that $\sigma \neq 1$. Define the function T as follows:

---

[7] The fact that $\breve{\sigma}^2 \neq 1$ is crucial. Obviously, if $\sigma_1 \neq \sigma_2$, the probability limit of $\breve{\sigma}$ can not be equal to 1 in this simplified version, because the probability limit of $\breve{\sigma}$ before this simplification was made can not be equal to σ_1 <u>and</u> σ_2.

$$T = r_1 \text{plim}\left[\frac{1}{n_1}\sum_{i\epsilon S_1} F(\varepsilon_{1i})Z_i\right] - r_0 \text{plim}\left[\frac{1}{n_0}\sum_{i\epsilon S_0} G(\varepsilon_{0i})Z_i\right] \tag{9'}$$

where:

$$F(\varepsilon_i) = \phi\left[\frac{Z_i\gamma(\sigma) - \rho\varepsilon_i}{\sqrt{1-\rho^2}}\right]\cdot\left[\Phi\left[\frac{Z_i\gamma(\sigma) - \rho\varepsilon_i}{\sqrt{1-\rho^2}}\right]\right]^{-1} = \frac{\phi(Z_i\gamma(\sigma)^* - \rho^*\varepsilon_i)}{\Phi(Z_i\gamma(\sigma)^* - \rho^*\varepsilon_i)}$$

$$G(\varepsilon_i) = \phi\left[\frac{Z_i\gamma(\sigma) - \rho\varepsilon_i}{\sqrt{1-\rho^2}}\right]\cdot\left[1 - \Phi\left[\frac{Z_i\gamma(\sigma) - \rho\varepsilon_i}{\sqrt{1-\rho^2}}\right]\right]^{-1} = \frac{\phi(Z_i\gamma(\sigma)^* - \rho^*\varepsilon_i)}{1 - \Phi(Z_i\gamma(\sigma)^* - \rho^*\varepsilon_i)}$$

$$\gamma(\sigma)^* = \frac{\gamma(\sigma)}{\sqrt{1-\rho^2}}$$

ε_{1i} and ε_{0i} are distinguished because they have distinct variances: $\varepsilon_{1i} \sim N(0,1)$ and $\varepsilon_{0i} \sim N(0,\sigma^2)$, $\sigma^2 \neq 1$. Clearly, plim $\tilde{T}$ = plim T = 0 in $\gamma(\sigma)$. Under the assumption that the error terms are uncorrelated across individuals we can write:

$$\text{var}\left[\frac{1}{n_1}\sum_{i\epsilon S_1} F(\varepsilon_{1i})Z_i\right] = \frac{1}{n_1^2}\sum_{i\epsilon S_1} \text{var}(F(\varepsilon_{1i})Z_i) = \frac{1}{n_1^2}\sum_{i\epsilon S_1} Z_iZ_i' \text{var}(F(\varepsilon_{1i})) \leq$$

$$\max_i\left[\frac{Z_iZ_i'}{n_1}\right]\sum_{i\epsilon S_1}\left[\frac{1}{n_1}E(F(\varepsilon_{1i})^2) - \left[\frac{E(F(\varepsilon_{1i}))}{\sqrt{n_1}}\right]^2\right]$$

Lemma 2.

$E(F(\varepsilon_{1i}))$ and $E(F(\varepsilon_{1i})^2)$ are finite.

The proof is given in the in the Appendix.

Lemma 3.

$E(G(\varepsilon_{0i})$ and $E(G(\varepsilon_{0i})^2)$ are finite.

Proof: Note that the Mills' ratio $\phi(x)/(1 - \Phi(x))$ equals $\phi(-x)/\Phi(-x)$ and employ Lemma 2. The fact that ε_{0i} has variance σ^2 ($\neq$ 1) does not cause complications because the density of ε_{0i} equals $\phi(\varepsilon_{0i}/\sigma)/\sigma$.

Assumption.

$$\frac{1}{n_1}\sum_{i\epsilon S_1} Z_i Z_i' \quad \text{and} \quad \frac{1}{n_0}\sum_{i\epsilon S_0} Z_i Z_i' \quad \text{are finite.}$$

This assumption implies that if the number of observations increases, the elements of the matrix $\Sigma Z_i Z_i'$ do not increase at a greater rate. This assumption is commonly made in the literature. Consequently $\max(Z_i Z_i'/n_1) < \infty$ and therefore, given Lemma's 1 and 2 the variance of the first and second term of (9') go to 0 if the number of observations is increased infinitely. So, $\gamma(\sigma)$ solves:

$$T = r_1 \left[\lim_{n_1\to\infty} E\left(\frac{1}{n_1}\sum_{i\epsilon S_1} Z_i F(\varepsilon_{1i})\right)\right] - r_0 \left[\lim_{n_0\to\infty} E\left(\frac{1}{n_0}\sum_{i\epsilon S_0} Z_i G(\varepsilon_{0i})\right)\right] = 0 \qquad (10)$$

where $\varepsilon_{1i} \sim N(0,1)$ and $\varepsilon_{0i} \sim N(0,\sigma^2)$. If $\sigma = 1$, (10) holds and $\text{plim}\tilde{\gamma} = \text{plim}\hat{\gamma} = \gamma$ (the heteroscedastic case (5)). For the consistency (inconsistency) of $\tilde{\gamma}$ we have to proof that $\gamma(\sigma)$ is independent (dependent) of σ if the number of observations is infinite or, in other words, $\partial\gamma(\sigma)/\partial\sigma = 0$ for all $\sigma > 0$ ($\partial\gamma(\sigma)/\partial\sigma \neq 0$ for all $\sigma > 0$). Consider the arbitrary element k of the vector T:

$$T_k = r_1 \left[\lim_{n_1\to\infty} E\left(\frac{1}{n_1}\sum_{i\epsilon S_1} Z_{ik} F(\varepsilon_{1i})\right)\right] - r_0 \left[\lim_{n_0\to\infty} E\left(\frac{1}{n_0}\sum_{i\epsilon S_0} Z_{ik} G(\varepsilon_{0i})\right)\right] = 0 \qquad (10')$$

Apply the implicit function theorem to (10'):

$$\frac{\partial\gamma_k(\sigma)}{\partial\sigma} = -\frac{\partial T_k/\partial\sigma}{\partial T_k/\partial\gamma_k(\sigma)} \quad \text{if} \quad \partial T_k/\partial\gamma_k(\sigma) \neq 0$$

where:

$$\frac{\partial T_k}{\partial\sigma} = -r_0 \lim_{n_0\to\infty} \frac{1}{n_0}\sum_{i\epsilon S_0} Z_{ik}\frac{\partial E(G(\varepsilon_{0i}))}{\partial\sigma} = -r_0 \lim_{n_0\to\infty} \frac{1}{n_0}\sum_{i\epsilon S_1} Z_{ik}\int_{-\infty}^{\infty} G(\varepsilon_{0i})\frac{\partial f(\varepsilon_{0i})}{\partial\sigma}\, d\varepsilon_{0i} =$$

$$\frac{r_0}{\sigma} \lim_{n_0\to\infty} \frac{1}{n_0}\sum_{i\epsilon S_0} Z_{ik} \int_{-\infty}^{\infty} G(\varepsilon_{0i}) f(\varepsilon_{0i}) \left[1-\frac{\varepsilon_{0i}^2}{\sigma^2}\right] d\varepsilon_{0i} = \frac{r_0}{\sigma} \lim_{n_0\to\infty} \frac{1}{n_0}\sum_{i\epsilon S_0} Z_{ik}\theta_{\sigma i}$$

and

$$\frac{\partial T_k}{\partial \gamma_k(\sigma)} = r_1 \lim_{n_1\to\infty} \frac{1}{n_1}\sum_{i\epsilon S_1} Z_{ik} \int_{-\infty}^{\infty} \frac{\partial F(\varepsilon_{1i})}{\partial \gamma_k(\sigma)} \phi(\varepsilon_{1i})\, d\varepsilon_{1i} - r_0 \lim_{n_0\to\infty} \frac{1}{n_0}\sum_{i\epsilon S_0} Z_{ik} \int_{-\infty}^{\infty} \frac{\partial G(\varepsilon_{0i})}{\partial \gamma_k(\sigma)} f(\varepsilon_{0i})\, d\varepsilon_{0i} =$$

$$r_1 \lim_{n_1\to\infty} \frac{1}{n_1}\sum_{i\epsilon S_1} Z_{ik}\theta^k_{1\gamma i} - r_0 \lim_{n_0\to\infty} \frac{1}{n_0}\sum_{i\epsilon S_0} Z_{ik}\theta^k_{0\gamma i}$$

$f(\varepsilon_{0i})$ is the density function of a normal distribution with expectation 0 and variance σ^2.

Proposition 1.

If there exists at least one $i \in S_0$ or at least one $i \in S_1$ for which $Z_{ik} \neq 0$, $\partial T_k/\partial\gamma_k(\sigma) < 0$ for all $\gamma_k(\sigma)$ and σ.

The proof is given in the Appendix.

Proposition 2.

If there exists at least one $i \in S_0$ such that $Z_{ik} \neq 0$ and if $\rho \neq 0$, $\theta_{\sigma i} < 0$ for all σ.

The proof is given in the Appendix.

Now assume that there exists an $i \in S_0$ such that $Z_i \neq 0$. From propositions 1 and 2 it follows that:

$$\frac{\partial \gamma_k(\sigma)}{\partial \sigma} \neq 0 \quad \text{if } \rho \neq 0$$

In other words, the homoscedastic ML-estimator $\tilde{\gamma}_k$ of the heteroscedastic model ($\sigma \neq 1$) is inconsistent under the assumption that α and ρ are consistently estimated. According to Proposition 2 it does not matter whether $\tilde{\rho}$ is consistent as long as $\tilde{\rho} \neq 0$. Therefore the assumption that $\tilde{\rho}$ is estimated consistently is superfluous. If $\tilde{\rho} = 0$ (or if it is simply assumed that $\rho = 0$), γ_1 is estimated consistently. Of course, this is in line with the frequently utilized 2-step estimation technique for selection models. For the present model the 2-step estimation of first estimating eq. (3b) by means of the probit technique and then estimating (3a) with ordinary least squares on the regressors X_i and a correction term created from the first step of the estimation (cf. Heckman, 1979) will yield a consistent estimate of γ despite the heteroscedasticity of the error term. However, in this case the 2-step estimate of α will be inconsistent (see further). We can conclude that $\tilde{\gamma}$ is an inconsistent estimator of γ if it is assumed that $\tilde{\alpha}$ is a consistent estimator of α.

Inconsistent estimation of α might in principle yield the consistency of $\tilde{\gamma}$. Suppose that $\text{plim}(\tilde{\alpha}) = \alpha + \mu(\sigma)$, where $\mu(\sigma) \neq 0$. Given Proposition 2, μ needs to be dependent on σ to establish the consistency of $\tilde{\gamma}$. $\mu(\sigma)$ solves:[8]

$$\frac{\partial T}{\partial \sigma} = 0 = r_1 \lim_{n_1 \to \infty} \sum_{i \in S_1} Z_i \frac{\partial E(F(\varepsilon_{1i}))}{\partial \sigma} \Big|_\gamma - r_0 \lim_{n_0 \to \infty} \sum_{i \in S_0} Z_i \frac{\partial E(G(\varepsilon_{0i}))}{\partial \sigma} \Big|_\gamma \tag{11}$$

Furthermore, $\mu(\sigma)$ should be such that $\tilde{\alpha} = \alpha + \mu(\sigma)$ solves (7b). Whether $\tilde{\alpha}$ is able to solve both of these equations simultaneously is a difficult problem. It becomes even more involved if we recognize that $\tilde{\rho}$ is likely to depend on σ also. However, it can be shown for a relevant case that $\tilde{\alpha}$ is not able to solve both (11) and (7b) simultaneously. Assume that $Z_i = X_i$. In that case $\tilde{\alpha}$ solves:

$$\sum_{i \in S_1} (y_i - X_i\tilde{\alpha})X_i + \sum_{i \in S_0} (y_i - X_i\tilde{\alpha})X_i = 0$$

that is, $\tilde{\alpha}$ is the ordinary least squares estimate of equation (3a).[9] Clearly, $\text{plim}(\tilde{\alpha})$ is independent of σ and therefore, it can not solve (11) for every σ. This conclusion can be made more general: construct the vector W_i by stacking the independent elements of X_i and Z_i.[10] Replace X_i and Z_i in (3) by W_i. The elements of α and γ corresponding to the elements of W_i not in X_i or in Z_i respectively should be estimated equal to 0. This artificial procedure does not prohibit us to conclude that $\tilde{\gamma}$ is an inconsistent estimate of γ irrespective of the consistency of

[8] In this case the first term of (10) also depends on σ through $\tilde{\alpha}$ which is part of ε_{1i}. $|_\gamma$ denotes that the expression is evaluated at γ.

[9] Note that this conclusion does not hold for the heteroscedastic model (6). The weights of the third and fourth elements of (6b) differ.

[10] At this point it is assumed for the moment that I_i is not an element of X_i.

both $\tilde{\alpha}$ and $\tilde{\rho}$. Unfortunately, this conclusion does not hold for the case that I_i is part of X_i: eq. (3b) can not be estimated if I_i is an element of Z_i. A Monte Carlo study will be performed to study this point (see section 3). Given the complicated structure of (11) and (7b) it appears to be unlikely that a $\mu(\sigma)$ exists with settles both these equations simultaneously.

Let us now turn to the estimation of α. First consider a 2-step estimation of model (3) while ignoring heteroscedasticity. In the second step the following equation is estimated with ordinary least squares:[11]

$$y_i = X_i\alpha + \rho\sigma\lambda_i + v_i$$

where:

$$\lambda_i = E(\varepsilon_i | I_i = 1) = E(\varepsilon_i | \nu_i \leq Z_i\gamma) \quad \text{if } I_i = 1$$

or

$$\lambda_i = E(\varepsilon_i | I_i = 0) = E(\varepsilon_i | Z_i\gamma > 0) \quad \text{if } I_i = 0$$

Consistent estimates of λ_i can be calculated from the first step, the probit estimation of (3b):

$$\lambda_i = -\frac{\phi(Z_i\mathring{\gamma})}{\Phi(Z_i\mathring{\gamma})} \quad \text{if } I_i = 1 \quad \text{and} \quad \lambda_i = \frac{\phi(Z_i\mathring{\gamma})}{1 - \Phi(Z_i\mathring{\gamma})} \quad \text{if } I_i = 0$$

where $\mathring{\gamma}$ is the probit estimate of γ. Ignoring the heteroscedasticity of v_i and assuming that it is normally distributed with mean 0 and variance σ_v^2, the OLS regression amounts to maximizing the following loglikelihoodfunction:

$$\log L = \frac{N}{2}\log(2\pi) - N\log\sigma_v - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - X_i\alpha - \delta\lambda_i}{\sigma_v}\right)^2$$

One of the first order conditions is can be written as:

[11] cf. Maddala, 1983, p. 120 or p. 224.

$$\frac{\partial logL}{\partial \alpha} = 0 = \frac{1}{\dot{\sigma}_{\upsilon}^{2}}\left[\sum_{i \epsilon S_1}(y_i - X_i\dot{\alpha})X_i + \sum_{i \epsilon S_0}(y_i - X_i\dot{\alpha})X_i + \dot{\delta}\sum_{i \epsilon S_1}(-\dot{\lambda}_i X_i) - \dot{\delta}\sum_{i \epsilon S_0}\dot{\lambda}_i X_i\right] \tag{12}$$

This method of estimation suffers from two different types of heteroscedasticity. As already noted υ_i is heteroscedastic. Due to the well known result that ignoring heteroscedasticity in a linear regression model does not hinder consistent estimation, this heteroscedasticity problem per se will not lead to an inconsistent estimate of α. However the second type does. The problem is that the coefficient of the correction term λ_i for $I_i = 1$ and $I_i = 0$ are restricted to be equal despite the heteroscedasticity of ε_i. The correct coefficients of the corrections terms are $\rho\sigma_1$ ($I_i = 1$) and $\rho\sigma_0$ ($I_i = 0$), which are only equal if $\sigma_1 = \sigma_0$ or $\rho = 0$. Consequently, this restriction prevents the consistent estimation of α by solving (12). Consistent estimates can be obtained in this case by not imposing this restriction in the 2-step estimation technique (see below). Assume for the moment that $X_i = Z_i$. Given the similarity of (12) and (7b) it will be clear that substituting consistent estimates of $F(\varepsilon_{1i})/\sqrt{(1-\rho^2)}$ and $G(\varepsilon_{0i})/\sqrt{(1-\rho^2)}$, assuming that there are available[12], and solving (7b) will not yield a consistent estimate of α. A more general conclusion can be reached by considering model (3) where W_i, the vector of independent elements of X_i and Z_i, is substituted for X_i and Z_i.[13] The third and fourth element of (7b) drop out. Assume W_i is arranged such that $W_i = (X_i, \dot{X})$, where $\dot{X}$ are the independent elements of Z_i not belonging to X_i. Denote the corresponding vector of parameters by: $\alpha^{*\prime} = (\tilde{\alpha}', \dot{\alpha}')$. α^* solves:[14]

$$\sum_{i \epsilon S_1}(y_i - W_i\alpha^*)W_i + \sum_{i \epsilon S_0}(y_i - W_i\alpha^*)W_i = 0$$

Therefore, the homoscedastic estimation is similar to regressing W_i on y_i.

Consider the ML-estimation of (3) while taking account of second type of the

---

[12] Consistent estimates of F(.) and G(.) can be obtained by carrying out a probit estimation on I_i while using both Z_i, y_i and X_i as regressors. $(\sqrt{(1-\rho^2)})^{-1}$ can be included in the coefficients of the second stage of the estimation (cf. footnote 15).

[13] Again, I_i is for the moment excluded from X_i.

[14] Of course, $\dot{\alpha}$ can be equal to 0.

heteroscedasticity discussed above. In this case the first order condition with respect to α^* is :[15]

$$\frac{\partial logL}{\partial \alpha^*} = 0 = \frac{1}{\sigma_1^2}\sum_{i\epsilon S_1}(y_i - W_i\alpha^*)W_i + \frac{1}{\sigma_0^2}\sum_{i\epsilon S_0}(y_i - W_i\alpha^*)W_i + \tag{13}$$

$$\delta_1\frac{1}{\sigma_1^2}\sum_{i\epsilon S_1}(-\lambda_i)W_i - \delta_0\frac{1}{\sigma_0^2}\sum_{i\epsilon S_0}\lambda_i W_i$$

The first order condition with respect to γ equals:

$$\frac{\partial logL}{\partial \gamma} = 0 = \sum_{i\epsilon S_1}F(\varepsilon_{1i})W_i - \sum_{i\epsilon S_0}F(\varepsilon_{0i})W_i = \sum_{i\epsilon S_1}(-\lambda_i)W_i - \sum_{i\epsilon S_0}\lambda_i W_i$$

Substituting this condition in (13) yields:

$$\frac{1}{\sigma_1^2}\sum_{i\epsilon S_1}(y_i - W_i\alpha^* - (\delta_1 - \frac{\sigma_1^2}{\sigma_0^2}\delta_0)\lambda_i)W_i + \frac{1}{\sigma_0^2}\sum_{i\epsilon S_0}(y_1 - W_i\alpha^*)W_i = 0$$

This first order condition corresponds to the regression model:

$$y_i = W_i\alpha^* + (\delta_1 - \delta_0)\lambda_i I_i + \omega_i \tag{14}$$

Recall at this point that we have already demonstrated that the homoscedastic maximum likelihood estimate of α of model (3) in which X_i and Z_i are replaced by W_i, is the ordinary least squares estimate of equation (3a) in which X_i is replaced by W_i. Consequently, if $\rho \neq 0$ and $\sigma_1 \neq \sigma_0$, simply estimating equation (3a), where y_i is regressed on W_i instead of X_i, will not lead to a consistent estimate of α due to missing variables bias. Only if W_i and λ_i are uncorrelated, consistent estimates are obtained by employing ordinary least squares. This is obviously not the case.

Finally, consider the case in which I_i is part of equation (3a). For simplicity look at the following model:

---

[15] $\delta_1 = \rho\sigma_1/\sqrt{(1-\rho^2)}$ and $\delta_0 = \rho\sigma_0/\sqrt{(1-\rho^2)}$.

$$y_i = \theta I_i + \varepsilon_i \quad \text{where } \varepsilon_i \sim N(0,1) \text{ if } I_i = 1 \text{ and } \varepsilon_i \sim N(0,\sigma^2),\ \sigma^2 \neq 1 \text{ if } I_i = 0$$

$$I_i^* = Z_i\gamma - \nu_i$$

Rewriting the first order conditions with respect to estimating θ for the homoscedastic and heteroscedastic maximum likelihood estimations yields (cf. eqs. (7b) and (6b)):

$$\tilde{\theta} = \frac{1}{n_1}\sum_{i\epsilon S_1} y_i + \tilde{\sigma}\tilde{\rho}\cdot\frac{1}{n_1}\sum_{i\epsilon S_1} F(\tilde{\varepsilon}_i) = \theta + \frac{1}{n_1}\sum_{i\epsilon S_1}\varepsilon_i + \tilde{\sigma}\tilde{\rho}\cdot\frac{1}{n_1}\sum_{i\epsilon S_1} F(\tilde{\varepsilon}_i)$$

$$\hat{\theta} = \frac{1}{n_1}\sum_{i\epsilon S_1} y_i + \hat{\sigma}\hat{\rho}\cdot\frac{1}{n_1}\sum_{i\epsilon S_1} F(\hat{\varepsilon}_i) = \theta + \frac{1}{n_1}\sum_{i\epsilon S_1}\varepsilon_i + \hat{\sigma}\hat{\rho}\cdot\frac{1}{n_1}\sum_{i\epsilon S_1} F(\hat{\varepsilon}_i)$$

The probability limit of $\hat{\theta}$ is θ. Consequently, $\tilde{\theta}$ is consistent only if:

$$\text{plim}\left[\tilde{\sigma}\tilde{\rho}\cdot\frac{1}{n_1}\sum_{i\epsilon S_1} F(\tilde{\varepsilon}_i)\right] = -\text{plim}\left[\frac{1}{n_1}\sum_{i\epsilon S_1}\varepsilon_i\right] = \rho\cdot\text{plim}\left[\frac{1}{n_1}\sum_{i\epsilon S_1} F(\hat{\varepsilon}_i)\right] \tag{14}$$

where $\varepsilon_i \sim N(0,1)$, $\hat{\varepsilon} \sim N(0,1)$ and $\tilde{\varepsilon} \sim N(0,\tilde{\sigma}^2)$. The inconsistency of θ can be proved by showing that the first term of (14) is independent of σ. However, this is a very complex task because both $\tilde{\sigma}$ and $\tilde{\rho}$ depend on σ. Furthermore, no explicit form of $\tilde{\sigma}$ and $\tilde{\rho}$ can be derived, we only know that they solve (7c) and (7d). To get some insight in the consistency of θ a Monte Carlo experiment will be carried out in the section 4.

3. Heteroscedasticity in the Heckman (1979) selectivity model.

Consider the model discussed in Heckman (1979):

$$y_i = X_i\alpha + \varepsilon_i \quad \text{observed if } I_i = 1 \tag{15a}$$

$$I_i^* = Z_i\gamma - \nu_i \tag{b}$$

where $I_i = 1$ if $I_i^* > 0$ and $I_i = 0$ if $I_i^* \leq 0$. Assume that there n_0 observations for which y_i is not observed and $N\text{-}n_0$ observations for which y_i is observed. Furthermore, for the observations for which y_i is observed the variance of ε_i is either σ_1^2 (n_1 observations) or σ_2^2 (n_2 observations) and $\sigma_1^2 \neq \sigma_2^2$. The group of n_1 (n_2) observations will be denoted by S_1 (S_2). The main goal of model (15) is the estimation of α, the addition of eq. (15b) to the model is only carried out to correct for selectivity. The loglikelihoodfunction of model (15) is:

$$\log L = \sum_{i \epsilon S_1} \log P(\varepsilon_{1i} = y_i - X_i\alpha,\ I_i^* > 0) + \sum_{i \epsilon S_2} \log P(\ \varepsilon_{2i} = y_i - X_i\alpha,\ I_i^* > 0) +$$

$$\sum_{i \epsilon S_0} \log P(I_i^* \leq 0)$$

where the additional subscript of ε_{ji} reflects the heteroscedasticity of this error term. Under the assumption of normally distributed error terms with expectation 0, $\text{var}(\varepsilon_{1i}) = \sigma_1^2$, $\text{var}(\varepsilon_{2i}) = \sigma_2^2$, $\text{var}(\nu_i) = 1$, $\text{cov}(\varepsilon_{1i},\nu_i) = \rho\sigma_1$ and $\text{cov}(\sigma_{2i},\nu_i) = \rho\sigma_2$, the loglikihoodfunction of model (15) can be written as:

$$\log L = -\frac{n_1+n_2}{2}\log 2\pi - n_1\log\sigma_1 - n_2\log\sigma_2 - 0.5\sum_{i \epsilon S_1}(\varepsilon_{1i}(\sigma_1))^2 - 0.5\sum_{i \epsilon S_2}(\varepsilon_{2i}(\sigma_2))^2 +$$

$$\sum_{i \epsilon S_1} \log\Phi\left(\frac{Z_i\gamma - \rho\varepsilon_{1i}(\sigma_1)}{\sqrt{1-\rho^2}}\right) + \sum_{i \epsilon S_2} \log\Phi\left(\frac{Z_i\gamma - \rho\varepsilon_{2i}(\sigma_2)}{\sqrt{1-\rho^2}}\right) + \sum_{i \epsilon S_0} \log\left(1 - \Phi(Z_i\gamma)\right)$$

where

$$\varepsilon_{1i}(\sigma_1) = \frac{y_i - X_i\alpha}{\sigma_1} \quad \text{and} \quad \varepsilon_{2i}(\sigma_2) = \frac{y_i - X_i\alpha}{\sigma_2} \qquad (16)$$

The loglikihoodfunction in the case that the heteroscedasticity of ε_i is ignored is obtained by substituting σ for σ_1 and σ_2 in (16). Maximizing this likelihoodfunction is equivalent to solving the system of first order derivatives with respect to the parameters of the model:

$$\frac{\partial \log L(\sigma)}{\partial \gamma} = 0 = \frac{1}{\sqrt{1-\rho^2}}\left[\sum_{i \epsilon S_1} F(\varepsilon_{1i}(\sigma))Z_i + \sum_{i \epsilon S_2} F(\varepsilon_{2i}(\sigma))Z_i\right] - \sum_{i \epsilon S_0} G(0)Z_i \qquad (17a)$$

$$\frac{\partial \log L(\sigma)}{\partial \alpha} = 0 = \frac{1}{\sigma}\sum_{i \epsilon S_1} \bar{\varepsilon}_{1i}(\sigma)X_i + \frac{1}{\sigma}\sum_{i \epsilon S_2} \bar{\varepsilon}_{2i}(\sigma)X_i + \frac{\rho^*}{\sigma}\sum_{i \epsilon S_1} F(\bar{\varepsilon}_{1i}(\sigma))X_i + \qquad (b)$$

$$\frac{\rho^*}{\sigma}\sum_{i\epsilon S_2} F(\bar{\varepsilon}_{2i}(\sigma))X_i$$

$$\frac{\partial \log L(\sigma)}{\partial \sigma} = 0 = -\frac{n_1+n_2}{\sigma} + \frac{1}{2\sigma}\sum_{i\epsilon S_1}(\bar{\varepsilon}_{1i}(\sigma))^2 + \frac{1}{2\sigma}\sum_{i\epsilon S_2}(\bar{\varepsilon}_{2i}(\sigma))^2 + \qquad \text{(c)}$$

$$\frac{\rho^*}{\sigma}\sum_{i\epsilon S_1} F(\bar{\varepsilon}_{1i}(\sigma))\bar{\varepsilon}_{1i}(\sigma) + \frac{\rho^*}{\sigma}\sum_{i\epsilon S_2} F(\bar{\varepsilon}_{2i}(\sigma))\bar{\varepsilon}_{2i}(\sigma)$$

$$\frac{\partial \log L(\sigma)}{\partial \rho} = 0 = \frac{1}{\sqrt{1-\rho^2}}\left[\sum_{i\epsilon S_1} F(\bar{\varepsilon}_{1i}(\sigma))\xi_i(\sigma) - \sum_{i\epsilon S_2} F(\bar{\varepsilon}_{2i}(\sigma))\xi_i(\sigma)\right] \qquad \text{(d)}$$

where

$$F(\varepsilon_i) = \phi\left[\frac{Z_i\gamma - \rho\varepsilon_i}{\sqrt{1-\rho^2}}\right]\cdot[\Phi\left[\frac{Z_i\gamma - \rho\varepsilon_i}{\sqrt{1-\rho^2}}\right]]^{-1} = \frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_i)}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_i)}$$

$$G(0) = \frac{\phi(Z_i\gamma)}{1 - \Phi(Z_i\gamma)}$$

$$\xi_i(\sigma_*) = \frac{1}{\sqrt{1-\rho^2}}\left[\rho\cdot\frac{Z_i\gamma - \rho\varepsilon_{*i}}{\sqrt{1-\rho^2}} - \varepsilon_{*i}\right]$$

$$\rho^* = \frac{\rho}{\sqrt{1-\rho^2}}$$

$$\gamma^* = \frac{\gamma}{\sqrt{1-\rho^2}}$$

Again the analysis will be concentrated on the consistency of the estimators of the most relevant parameters of the model: α and γ. To prove the consistency of the estimates resulting from solving (17) we need to prove that (17) also holds, while acknowledging the heteroscedasticity of the error term, if the number of observations goes to infinity. Denote the solutions of (17) by: $\bar{\gamma}$, $\bar{\alpha}$, $\bar{\sigma}$ and $\bar{\rho}$. Starting with (17a) assume that if the number of observations gets very large, the proportions of the observations in each set S_1, S_2 and S_0 got to the constant r_1, r_2 and r_0. Eq.

(17a) can be written as:

$$\frac{1}{\sqrt{1-\bar{\rho}^2}}\left[\frac{n_1}{N}\frac{1}{n_1}\sum_{i\epsilon S_1}\bar{F}(\bar{\varepsilon}_{1i}(\bar{\sigma}))Z_i + \frac{n_2}{N}\frac{1}{n_2}\sum_{i\epsilon S_2}\bar{F}(\bar{\varepsilon}_{2i}(\bar{\sigma}))Z_i\right] - \frac{n_0}{N}\frac{1}{n_0}\sum_{i\epsilon S_0}\bar{G}(0)Z_i = 0 \tag{18}$$

where the bars indicate that the function is evaluated in $\bar{\gamma}$, $\bar{\alpha}$, $\bar{\sigma}$ and $\bar{\rho}$. If the number of observations becomes boundless we know that the probability limits of $\bar{\gamma}$, $\bar{\alpha}$, $\bar{\sigma}$ and $\bar{\rho}$ solve:

$$\bar{T} = \text{plim}\left[\frac{1}{\sqrt{1-\bar{\rho}^2}}\right]\left[r_1\text{plim}\left[\frac{1}{n_1}\sum_{i\epsilon S_1}\bar{F}(\bar{\varepsilon}_{1i}(\bar{\sigma}))Z_i\right] + r_2\text{plim}\left[\frac{1}{n_2}\sum_{i\epsilon S_2}\bar{F}(\bar{\varepsilon}_{2i}(\bar{\sigma}))Z_i\right]\right] - \tag{19}$$

$$r_0\text{plim}\left[\frac{1}{n_0}\sum_{i\epsilon S_0}\bar{G}(0)Z_i\right] = 0$$

where $\bar{\varepsilon}_{1i} \sim N(0,(\sigma_1/\bar{\sigma})^2)$ and $\bar{\varepsilon}_{2i} \sim N(0,(\sigma_2/\bar{\sigma})^2)$. The essential characteristics of the specification is that $\sigma_1 \neq \sigma_2$. Again, to simplify things, we will assume again that $\bar{\varepsilon}_{1i} \sim N(0,1)$ and $\bar{\varepsilon}_{2i} \sim N(0,\breve{\sigma}^2)$, $\breve{\sigma}^2 \neq 1$. T consists of five different random variables: $\bar{\gamma}$, $\bar{\alpha}$, $\bar{\sigma}$, $\bar{\rho}$ and y_i. Assume that $\bar{\alpha}$ and $\bar{\rho}$ are consistent estimates of α and ρ and that plim $\breve{\sigma} = \sigma \neq 1$ and plim $\bar{\gamma} = \gamma(\sigma)$. Define the function T as follows:

$$T = \frac{1}{\sqrt{1-\rho^2}}\left[r_1\text{plim}\left[\frac{1}{n_1}\sum_{i\epsilon S_1}F(\varepsilon_{1i})Z_i\right] + r_2\text{plim}\left[\frac{1}{n_2}\sum_{i\epsilon S_2}F(\varepsilon_{2i})Z_i\right]\right] - \tag{19'}$$

$$r_0\text{plim}\left[\frac{1}{n_0}\sum_{i\epsilon S_0}G(0)Z_i\right]$$

where F(.) and G(.) are defined as in (9'), $\varepsilon_{1i} \sim N(0,1)$ and $\varepsilon_{2i} \sim N(0,\sigma^2)$, $\sigma^2 \neq 1$. Obviously, plim T = plim T = 0 in $\gamma(\sigma)$. Analogously to the previous section and under the assumption that the elements of the matrix $\Sigma Z_i Z_i'$ increase in a moderate enough fashion (cf. the Assumption in section 2) it can be proved that $\gamma(\sigma)$ solves:

$$T = \frac{1}{\sqrt{1-\rho^2}}\left[r_1 \lim_{n_1\to\infty} E\left[\frac{1}{n_1}\sum_{i\epsilon S_1} F(\varepsilon_{1i})Z_i\right] + r_2 \lim_{n_2\to\infty} E\left[\frac{1}{n_2}\sum_{i\epsilon S_2} F(\varepsilon_{2i})Z_i\right]\right] - \tag{20}$$

$$r_0 \lim_{n_0\to\infty}\left[\frac{1}{n_0}\sum_{i\epsilon S_0} G(0)Z_i\right] = 0$$

For the consistency (inconsistency) of $\bar{\gamma}$ we have to prove that $\gamma(\sigma)$ is independent (dependent) of σ if the number of observations is infinite. Apply the implicit function theorem to an arbitrary element k of the vector T:

$$\frac{\partial\gamma_k(\sigma)}{\partial\sigma} = -\frac{\partial T_k/\partial\sigma}{\partial T_k/\partial\gamma_k(\sigma)} \quad \text{if} \quad \partial T_k/\partial\gamma_k(\sigma)\neq 0$$

where:

$$\frac{\partial T_k}{\partial\sigma} = r_2 \lim_{n_2\to\infty}\frac{1}{n_2}\sum_{i\epsilon S_2} Z_{ik}\frac{\partial E(F(\varepsilon_{2i}))}{\partial\sigma} = r_2 \lim_{n_2\to\infty}\frac{1}{n_2}\sum_{i\epsilon S_2} Z_{ik}\int_{-\infty}^{\infty} F(\varepsilon_{2i})\frac{\partial f(\varepsilon_{2i})}{\partial\sigma}\, d\varepsilon_{2i} =$$

$$-\frac{r_2}{\sigma} \lim_{n_2\to\infty}\frac{1}{n_2}\sum_{i\epsilon S_2} Z_{ik}\int_{-\infty}^{\infty} F(\varepsilon_{2i})f(\varepsilon_{2i})\left[1-\frac{\varepsilon_{2i}^2}{\sigma^2}\right] d\varepsilon_{2i} = -\frac{r_2}{\sigma} \lim_{n_2\to\infty}\frac{1}{n_2}\sum_{i\epsilon S_2} Z_{ik}\psi_{\sigma i}^{*}$$

and

$$\frac{\partial T_k}{\partial\gamma_k(\sigma)} = r_1 \lim_{n_1\to\infty}\frac{1}{n_1}\sum_{i\epsilon S_1} Z_{ik}\int_{-\infty}^{\infty}\frac{\partial F(\varepsilon_{1i})}{\partial\gamma_k(\sigma)}\phi(\varepsilon_{1i})\, d\varepsilon_{1i} + r_2 \lim_{n_2\to\infty}\frac{1}{n_2}\sum_{i\epsilon S_2} Z_{ik}\int_{-\infty}^{\infty}\frac{\partial F(\varepsilon_{2i})}{\partial\gamma_k(\sigma)}f(\varepsilon_{2i})\, d\varepsilon_{2i} -$$

$$r_0 \lim_{n_0\to\infty}\frac{1}{n_0}\sum_{i\epsilon S_0} Z_{ik}\frac{\partial G(0)}{\partial\gamma_k(\sigma)}$$

$$= r_1 \lim_{n_1\to\infty}\frac{1}{n_1}\sum_{i\epsilon S_1} Z_{ik}\psi_{1\gamma i}^{k} + r_2 \lim_{n_2\to\infty}\sum_{i\epsilon S_2} Z_{ik}\psi_{2\gamma i}^{k} - r_0 \lim_{n_0\to\infty}\sum_{i\epsilon S_0} Z_{ik}\psi_{0\gamma i}^{k}$$

$f(\varepsilon_{2i})$ is the density function of a normal distribution with expectation 0 and variance σ^2.

Proposition 3.

If there exists at least one $i \in S_0$, at least one $i \in S_1$ or at least one $i \in S_2$ for which $Z_{ik} \neq 0$, $\partial T_k/\partial\gamma_k(\sigma) < 0$ for all $\gamma_k(\sigma)$ and σ.

Proof:

a). $\psi^k_{1\gamma i} = \Xi_{ik}Z_{ik}$ with $\Xi_{ik} < 0$, due to Proposition 1.

b). $\psi^k_{2\gamma i} = \Omega_{ik}Z_{ik}$ with $\Omega_{ik} > 0$, due to Proposition 1 and the property of the inverse Mills' ratio: $\phi(x)/\Phi(x) = \phi(-x)/(1-\Phi(-x))$.

c). $\psi^k_{0\gamma i} = \partial G(0)/\partial\gamma_k(\sigma) = \partial G(0)/\partial(Z_{ik}\gamma_k(\sigma))Z_{ik} = \Psi_{ik}Z_{ik}$ with $\Psi_{ik} > 0$ (see the proof of Lemma 1) .

Combining a), b) and c) leads to the specified result.

Proposition 4.

If there exists at least one $i \in S_2$ such that $Z_{ik} \neq 0$ and if $\rho \neq 0$, $\psi_{\sigma i} > 0$ for all σ.

Proof: Due to the property of the inverse Mills' ratio that $\phi(x)/\Phi(x) = \phi(-x)/(1-\Phi(-x))$, we can make use of Proportion 2.

Assume that there exist an $i \in S_2$ such that Z_{ik} is not equal to 0. It follows from Propositions 3 and 4 that:

$$\frac{\partial\gamma_k(\sigma)}{\partial\sigma} \neq 0 \quad \text{if } \rho\neq 0$$

So, under the assumption that both $\bar{\alpha}$ and $\bar{\rho}$ are consistent, $\bar{\gamma}$ is inconsistent if $\rho \neq 0$. Again, in order to obtain this result it does not matter whether $\bar{\rho}$ is consistent as long as $\bar{\rho} \neq 0$. Consequently, the assumption that $\bar{\rho}$ is a consistent estimate of ρ is dispensable. Note however, that if $\bar{\rho}$ is inconsistent it is likely to be related to σ and therefore, the direction of the bias of $\bar{\gamma}$ which can be deduced from Proposition 4 need not to be correct. Furthermore, note that if we would have started with the incorrect assumption that $\rho = 0$, consistent estimates of γ would have been obtained. Of course, this is no surprise given the properties of the 2-step estimation technique discussed in Heckman (1979).

To investigate the consistency of $\bar{\alpha}$ consider the 2-step estimation of model (15) while ignoring the heteroscedasticity of the error term ε_i. Just like in the previous section this amounts to imposing an incorrect restriction on the coefficients of the second step equation. While ignoring heteroscedasticity the following model is estimated with OLS:

$$y_i = X_i\alpha + \theta\lambda_i + v_i$$

where

$$\lambda_i = E(\varepsilon_i | \nu_i \leq Z_i\gamma)$$

and ε_i is assumed to be normally distributed with mean 0 and variance σ^2, whereas taking full account of the heteroscedasticity of ε_i would require the estimation of

$$y_i = X_i\alpha + \theta_1\lambda_{1i} + \theta_2\lambda_{2i} + v_i$$

where

$$\lambda_{1i} = E(\varepsilon_{1i} | \nu_i \leq Z_i\gamma) \text{ if } i\epsilon S_1 \quad \text{and} \quad \lambda_{1i} = 0 \text{ if } i\epsilon S_2$$

$$\lambda_{2i} = E(\varepsilon_{2i} | \nu_i \leq Z_i\gamma) \text{ if } i\epsilon S_2 \quad \text{and} \quad \lambda_{2i} = 0 \text{ if } i\epsilon S_1$$

and $\varepsilon_{1i} \sim N(0,\sigma_1^2)$ and $\varepsilon_{2i} \sim N(0,\sigma_2^2)$, with ordinary least squares by substituting consistent estimates of λ_{1i} and λ_{2i}. Clearly the imposition of this incorrect restriction will lead to a inconsistent estimate of α. The ML-model (15) can be considered to suffer from exactly the same problem as the model discussed in Section 2.

What we have proved for model (3) is that if α is consistently estimated, $\bar{\gamma}$ is inconsistent and if γ is consistently estimated, $\bar{\alpha}$ is inconsistent. Unlike the model discussed in the previous section, it is unclear how to prove that $\bar{\gamma}$ or $\bar{\alpha}$ are unconditionally inconsistent. The method used in section 2 can not be applied because we can not eliminate the summation over the observations in S_0. To study this point a Monte Carlo experiment will be conducted in the next section.

4. Some Monte Carlo evidence.

To investigate the remaining problems and to get some insight in the bias of the estimates a Monte Carlo experiment will be carried out. Two models will be considered. The first model is discussed in section 2 and has the following structure:

$$y_i = \alpha_0 + x_{1i}\alpha_1 + \theta I_i + \varepsilon_i$$

$$I_i^* = \gamma_0 + x_{1i}\gamma_1 - \nu_i$$

where $I_i = 1$ if $I_i^* \geq 0$ and $I_i = 0$ otherwise, x_{1i} is a scalar randomly drawn from a uniform (0,1)-distribution. The coefficients were set to: $\alpha_0 = 1$, $\alpha_1 = 1$, $\theta = 1$, $\gamma_0 = -1$ and $\gamma_1 = 1$. The normally distributed error term ε_i has mean 0 and variance 0.5 if $I_i = 0$ and 2 if $I_i = 1$. Both these variances will have to be estimated. The mean and variance of the normally distributed error term ν_i are 0 and 1. The correlation between ε_i and ν_i is put to 0.8. 1000 independent data sets of 1000 observations were created and the model was estimated with homoscedastic and heteroscedastic Maximum Likelihood. Table 1 gives information on the results:

-INSERT TABLE 1-

This table clearly illustrates that the homoscedastic estimation results are biased. In particular, the coefficient of θ is biased strongly. The true value (1) is even much smaller than the minimum of the homoscedastic estimates (2.132). Given the importance of this parameter estimate (cf. the introduction), heteroscedasticity of the error term of the linear equation should be a common procedure in models like model (1). The quality of the homoscedastic estimate of the correlation between the error terms of the model is also very poor: again, the true value (0.8) does not lie in the range of estimated coefficients (0.891-0.982).

To investigate the bias of the estimates of the model discussed in Section 3 consider the following specification:

$$y_i = \alpha_0 + x_i\alpha_1 + \varepsilon_i \quad \text{observed if } I_i = 1$$

$$I_i^* = \gamma_0 + x_i\gamma_1 - \nu_i$$

where $I_i = 1$ if $I_i^* > 0$ and $I_i = 0$ if $I_i^* \leq 0$. The variance of ε_i is either σ_1^2 or σ_2^2. x_i is drawn from the uniform (0,1)-distribution. The coefficients were set to: $\alpha_0 = 1$, $\alpha_1 = 1$, $\gamma_0 = -1$ and $\gamma_1 = 1$. The normally distributed error term ε_i has mean 0 and variance 0.5 or 2.0. Which of these variances applies to an individual observation is decided upon by a simple selection rule: a standard normal random variable is drawn and if it exceeds 0 the variance was put to 2.0 and otherwise to 0.5. The mean and variance of the normally distributed error term ν_i are 0 and 1. The correlation between ε_i and ν_i is put to 0.8. 1000 independent data sets of 1000 observations were created and the model was estimated with homoscedastic and heteroscedastic Maximum Likelihood. Table 2 gives the results.

-INSERT TABLE 2-

The estimation results of the homoscedastic maximum likelihood estimation are particularly poor for the parameters of the linear equation. The range of estimate α_0 does not even cover the true value of the coefficient. Given the fact that the estimation of the linear equation is the primary goal in this model this is a very worrisome result. Heteroscedasticity has a very strong impact on the quality of the estimation results. The range of the estimates of ρ does not contain the true value also. The quality of the estimates of the probit-type equation are much better, but still there seems to exist a negative bias.

5. CONCLUSION.

The effect of ignoring heteroscedasticity of the error term of the linear equation in a simultaneous equation model consisting of a linear equation and a qualitative variables equation is that the parameter estimates are no longer consistent. The Monte Carlo analysis performed in Section 4 demonstrates that the bias of the estimates is quite substantial. It should therefore be common practice to test the linear equation of the simultaneous model against heteroscedasticity.

Finally, it should be noted that a two-stage estimation method such as the one described in Heckman (1979) has an advantage compared to homoscedastic maximum likelihood estimation of a heteroscedastic model: the coefficients of the qualitative variable equation will be estimated consistently.

References.

Björklund, A. and R. Moffitt, 1987, The estimation of wage gains and welfare gains in self-selection, *The Review of Economics and Statistics*, 69, pp. 42-49.

Cramer, J.S., 1986, *Econometric applications of Maximum Likelihood methods*, Cambridge University Press, Cambridge (Mass.).

Goddeeris, J.H., 1988, Compensating differentials and self-selection: an application to lawyers, *Journal of Political Economy*, 96, pp. 411-428.

Heckman, J.J., 1979, Sample selection bias as a specification error, *Econometrica*, 47, pp. 153-161.

Hurd, M., 1979, Estimation in truncated samples when there is heteroscedasticity, *Journal of Econometrics*, 11, pp. 247-258.

Johnson, N.L. and S. Kotz, 1970, *Distributions in Statistics: Continuous Univariate Distributions-2*, John Wiley & Sons, New York.

Maddala, G.S., 1983, *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge (Mass.).

Willis, R. and S. Rosen, 1979, Education and self-selection, *Journal of Political Economy*, 87, pp. S7-S36.

Yatchew, A. and Z. Griliches, 1984, Specification error in probit models, *The Review of Economics and Statistics*, 66, pp. 134-39.

Table 1: Characteristics of 1000 replications of the selectivity model.

	homoscedastic ML				heteroscedastic ML			
coefficient	mean	variance	min	max	mean	variance	min	max
α_0 (=1)	1.066	0.01598	0.702	1.490	1.007	0.01213	0.714	1.396
α_1 (=1)	0.609	0.04596	-0.044	1.345	1.008	0.00808	0.691	1.280
θ (=1)	2.712	0.02075	2.132	3.184	0.984	0.02519	0.281	1.415
γ_0 (=-1)	-1.098	0.00569	-1.393	-0.842	-1.003	0.00730	-1.327	-0.707
γ_1 (=1)	0.779	0.01829	0.323	1.287	1.004	0.01956	0.563	1.476
σ	1.947	0.00416	1.751	2.170				
σ_1 (=0.5)					0.498	0.00128	0.355	0.627
σ_2 (=2)					1.993	0.00781	1.672	2.338
ρ (=0.8)	0.952	0.00018	0.891	0.982	0.792	0.00345	0.395	0.934

True parameter values between parentheses.

Table 2: Characteristics of 1000 replications of the Heckman selectivity model.

	homoscedastic ML				heteroscedastic ML			
coefficient	mean	variance	min	max	mean	variance	min	max
α_0 (=1)	2.712	0.05946	1.855	3.467	0.997	0.01124	0.682	1.413
α_1 (=1)	0.029	0.09739	-1.264	1.166	1.004	0.01446	0.570	1.410
γ_0 (=-1)	-0.966	0.00782	-1.311	-0.703	-1.002	0.00798	-1.291	-0.728
γ_1 (=1)	0.013	0.01954	0.483	1.552	1.002	0.01988	0.549	1.487
σ	2.127	0.01496	1.763	2.480				
σ_1 (=0.5)					0.499	0.00157	0.385	0.640
σ_2 (=2)					1.998	0.01040	1.657	2.335
ρ (=0.8)	0.972	0.00011	0.922	0.994	0.798	0.00175	0.653	0.912

True parameter values between parentheses.

APPENDIX

A note on notation:
Throughout this Appendix the inverse Mills'ratio will be denoted by:

$$H(x) = \frac{\phi(x)}{1 - \Phi(x)}$$

the related function M(x) is defined as:

$$M(x) = \frac{\phi(x)}{\Phi(x)}$$

Proof of Lemma 1.

$$G(\varepsilon) = \frac{\phi(Z\gamma^* - \rho^*\varepsilon)}{1 - \Phi(Z\gamma^* - \rho^*\varepsilon)}$$

Clearly,

$$\frac{\partial G(\varepsilon)}{\partial \varepsilon} = -\rho^* \frac{\partial H(x)}{\partial x} \quad \text{and} \quad \frac{\partial^2 G(\varepsilon)}{\partial \varepsilon^2} = \rho^{*2} \frac{\partial^2 H(x)}{\partial x^2}$$

Using the results that H(x) > 0 and H(x) > x (cf. Johnson and Kotz, 1970, p. 279):

$$\frac{\partial H(x)}{\partial x} = H(x)(H(x) - x) > 0$$

This proves (a) and (b). The second derivative of H(x) equals:

$$\frac{\partial^2 H(x)}{\partial x^2} = \frac{\partial H(x)}{\partial x}(H(x) - x) + H(x)\left[\frac{\partial H(x)}{\partial x} - 1\right] = H(x)\Big((H(x) - x)^2 + H(x)(H(x) - x) - 1\Big)$$

This expression exceeds 0 for all x if

$$y(x)^2 + (y(x)(x + y(x)) - 1 \qquad \text{(A1)}$$

exceeds 0, where y(x) = H(x) - x. The following properties of y(x) are relevant here:

(a) y(x) = H(x) - x > 0 because H(x) > x (cf. Johnson and Kotz, 1970, p. 279);

(b) $y(x) \to 0$ if $x \to \infty$ because $H(x) \to x$ if $x \to \infty$ (x < H(x) < x + (1/x), cf. Johnson and Kotz, 1970, p. 279);

(c) $\partial y(x)/\partial x < 0$, because $\partial y(x)/\partial x = \partial H(x)/\partial x - 1 = H(x)(H(x) - x) - 1 = -(1 + H(x)x - H(x)^2) < 0$ because $1 + H(x)x - H(x)^2$ is the variance of a unit normal variable truncated from below at x (cf. Johnson and Kotz, 1970, p. 278).

Thus, y(x) is a strictly positive decreasing function of x. Taking the first derivative of (A1) we get:

$$\frac{\partial y(x)}{\partial x}\Big(2y(x) + (x + y(x)) + y(x)\Big)$$

which, due to the properties of y(x) and the fact that x + y(x) = H(x) > 0, is negative. So, (A1) is a decreasing function of x. Finally, by noting that:

$$\lim_{x\to\infty}(y(x)^2 + y(x)(x + y(x)) - 1) = 0$$

because of the property (b) of y(x) ($y(x)(x + y(x)) = \partial H/\partial x \to 1$ if $x \to \infty$), we have established that (A1) exceeds 0 and therefore that $\partial H^2/\partial x > 0$ and consequently, $\partial G^2/\partial \varepsilon > 0$.

□

Proof of Lemma 2.

$$E(F(\varepsilon_{1i})^\theta) = \int_{-\infty}^{\infty}\left[\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}\right]^\theta \phi(\varepsilon_{1i})\, d\varepsilon_{1i} \qquad \theta = 1,2 \qquad \text{(A1)}$$

$$\lim_{\varepsilon_{1i}\to-\infty}\left[\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^{\prime} - \rho^*\varepsilon_{1i})}\right]^\theta \phi(\varepsilon_{1i}) = \frac{0}{1}0 = 0 \qquad \theta = 1,2$$

By applying L'Hôpital's rule we find:

$$\lim_{\varepsilon_{1i}\to\infty}\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})} = \lim_{\varepsilon_{1i}\to\infty}\frac{\rho^*(Z_i\gamma^* - \rho^*\varepsilon_{1i})\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{-\rho^*\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})} = \lim_{\varepsilon_{1i}\to\infty}-(Z_i\gamma^* - \rho^*\varepsilon_{1i})$$

$$\lim_{\varepsilon_{1i}\to\infty}\left[\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}\right]\phi(\varepsilon_{1i}) = \lim_{\varepsilon_{1i}\to\infty}-(Z_i\gamma^* - \rho^*\varepsilon_{1i})\phi(\varepsilon_{1i}) = 0$$

because $x^q\exp(-x) \to 0$ if $x \to \infty$ and $q \geq 0$.
Again, by implementing L'Hôpital's rule we find:

$$\lim_{\varepsilon_{1i}\to\infty}\left[\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}\right]^2 = \lim_{\varepsilon_{1i}\to\infty}\frac{2\rho^*(Z_i\gamma^* - \rho^*\varepsilon_{1i})\phi^2(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{-2\rho^*\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})} = \lim_{\varepsilon_{1i}\to\infty}-(Z_i\gamma^* - \rho^*\varepsilon_{1i})\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}$$

$$\lim_{\varepsilon_{1i}\to\infty}\left[\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}\right]^2\phi(\varepsilon_{1i}) = \lim_{\varepsilon_{1i}\to\infty}-(Z_i\gamma^* - \rho^*\varepsilon_{1i})\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}\phi(\varepsilon_{1i}) = 0$$

due to the previous result.
Both $\phi(.)$ and $\Phi(.)$ are bounded strictly positive and continuous functions and therefore:

$$\max_{\varepsilon_{1i}}\left[\frac{\phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}{\Phi(Z_i\gamma^* - \rho^*\varepsilon_{1i})}\right]^\theta\phi(\varepsilon_{1i}) < \infty \qquad \theta = 1,2$$

Given these results (A1) is finite for both $\theta = 1$ and $\theta = 2$.

□

Proof of Proposition 1.

$$\theta^k_{1\gamma i} = \int_{-\infty}^{\infty} \frac{\partial F(\varepsilon_{1i})}{\partial \gamma_k(\sigma)}\phi(\varepsilon_{1i})\, d\varepsilon_{1i} \quad \text{and} \quad \theta^k_{0\gamma i} = \int_{-\infty}^{\infty} \frac{\partial G(\varepsilon_{0i})}{\partial \gamma_k(\sigma)} f(\varepsilon_{0i})\, d\varepsilon_{0i} \tag{A3}$$

where

$$\frac{\partial F(\varepsilon_{1i})}{\partial \gamma_k(\sigma)} = \frac{1}{\sqrt{1-\rho^2}}\frac{\partial F(\varepsilon_{1i})}{\partial \gamma_k^*} = -\frac{1}{\sqrt{1-\rho^2}}F(\varepsilon_{1i})(\gamma_k^* - \rho^*\varepsilon_{1i} + F(\varepsilon_{1i}))Z_{ik} = \Xi_{ik}Z_{ik} \tag{A3a}$$

$$\frac{\partial G(\varepsilon_{0i})}{\partial \gamma_k(\sigma)} = \frac{1}{\sqrt{1-\rho^2}}\frac{\partial G(\varepsilon_{0i})}{\partial \gamma_k^*} = \frac{1}{\sqrt{1-\rho^2}}G(\varepsilon_{0i})(G(\varepsilon_{0i}) - (\gamma_k^* - \rho^*\varepsilon_{0i}))Z_{ik} = \Omega_{ik}Z_{ik} \tag{A3b}$$

The function $x + M(x)$ exceeds zero if $\Phi(x)x + \phi(x)$ exceeds 0. This holds because it is an increasing function of x (first derivative $\Phi(x) > 0$) and $\Phi(x)x + \phi(x) \to 0$ if $x \to -\infty$. Similarly the function $H(x) - x$ exceeds zero if $\phi(x) -(1-\Phi(x))x$ exceeds 0. This function is decreasing in x (first derivative $-(1-\Phi(x)) < 0$) and $\phi(x) - (1-\Phi(x))x \to 0$ if $x \to \infty$. Consequently Ξ_{ik} is strictly negative and Ω_{ik} is a strictly positive function of $\tilde{\gamma}$ and therefore:

$$r_1 \lim_{n_1\to\infty}\frac{1}{n_1}\sum_{i\in S_1} Z_{ik}\theta^k_{1\gamma i} = r_1 \lim_{n_1\to\infty}\frac{1}{n_1}\sum_{i\in S_1} Z_{ik}\Xi_{ik}Z_{ik} < 0 \quad \text{and}$$

$$r_0 \lim_{n_0\to\infty}\frac{1}{n_0}\sum_{i\in S_0} Z_{ik}\theta^k_{0\gamma i} = r_0 \lim_{n_0\to\infty}\frac{1}{n_0}\sum_{i\in S_0} Z_{ik}\Omega_{ik}Z_{ik} > 0$$

if $Z_{ik} \neq 0$ for all i. Consequently, $\partial T_k/\partial\gamma_k(\sigma) < 0$.

□

Proof of Proposition 2.

Define

$$h(\varepsilon_{0i}) = f(\varepsilon_{0i})\left(1 - \frac{\varepsilon_{0i}^2}{\sigma^2}\right)$$

where f(.) is the normal density with expectation 0 and variance σ^2. This function is symmetric in ε_{0i} ($h(\varepsilon_{0i}) = h(-\varepsilon_{0i})$), $h(\varepsilon_{0i}) = 0$ if and only if $\varepsilon_{0i} = \sigma$, $h(\varepsilon_{0i}) > 0$ if $-\sigma \leq \varepsilon_{0i} < \sigma$, $h(\varepsilon_{0i}) < 0$ if $|\varepsilon_{0i}| > \sigma$ and

$$\int_{-\infty}^{\infty} h(\varepsilon_{0i})\, d\varepsilon_{0i} = \int_{-\infty}^{\infty} f(\varepsilon_{0i})\, d\varepsilon_{0i} - \frac{1}{\sigma^2}\int_{-\infty}^{\infty} \varepsilon_{0i}^2 f(\varepsilon_{0i})\, d\varepsilon_{0i} = 0$$

Because of the symmetry of $h(\varepsilon_{0i})$

$$\int_0^{\infty} h(\varepsilon_{0i})\, d\varepsilon_{0i} = \int_0^{\sigma} h(\varepsilon_{0i})\, d\varepsilon_{0i} + \int_{\sigma}^{\infty} h(\varepsilon_{0i})\, d\varepsilon_{0i} = 0$$

and therefore:

$$\int_{0}^{\sigma} h(\varepsilon_{0i})\, d\varepsilon_{0i} = -\int_{\sigma}^{\infty} h(\varepsilon_{0i})\, d\varepsilon_{0i} \tag{A4}$$

Consider

$$\theta_{\sigma i} = \int_{-\infty}^{\infty} G(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i}$$

Making use of the properties of h(.) and G(.) we can deduce:

$$\theta_{\sigma i} = \int_{-\infty}^{0} G(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i} + \int_{0}^{\infty} G(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i} =$$

$$= \int_{\infty}^{0} -G(-\varepsilon_{0i})h(-\varepsilon_{0i})\, d(-\varepsilon_{0i}) + \int_{0}^{\infty} G(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i} =$$

$$= \int_{0}^{\infty} (G(-\varepsilon_{0i}) + G(\varepsilon_{0i}))h(\varepsilon_{0i})\, d\varepsilon_{0i} = \int_{0}^{\infty} G^{*}(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i}$$

The function $G^{*}(\varepsilon_{0i})$ has the following properties:

$$G^{*}(\varepsilon_{0i}) = G^{*}(-\varepsilon_{0i})$$

$$G^{*}(\varepsilon_{0i},\rho^{*}) = G^{*}(\varepsilon_{0i},-\rho^{*}) \qquad (G^{*}(x) \text{ is symmetric in } \rho)$$

$$\frac{\partial G^{*}(\varepsilon_{0i})}{\partial \varepsilon_{0i}} = \rho^{*}[(\bar{\gamma} - \rho^{*}\varepsilon_{0i})G(\varepsilon_{0i}) - (\bar{\gamma} + \rho^{*}\varepsilon_{0i})G(-\varepsilon_{0i}) +$$

$$G(-\varepsilon_{0i})^{2} - G(\varepsilon_{0i})^{2}] = -\frac{\partial G^{*}(-\varepsilon_{0i})}{\partial \varepsilon_{0i}}$$

$$\frac{\partial G^{*}(\varepsilon_{0i})}{\partial \varepsilon_{0i}} = 0 \qquad \text{if } \varepsilon_{0i} = 0$$

$$\frac{\partial G^{*}(\varepsilon_{0i})}{\partial \varepsilon_{0i}} > 0 \qquad \text{if } \varepsilon_{0i} > 0$$

$$\frac{\partial G^{*}(\varepsilon_{0i})}{\partial \varepsilon_{0i}} < 0 \qquad \text{if } \varepsilon_{0i} < 0$$

These last two results follow from Lemma 1. Consequently $G^{*}(\varepsilon_{0i})$ is a strictly increasing function for $\varepsilon_{0i} > 0$. Splitting $\theta_{\sigma i}$ up we get:

$$\theta_{\sigma i} = \int_0^{\sigma} G^*(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i} + \int_{\sigma}^{\infty} G^*(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i}$$

$$< \int_0^{\sigma} G^*(\varepsilon_{0i})h(\varepsilon_{0i})\, d\varepsilon_{0i} + \min_{\varepsilon_{0i}\in[\sigma,\infty)} G^*(\varepsilon_{0i}) \int_{\sigma}^{\infty} h(\varepsilon_{0i})\, d\varepsilon_{0i}$$

because $h(\varepsilon_{0i}) < 0$ and $G^*(\varepsilon_{0i}) > 0$ for $\varepsilon_{0i} > \sigma$. Because $G^*(\varepsilon_{0i})$ is an increasing function beyond $\varepsilon_{0i} = 0$, we know $\min G^*(\varepsilon_{0i}) = G^*(\sigma)$ for $\varepsilon_{0i} \geq \sigma$. Using (A4) we get:

$$\theta_{\sigma i} < \int_0^{\sigma} (G^*(\varepsilon_{0i}) - G^*(\sigma))h(\varepsilon_{0i})\, d\varepsilon_{0i} < 0.$$

because $h(\varepsilon_{0i}) > 0$ and $G^*(\varepsilon_{0i}) < G^*(\sigma)$ for $\varepsilon_{0i} \in (0,\sigma)$.

□