UNIVERSITÄT OSNABRÜCK

Working Papers of the Institute of Empirical Economic Research

Testing the normality assumption in an ordered probit model using an artificial regression – some results for the LM-test

Joachim Wilde, Sarah Forstinger

Working Paper No. 127
March 2026

Osnabrueck University
Institute of Empirical Economic Research
Rolandstr. 8 · 49069 Osnabrück · Germany

School of Business Administration and Economics
Institute of Empirical Economic Research
www.wiwi.uni-osnabrueck.de

Testing the normality assumption in an ordered probit model using an artificial regression– some results for the LM-test

Joachim Wilde and Sarah Forstinger

Osnabrueck University, Department of Economics, Rolandstr. 8, 49069 Osnabrueck, Germany

(corresponding author: Joachim.Wilde@uni-osnabrueck.de)

March 2026

Abstract

The key assumption of normally distributed error terms is usually not tested in empirical practice when using ordered probit models. Therefore, an artificial regression version of the LM test against the class of Pearson distributions is derived that can be implemented more easily than the well-known matrix version. A comprehensive simulation study analyses the properties of the LM test and of the t-statistics in the artificial regression that correspond to skewness and fat tails, respectively. For most designs a large power against skewness and a moderate power against fat tails are found. However, the t-statistics against skewness and fat tails exhibit notable size distortions. Therefore, new double indicators are proposed. The simulation results indicate that the double indicators avoid the size distortions and exhibit power characteristics similar to the original statistics for most designs.

Keywords:

ordered probit model, normality assumption, Lagrange multiplier test, artificial regression

JEL classification: C25

1. Introduction

Testing the normality assumption is more important for probit models than for linear regression models because nonnormal disturbances hamper consistency of the standard (ML) estimator of the probit model. Thus, it should be tested before the results of a probit model are evaluated. A rather general alternative hypothesis for such a test is the class of Pearson distributions (cf. Ord 1972, ch. 1 for a detailed description of this class of distributions). It contains also skew distributions and distributions with fat tails. The normal distribution is a special case with symmetry 0 and kurtosis 3. The class of Pearson distributions was used by Bera/ Jarque/ Lee (1984) for testing the normality assumption in a binary probit model. Since the class is characterized by differential equations, a ML estimation under the alternative is not easily available. Thus, the LM principle is attractive because it only requires estimation under the null hypothesis of normally distributed disturbances, i.e. for the standard probit model. However, calculating the information matrix in the test-statistic requires expectations under the alternative. For the binary probit these expectations are given by Bera/ Jarque/ Lee (1984), for the ordinary probit with more than two categories general formulas are given by Johnson (1996) and Glewwe (1997).

Although all ingredients for the LM-test are available testing normality in applied ordered probit models is usually missing. The test is still not implemented in standard software like Stata. Furthermore, the formulas for a matrix implementation are complicated, which presumably has a deterrent effect, too. Thus, using an artificial linear regression for calculating the test-statistic may be useful. Two versions of calculating the LM-statistic via an artificial linear regression are suggested in the literature: The first one estimates the information matrix by the outer-product-of-the-gradient (OPG). It is easy to implement, but it behaves poorly in small and even medium samples (cf. Davidson/ MacKinnon 1993, p. 477). The second one is equivalent to the original test-statistic. Thus, the properties are the same as those of the original test-statistic, and it should be preferred over the OPG-version. It was derived for the binary probit in Wilde (2008), whereas a solution for the ordered probit model is still open. Johnson (1996) doesn’t consider an artificial regression. Glewwe (1997) only uses the OPG-version and obtained the expected poor results in his simulation study – even for a sample size of 1000 (p. 10). Murphy (1994) mentions the possibility of an artificial regression but didn’t give any formulas for the case of more than two categories.

The following paper closes this gap and derives an artificial regression version of the LM test on normality against the class of Pearson distributions that is equivalent to the original test-statistic. As in the binary case the artificial regression contains a regressor that controls for a skew distribution and a regressor that controls for fat tails. Thus, if the normal distribution is rejected the t-statistics of these regressors may be useful to decide in which direction(s) the distribution differs from the normal one. A broad simulation study analyzes the properties of the LM test as a whole and of the t-statistics of the two relevant regressors. To the best of our knowledge, the latter aspect has not been analyzed in the literature even for binary probit models so far. We observe size distortions for the t-statistics, i.e. they cannot be used as the sole criterion to test for skewness or fat tails. However, we derive a double indicator using the t-statistics and their ratio that shows whether skewness or fat tails is the main problem if the LM test rejects the normal distribution. An R code for calculating the test statistics and the indicator is given in the appendix.

2. Artificial regression for an ordered probit model with three categories

We define an ordered probit model as usual by using a linear regression model for a latent variable that is linked via a threshold condition to the observed endogenous outcome:

$$y_i^* = \beta x_i + u_i, \quad i = 1, \ldots, N,$$

$$y_i = \begin{cases} 1, & \text{if } y_i^* \le \mu_1 \\ 2, & \text{if } \mu_1 < y_i^* \le \mu_2 \\ 3, & \text{if } \mu_2 < y_i^* \end{cases} \tag{2.1}$$

y_i^* is a latent continuous variable, x_i an exogenous variable, β the corresponding parameter, and μ_1 and μ_2 the threshold parameters. For ease of exposition we take three categories and one exogenous variable here. Generalized formulas for an arbitrary number of categories and exogenous variables are given in appendix 1. The disturbances u_i are assumed to be independent normally distributed with zero expectation and variance 1 (given x_i). The latter value is due to the usual normalization of parameters in a probit model because all parameters are only identified up to a positive scalar. For a similar reason either the constant or a threshold parameter has to be normalized to zero (cf. Greene and Hensher 2010: 107). We normalize the constant to zero because this facilitates the interpretation of the simulation design.

The loglikelihood function for the Maximum Likelihood estimator of the model above is:

$$\ln L(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{3} y_{ij} \ln\left(p_{ij}(\theta)\right), \; y_{ij} = \begin{cases} 1, & \text{if } y_i = j \\ 0, & \text{otherwise} \end{cases}, \; p_{ij}(\theta) = P\left(y_{ij} = 1 \mid \theta, x_i\right), j = 1, 2, 3 \tag{2.2}$$

$\theta = (\beta, \mu_1, \mu_2)'$ the vector of unknown parameters,

$$p_{i1}(\theta) = \Phi(\mu_1 - \beta x_i), \; p_{i2}(\theta) = \Phi(\mu_2 - \beta x_i) - \Phi(\mu_1 - \beta x_i), \; p_{i3}(\theta) = 1 - \Phi(\mu_2 - \beta x_i) \tag{2.3}$$

$\Phi(\ldots)$ denoting the cumulative distribution function of a standard normal distribution.

The LM-statistic is based on the loglikelihood function under the alternative, i.e. under the general class of Pearson distributions. Whereas the structure of the loglikelihood function in (2.2) is the same under the null and the alternative, θ and the formulas in (2.3) are different. Now $\theta = (\beta, \mu_1, \mu_2, c_1, c_2)'$, c_1 a parameter of the Pearson class that measures skewness and c_2 a parameter of the Pearson class that measures small or fat tails. c_1 is zero for a symmetric distribution, c_2 is zero for a symmetric distribution with kurtosis 3 (cf. Johnson/ Kotz/ Balakrishnan 1994, p. 22). Thus, under normality both parameters are equal to zero, and the hypotheses tested are:

H_0: $c_1 = c_2 = 0$ against H_1: c_1 or $c_2 \neq 0$.

Using the information matrix equality, the LM-statistic can be calculated as follows (cf. Wilde 2008, p. 120):

$$LM = \left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \theta}\right)' \left[E_{H_0}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \theta}\right)\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \theta}\right)'\right]^{-1}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \theta}\right), \qquad (2.4)$$

ln L the loglikelihood function under H_1, $\hat{\theta}_r = (\hat{\beta}, \hat{\mu}_1, \hat{\mu}_2, 0, 0)'$ the restricted estimator under H_0 (i.e., $\hat{\beta}$, $\hat{\mu}_1$ and $\hat{\mu}_2$ are the ML-estimates of an ordinal probit), and E_{H_0} the expectation over the y_{ij}'s given $c_1 = c_2 = 0$ and given x_i.

Let

$$\hat{\mu}_{im} = \hat{\mu}_m - \hat{\beta}x_i,\ \hat{p}_{i1} = \Phi(\hat{\mu}_{i1}),\ \hat{p}_{i2} = \Phi(\hat{\mu}_{i2}) - \Phi(\hat{\mu}_{i1}),\ \hat{p}_{i3} = 1 - \Phi(\hat{\mu}_{i2}), \text{ and}$$
$$\hat{\varphi}_{im} = \varphi(\hat{\mu}_{im}),\ m = 1, 2, \qquad (2.5)$$

$\varphi(\ldots)$ the density function of a standard normal distribution. Then, the gradient vector equals (cf. Johnson 1996, pp. 215-216):

$$\frac{\partial \ln L}{\partial \mu_1}(\hat{\theta}_r) = \sum_{i=1}^{N}\left[\frac{y_{i1}}{\hat{p}_{i1}} - \frac{y_{i2}}{\hat{p}_{i2}}\right]\hat{\varphi}_{i1} \qquad (2.6a)$$

$$\frac{\partial \ln L}{\partial \mu_2}(\hat{\theta}_r) = \sum_{i=1}^{N}\left[\frac{y_{i2}}{\hat{p}_{i2}} - \frac{y_{i3}}{\hat{p}_{i3}}\right]\hat{\varphi}_{i2} \qquad (2.6b)$$

$$\frac{\partial \ln L}{\partial \beta}(\hat{\theta}_r) = \sum_{i=1}^{N}\left[y_{i1}\frac{-\hat{\varphi}_{i1}}{\hat{p}_{i1}} + y_{i2}\frac{\hat{\varphi}_{i1} - \hat{\varphi}_{i2}}{\hat{p}_{i2}} + y_{i3}\frac{\hat{\varphi}_{i2}}{\hat{p}_{i3}}\right]x_i \qquad (2.6c)$$

$$\frac{\partial \ln L}{\partial c_1}(\hat{\theta}_r) = -\frac{1}{3}\sum_{i=1}^{N}\left[y_{i1}\frac{-(\hat{\mu}_{i1}^2 - 1)\hat{\varphi}_{i1}}{\hat{p}_{i1}} + y_{i2}\frac{(\hat{\mu}_{i1}^2 - 1)\hat{\varphi}_{i1} - (\hat{\mu}_{i2}^2 - 1)\hat{\varphi}_{i2}}{\hat{p}_{i2}} + y_{i3}\frac{(\hat{\mu}_{i2}^2 - 1)\hat{\varphi}_{i2}}{\hat{p}_{i3}}\right] \qquad (2.6d)$$

$$\frac{\partial \ln L}{\partial c_2}(\hat{\theta}_r) = \frac{1}{4}\sum_{i=1}^{N}\left[y_{i1}\frac{-\hat{\mu}_{i1}(3 + \hat{\mu}_{i1}^2)\hat{\varphi}_{i1}}{\hat{p}_{i1}} + y_{i2}\frac{\hat{\mu}_{i1}(3 + \hat{\mu}_{i1}^2)\hat{\varphi}_{i1} - \hat{\mu}_{i2}(3 + \hat{\mu}_{i2}^2)\hat{\varphi}_{i2}}{\hat{p}_{i2}} + y_{i3}\frac{\hat{\mu}_{i2}(3 + \hat{\mu}_{i2}^2)\hat{\varphi}_{i2}}{\hat{p}_{i3}}\right] \qquad (2.6e)$$

Denoting $\mu = (\mu_1\ \mu_2)'$, the information matrix can be partitioned as follows:

$$E_{H_0}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \theta}\right)\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \theta}\right)'$$

$$= E_{H_0}\begin{pmatrix}
\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2} \\
\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2} \\
\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2} \\
\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}
\end{pmatrix}$$

$$=: \begin{pmatrix}
J_{\beta\beta} & J_{\beta\mu} & J_{\beta c_1} & J_{\beta c_2} \\
J_{\beta\mu}' & J_{\mu\mu} & J_{\mu c_1} & J_{\mu c_2} \\
J_{\beta c_1} & J_{\mu c_1}' & J_{c_1 c_1} & J_{c_1 c_2} \\
J_{\beta c_2} & J_{\mu c_2}' & J_{c_1 c_2} & J_{c_2 c_2}
\end{pmatrix}. \tag{2.7}$$

Following Glewwe (1997, pp. 5-6) the partial matrices are:[1]

$$J_{\beta\beta} = \sum_{i=1}^{N} x_i^2\left(\frac{\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \frac{(\hat{\varphi}_{i1}-\hat{\varphi}_{i2})^2}{\hat{p}_{i2}} + \frac{\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right) \tag{2.8a}$$

$$J_{\beta\mu} = \left(\sum_{i=1}^{N} x_i\left(\frac{-\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} - \frac{(\hat{\varphi}_{i1}-\hat{\varphi}_{i2})\hat{\varphi}_{i1}}{\hat{p}_{i2}}\right) \quad \sum_{i=1}^{N} x_i\left(\frac{(\hat{\varphi}_{i1}-\hat{\varphi}_{i2})\hat{\varphi}_{i2}}{\hat{p}_{i2}} - \frac{\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right)\right) \tag{2.8b}$$

$$J_{\beta c_1} = -\frac{1}{3}\sum_{i=1}^{N} x_i\left(\frac{(\hat{\mu}_{i1}^2-1)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \frac{(\hat{\varphi}_{i1}-\hat{\varphi}_{i2})\left((\hat{\mu}_{i1}^2-1)\hat{\varphi}_{i1}-(\hat{\mu}_{i2}^2-1)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}} + \frac{(\hat{\mu}_{i2}^2-1)\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right) \tag{2.8c}$$

$$J_{\beta c_2} = \frac{1}{4}\sum_{i=1}^{N} x_i\left(\frac{\hat{\mu}_{i1}(3+\hat{\mu}_{i1}^2)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \frac{(\hat{\varphi}_{i1}-\hat{\varphi}_{i2})\left(\hat{\mu}_{i1}(3+\hat{\mu}_{i1}^2)\hat{\varphi}_{i1}-\hat{\mu}_{i2}(3+\hat{\mu}_{i2}^2)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}} + \frac{\hat{\mu}_{i2}(3+\hat{\mu}_{i2}^2)\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right) \tag{2.8d}$$

[1] The notation is somewhat different because Glewwe normalizes the first threshold instead of the constant to zero. Furthermore, Glewwe also uses terms containing e.g. $\varphi(\infty)$ for a general notation, while we explicitly set these terms to 0.

$$J_{\mu\mu}=\begin{pmatrix}\sum_{i=1}^{N}\left(\frac{\hat{\varphi}_{i1}^2}{\hat{p}_{i1}}+\frac{\hat{\varphi}_{i1}^2}{\hat{p}_{i2}}\right) & -\sum_{i=1}^{N}\frac{\hat{\varphi}_{i1}\hat{\varphi}_{i2}}{\hat{p}_{i2}}\\ -\sum_{i=1}^{N}\frac{\hat{\varphi}_{i1}\hat{\varphi}_{i2}}{\hat{p}_{i2}} & \sum_{i=1}^{N}\left(\frac{\hat{\varphi}_{i2}^2}{\hat{p}_{i2}}+\frac{\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right)\end{pmatrix} \tag{2.8e}$$

$$J_{\mu c_1}=\begin{pmatrix}-\frac{1}{3}\sum_{i=1}^{N}\left(\frac{-\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}}-\frac{\hat{\varphi}_{i1}\left(\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}-\left(\hat{\mu}_{i2}^2-1\right)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}}\right)\\ -\frac{1}{3}\sum_{i=1}^{N}\left(\frac{\hat{\varphi}_{i2}\left(\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}-\left(\hat{\mu}_{i2}^2-1\right)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}}-\frac{\left(\hat{\mu}_{i2}^2-1\right)\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right)\end{pmatrix} \tag{2.8f}$$

$$J_{\mu c_2}=\begin{pmatrix}\frac{1}{4}\sum_{i=1}^{N}\left(\frac{-\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}}-\frac{\hat{\varphi}_{i1}\left(\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}-\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}}\right)\\ \frac{1}{4}\sum_{i=1}^{N}\left(\frac{\hat{\varphi}_{i2}\left(\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}-\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}}-\frac{\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right)\end{pmatrix} \tag{2.8g}$$

$$J_{c_1c_1}=\frac{1}{9}\sum_{i=1}^{N}\left(\frac{\left(\hat{\mu}_{i1}^2-1\right)^2\hat{\varphi}_{i1}^2}{\hat{p}_{i1}}+\frac{\left(\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}-\left(\hat{\mu}_{i2}^2-1\right)\hat{\varphi}_{i2}\right)^2}{\hat{p}_{i2}}+\frac{\left(\hat{\mu}_{i2}^2-1\right)^2\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right) \tag{2.8h}$$

$$J_{c_1c_2}=-\frac{1}{12}\sum_{i=1}^{N}\left(\frac{\left(\hat{\mu}_{i1}^2-1\right)\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}}+\frac{\left(\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}-\left(\hat{\mu}_{i2}^2-1\right)\hat{\varphi}_{i2}\right)\left(\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}-\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}}\right.$$
$$\left.+\frac{\left(\hat{\mu}_{i2}^2-1\right)\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right) \tag{2.8i}$$

$$J_{c_2c_2}=\frac{1}{16}\sum_{i=1}^{N}\left(\frac{\hat{\mu}_{i1}^2\left(3+\hat{\mu}_{i1}^2\right)^2\hat{\varphi}_{i1}^2}{\hat{p}_{i1}}+\frac{\left(\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}-\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}\right)^2}{\hat{p}_{i2}}+\frac{\hat{\mu}_{i2}^2\left(3+\hat{\mu}_{i2}^2\right)^2\hat{\varphi}_{i2}^2}{\hat{p}_{i3}}\right) \tag{2.8j}$$

Using these expressions, the procedure described in Wilde (2008, p. 120) can be generalized to three categories, i.e.: Define an artificial regression

$$r = R\delta + \varepsilon \tag{2.9}$$

such that

$$R'r=\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial\theta} \text{ and } R'R=E_{H_0}\left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial\theta}\right)\left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial\theta}\right)'.$$

Thereby

$$LM = r'R\left(R'R\right)^{-1}R'r\,,$$

and letting $\hat{\hat{\delta}}$ be the OLS estimator of δ in (2.9) the test-statistic in (2.4) can be calculated using the explained sum of squares after an OLS estimation of (2.9):

$$LM = r'R\hat{\hat{\delta}} = r'\hat{r} = \hat{r}'\hat{r} \tag{2.10}$$

Although the basic idea remains unchanged from the binary probit case, the formulas become somewhat more complicated. This is due to the fact that in many formulas of (2.6) and (2.8) three summands are needed for each individual. Accordingly, r and R contain three rows instead of one row for each i due to the three random variables y_{i1}, y_{i2} and y_{i3}. In the case of the binary probit model, this can be avoided by using $y_{2i} = 1-y_{1i}$, so that only one random variable is needed there. Accordingly, the original two summands can be combined into one summand and r and R contain only one row for each i. Such a reduction to one summand and one line is no longer possible for 3 or more categories.

Denote with r_i and R_i the rows of r and R for the i-th observation. Then,

$$r_i = \begin{pmatrix} \frac{y_{i1}}{\sqrt{\hat{p}_{i1}}} \\ \frac{y_{i2}}{\sqrt{\hat{p}_{i2}}} \\ \frac{y_{i3}}{\sqrt{\hat{p}_{i3}}} \end{pmatrix},$$

$$R_i = \begin{pmatrix} \frac{-\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}}x_i & \frac{\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} & 0 & \frac{1}{3}\frac{\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} & -\frac{1}{4}\frac{\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} \\ \frac{\hat{\varphi}_{i1}-\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}}x_i & \frac{-\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i2}}} & \frac{\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} & -\frac{1}{3}\frac{\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}-\left(\hat{\mu}_{i2}^2-1\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} & \frac{1}{4}\frac{\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}-\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} \\ \frac{\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}}x_i & 0 & \frac{-\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} & -\frac{1}{3}\frac{\left(\hat{\mu}_{i2}^2-1\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} & \frac{1}{4}\frac{\hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} \end{pmatrix} \tag{2.11a}$$

$$r = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix},\; R = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{pmatrix}. \tag{2.11b}$$

Using these formulas and algebraic transformations, it can be shown that R′r is equal to the gradient vector, and R′R is equal to the information matrix. All components of r and R are simple functions of variables and parameter estimates and can be calculated easily. For each individual three values have to be calculated and set one below the other. This can be done, for example, by using a well-known function of panel data analysis: In panel data analysis the values of variables for different waves are often given as separate columns. This so-called wide format is then converted by a function into the so-called long format, in which for each individual the different waves are set one below the other. In our case the "waves" are the values for the different categories. Thus, for each variable of the artificial regression three columns are calculated separately (including a column of zeros for the two variables related to the threshold parameters), and afterwards the variables are set among each other by the panel data function to obtain the format required for (2.11).

Since the product of r_i and the first three columns of R_i equals the gradient according to β, μ_1 and μ_2 (cf. 2.6a-c and 2.11), and this gradient is zero for the estimate under H_0, any explanatory power of the right-hand side of the artificial regression is caused by the last two columns. These columns refer to the parameters c_1 and c_2. Thus, the t-statistics for these variables, hereafter referred to as t_{c1} and t_{c2}, can be useful to decide whether the error terms are skewed (t_{c1}) or have fat tails (t_{c2}).

3. Design and procedure of the simulation study

Our simulation is based on model (2.1). In our baseline scenario β is equal to one, the x_i's are drawn independently from a N(0, 1) distribution and are fixed across all replications, and the u_i's are drawn independently from a N(0, 1) distribution in each of the 5000 replications. μ_1 and μ_2 are used to control the distribution of the values of y_i. In our baseline scenario we assume a uniform distribution. Since $y_i^* \sim N(0, 2)$, μ_1 is chosen as 33% quantile, and μ_2 is chosen as 67% quantile of this distribution. For N we choose N=400 (medium sample size) and N=2000 (large sample size). Since the test is an asymptotic test, for N=2000 everything should be fine, and N=400 is an indicator how fast the asymptotic results are reached.

Using the simulated data, the following calculations are done for each replication:

1) The ordered probit model (2.1) is estimated via the Maximum Likelihood method, and the estimates $\hat{\beta}$, $\hat{\mu}_1$, and $\hat{\mu}_2$ are saved.
2) Using $\hat{\varphi}_{im}$ and $\hat{p}_{ij}$ (j = 1, 2, 3) as in (2.5), r is computed as in (2.11), and the following explanatory variables for the artificial regression are computed (cf. 2.11a):

$$R_\beta = \begin{pmatrix} R_{1\beta} \\ \vdots \\ R_{N\beta} \end{pmatrix}, \quad R_{i\beta} = \begin{pmatrix} \dfrac{-\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} x_i \\ \dfrac{\hat{\varphi}_{i1} - \hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} x_i \\ \dfrac{\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} x_i \end{pmatrix}, \ i = 1, \ldots, N$$

$$R_{\mu_1} = \begin{pmatrix} R_{1\mu_1} \\ \vdots \\ R_{N\mu_1} \end{pmatrix}, \; R_{i\mu_1} = \begin{pmatrix} \dfrac{\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} \\ \dfrac{-\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i2}}} \\ 0 \end{pmatrix}, \; R_{\mu_2} = \begin{pmatrix} R_{1\mu_2} \\ \vdots \\ R_{N\mu_2} \end{pmatrix}, \; R_{i\mu_2} = \begin{pmatrix} 0 \\ \dfrac{\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} \\ \dfrac{-\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} \end{pmatrix}$$

$$R_{c_1} = \begin{pmatrix} R_{1c_1} \\ \vdots \\ R_{Nc_1} \end{pmatrix}, \; R_{ic_1} = \begin{pmatrix} \dfrac{1}{3}\dfrac{\left(\hat{\mu}_{i1}^2 - 1\right)\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} \\ -\dfrac{1}{3}\dfrac{\left(\hat{\mu}_{i1}^2 - 1\right)\hat{\varphi}_{i1} - \left(\hat{\mu}_{i2}^2 - 1\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} \\ -\dfrac{1}{3}\dfrac{\left(\hat{\mu}_{i2}^2 - 1\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} \end{pmatrix}$$

$$R_{c_2} = \begin{pmatrix} R_{1c_2} \\ \vdots \\ R_{Nc_2} \end{pmatrix}, \; R_{ic_2} = \begin{pmatrix} -\dfrac{1}{4}\dfrac{\hat{\mu}_{i1}\left(3 + \hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} \\ \dfrac{1}{4}\dfrac{\hat{\mu}_{i1}\left(3 + \hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1} - \hat{\mu}_{i2}\left(3 + \hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} \\ \dfrac{1}{4}\dfrac{\hat{\mu}_{i2}\left(3 + \hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} \end{pmatrix}$$

3) The artificial regression equation

$$r = R_{\beta}\delta_1 + R_{\mu_1}\delta_2 + R_{\mu_2}\delta_3 + R_{c_1}\delta_4 + R_{c_2}\delta_5 + \varepsilon$$

is estimated via the Least Squares method, and the prediction

$$\hat{r} = R_{\beta}\hat{\delta}_1 + R_{\mu_1}\hat{\delta}_2 + R_{\mu_2}\hat{\delta}_3 + R_{c_1}\hat{\delta}_4 + R_{c_2}\hat{\delta}_5,$$

and the t-statistics t_{c1} and t_{c2} testing H_0: $\delta_4 = 0$ and H_0: $\delta_5 = 0$ are saved.

4) The LM statistic $\hat{r}'\hat{r}$ is computed, and the test results for the LM test and the two t-tests are evaluated using the traditional levels of significance ($\alpha = 1\%$, $\alpha = 5\%$, $\alpha = 10\%$).

5) We calculate two additional indicators for skewness and fat tails. The first indicator checks whether t_{c1} exceeds the critical value at a significance level of 5% *and* is in absolute value larger than 1.5 times the absolute value of t_{c2}. The factor 1.5 was derived from simulations in front. If both criteria are fulfilled we conclude that skewness is a reason for rejecting normality. The second indicator does it the other way around. It checks whether t_{c2} exceeds the critical value at a significance level of 5% *and* is in absolute value larger than 1.5 times the absolute value of t_{c1}. Again, if both criteria are fulfilled we conclude that fat tails are a reason for rejecting normality. Afterwards we evaluate whether decisions using these indicators comply with the 5% level.

Due to the specific structure of r the disturbances of the artificial regression are heteroscedastic. Therefore, we compute several versions of the t-statistics: the first uses the traditional standard errors, the second uses the robust standard errors of White, and the third the robust HC3 standard errors. Since the robust versions are asymptotic standard errors, their results should be better for N=2000 than for N=400.

In the baseline scenario the LM test should meet the nominal size at least for large sample sizes, whereas this is an open question for the t-tests. Since the baseline scenario is a rather symmetric scenario several sensitivity analyzes are carried out to avoid that the results are only valid in “nice” scenarios. The following variations are analyzed:

- y_i is no longer uniformly distributed. This is done by changing the values of the μ_m's ($m = 1, 2$). They are chosen as different combinations of quantiles of a $N(0, 2)$ distribution, so that each of the categories is the weakest and strongest category in at least one design. Rather extreme cases like μ_1 = 1% quantile are not considered because in these cases one category may remain empty for some replications in medium samples.
- The relationship between the “explained” variation (= the variation of x_i) and the “unexplained” variation (= the variation of u_i) is changed by varying $\sigma_u^2 = Var(u_i)$. The quantiles are then taken from a $N(0, 1+\sigma_u^2)$ distribution.

Fixing $E(x_i)$ is not restrictive because a shift of $E(x_i)$ is equivalent to a shift of both threshold parameters by the same value. We therefore don't vary $E(x_i)$. Since the alternative of the test is the class of Pearson distributions, the distributions in the simulation study should belong to this class. We distinguish three cases: a symmetric alternative with fat tails, i.e. only the assumption of a kurtosis of 3 is violated; a skew alternative with kurtosis 3, i.e. only the assumption of symmetry is violated; and a skew alternative with a kurtosis different from 3, where both parameters are different from those of a normal distribution. This case distinction allows us to check the power of the test for different types of alternatives. Furthermore, we can check whether t_{c1} and t_{c2} identify the origin of the violation of the normal distribution and, if so, what is the power of these tests. This is done using the test statistics themselves and the additional indicators derived from the test statistics.

Glewwe (1997) uses a t(5) and a lognormal distribution for his simulations. T-distributions belong to the system of Pearson distributions (cf. Johnson et.al. 1995, p. 363) and are, therefore, suitable candidates for alternative distributions. However, the lognormal distribution belongs to the system of Johnson distributions and not to the system of Pearson distributions (cf. Ord 1972, p. 37). Thus, we use a member of the Pearson system instead to test against a skew alternative with a kurtosis different from 3.

As representatives for the first case of a symmetric distribution with fat tails we use t(k) distributions. The kurtosis of a t(k) distribution depends directly on k. It is infinite for $k<5$, and it is equal to $3+6/(k-4)$ for $k \geq 5$ (Balakrishnan et.al. 2006, p. 8554). Table 1 shows the kurtosis for those k that we use in the simulation study:

k	3	5	7	9
kurtosis	infinite	9	5	4.2

table 1: kurtosis of different t(k) distributions

k=3 and k=5 represent symmetric distributions with very fat tails, one with a well-defined kurtosis and one with a theoretically infinite kurtosis. For k=9 the kurtosis is equal to the kurtosis of a logistic distribution (cf. Müller 1991, p. 561), which shapes ordered logit models. Thus, the simulation results for a t(9) distribution may indicate whether the difference to a logistic distribution can be detected by the test.

The t-distributed random numbers can be drawn directly in R, no transformation is needed. Since the expectation is zero for all k, the corresponding assumption for the disturbances is fulfilled automatically. The variance is equal to $k/(k-2)$ and thus, different from one. This affects our simulation in that we design the distribution of y_i via the quantiles of the distribution of y_i^*. Since the theoretical distribution is complicated we simulate the distribution and use the empirical quantiles of the simulated distribution.[2]

For the second type of an alternative distribution we need a skew alternative with kurtosis equal to 3. This is not a standard case. However, Ord (1972, p. 10) shows graphically, how the different types of Pearson distributions are characterized by different combinations of values of skewness and kurtosis. In his graph, β_1 is equal to squared skewness and β_2 is equal to kurtosis. It is easily seen that only type I distributions of the Pearson system allow skew distributions with a kurtosis equal to 3. The values of the skewness are limited by the condition $\beta_2-\beta_1-1 \geq 0$. This implies a skewness smaller than $\sqrt{2}$ in absolute value if the kurtosis is equal to 3. Thus, huge values of the skewness are not possible, but a skewness of $\pm\sqrt{2}$ is still substantially different from a symmetric distribution.

Pearson distributions of type I belong to the class of beta type I distributions (cf. Nagahara 2004, p. 7, Johnson et.al. 1995, p. 210). Beta distributions can also be drawn with R. However, Nagahara (2004, pp. 7+13) gives an alternative way to draw these random numbers. We use his way because his procedure allows us to set the mean to 0, the variance to 1 (or an alternative value), the kurtosis to 3 and the skewness to values between $-\sqrt{2}$ and $\sqrt{2}$. Afterwards, the parameters of the distribution are calculated automatically, and the random numbers are drawn using these parameters. To be more precise: Given a mean of 0 and a kurtosis of 3 and var being the variance of u, the steps of the procedure are:

$$r = 2\left(2-\text{skewness}^2\right)/\text{skewness}^2$$

[2] Since x_i is still drawn from a normal distribution, the distribution of y_i^* is a sum of two random variables with different types of distributions. The theoretical distribution of this sum is not easy to determine. Thus, we use quantiles of the simulated distribution of y_i^*. For this purpose, we draw 100 million random numbers of each single distribution, calculate the sum of both, and use the empirical quantiles of this sum.

$$r_3 = 0.5r + 0.5r(r+2)\sqrt{\text{skewness}^2 \big/ \left(\text{skewness}^2 (r+2)^2 + 16(r+1)\right)}$$

$$r_4 = 0.5r - 0.5r(r+2)\sqrt{\text{skewness}^2 \big/ \left(\text{skewness}^2 (r+2)^2 + 16(r+1)\right)}$$

$$q = \max[r_3,\, r_4] \text{ if skewness} > 0 \text{ and } \min[r_3,\, r_4] \text{ if skewness} < 0$$

$$p = \min[r_3,\, r_4] \text{ if skewness} > 0 \text{ and } \max[r_3,\, r_4] \text{ if skewness} < 0$$

$$b = (p+q)\sqrt{\text{var}\cdot(p+q+1)/(pq)}$$

$$a = -bp/(p+q)$$

$$\text{draw } X \sim \text{Gamma}(p) \text{ and } Y \sim \text{Gamma}(q)$$

$$u = b\left(X/(X+Y)\right) + a \text{ (Nagahara 2004, pp. 7+13)}$$

p and q are the parameters of the Pearson I (beta I) distribution of u, a and b define the support of u, i.e. u varies within the interval [a, a+b] (Nagahara 2004, p. 7). Different values of skewness lead to the following parameters and supports:

skewness	−1.4	−1	−0.75	−0.5	0.5	0.75	1	1.4
r	0.04	2	5.11	14	14	5.11	2	0.04
q	0.009	0.5	1.34	3.79	10.21	3.77	1.5	0.03
p	0.03	1.5	3.77	10.21	3.79	1.34	0.5	0.009
a	−1.96	−3	−4.14	−6.36	−2.36	−1.48	−1	−0.53
a+b	0.53	1	1.48	2.36	6.36	4.14	3	1.96

Table 2: Parameters and support of different Pearson I distributions

It can be seen that the distribution behaves symmetrically around 0 for different values of the skewness. Furthermore, the support is the smaller, the larger the skewness is in absolute value.

The Pearson I distribution could be also used for our third case, a skew distribution with kurtosis different from 3. However, less restricted is the class of Pearson III distributions. For a kurtosis > 3 this distribution is also known as gamma distribution (cf. Nagahara 2004, p. 9). The standard version of a gamma distribution is right skewed (cf. Müller 1991, p. 557), which may be more relevant for applications than a left skewed distribution. If a left skewed distribution is needed a random variable from a standard gamma distribution can just be multiplied with −1. Furthermore, the standard version of a gamma distribution has a strictly positive support (ibid.). This would contradict our assumption of a zero mean. Thus, after drawing the random numbers from a gamma distribution, the mean has to be subtracted. The relation between skewness and kurtosis of a gamma distribution is defined by the equation $2\beta_2-3\beta_1-6=0$, again β_1 equal to squared skewness and β_2 equal to kurtosis (Nagahara 2004, p. 9, Ord 1972, p. 10).[3] Table 3 shows some pairs of skewness and kurtosis resulting from the equation above:

[3] The smallest possible value of the kurtosis is 3, thus, the normal distribution is a kind of lower bound of Pearson III distributions in terms of the kurtosis (cf. Ord 1972, p. 10).

skewness	−1	0.5	0.75	1	1.5	2	3	4	5	6
kurtosis	4.5	3.38	3.84	4.5	6.38	9	16.5	27	40.5	57

Table 3: Skewness and kurtosis of the selected gamma distributions

Again, using the formulas of Nagahara (2004) allows us to fix mean and variance of the disturbances and calculate the parameter of the gamma distribution depending on our choice of the skewness. Again, var being the variance of u, the steps of the procedure are:

$$p = 4/\text{skewness}^2$$

$$b = \sqrt{\text{var}/p}$$

$$\text{draw } X \sim \text{Gamma}(p)$$

$u = bX - pb$ (Nagahara 2010, pp. 9+13, with $\alpha = p$, $\beta = b$, and $\gamma = -pb$).

All simulations were done for an ordered probit model with 3 categories. Additionally, a robustness check was done for an ordered probit model with 4 categories:

$$y_i^* = \beta x_i + u_i, \quad i = 1, \ldots, N,$$

$$y_i = \begin{cases} 1, & \text{if } y_i^* \le \mu_1 \\ 2, & \text{if } \mu_1 < y_i^* \le \mu_2 \\ 3, & \text{if } \mu_2 < y_i^* \le \mu_3 \\ 4, & \text{if } \mu_3 < y_i^* \end{cases} \tag{3.1}$$

4. Results of the simulation study

4.1 Compliance with the nominal size under normality

In all designs of this chapter we assume that the error terms are normally distributed. Thus, the LM test should reject H_0 only for α percent of the replications. The same should happen for t_{c1} and t_{c2} if these test statistics are informative for skewness and fat tails. In our baseline scenario no size distortions are observed for both sample sizes. However, the nominal size is less well achieved for N=400, indicating a lower power of the test for this sample size. Table 4 shows the share of rejections for 5000 replications, the corresponding threshold values can be found in appendix 2:

N	$u_i \sim$	q1	q2	LM 1%	LM 5%	LM 10%	t_{c1} 1%	t_{c1} 5%	t_{c1} 10%	t_{c2} 1%	t_{c2} 5%	t_{c2} 10%
2000	N(0,1)	0.33	0.67	0.0122	0.0524	0.1006	0.0108	0.049	0.1022	0.0098	0.0444	0.0908
400	N(0,1)	0.33	0.67	0.0144	0.0424	0.0832	0.007	0.045	0.0912	0.008	0.038	0.0798

table 4: share of rejections, baseline scenario, q1 = quantile determining μ_1, q2 = quantile determining μ_2

First of all, we vary the distribution of y_i and assume a highly uneven distribution across the categories. We consider a narrow middle class (quantiles 0.45 and 0.55) and distributions that are dominated by one of the categories (quantiles 0.1/0.9, 0.8/0.9, and 0.1/0.2). The results are rather similar to those of the baseline scenario, although concerning the power some variation can be observed for the t-statistics (see table 5).

N	u_i ~	q1	q2	LM 1%	LM 5%	LM 10%	t_{c1} 1%	t_{c1} 5%	t_{c1} 10%	t_{c2} 1%	t_{c2} 5%	t_{c2} 10%
2000	N(0,1)	0.45	0.55	0.011	0.0478	0.0968	0.01	0.0488	0.0978	0.0096	0.0502	0.0946
		0.10	0.90	0.0104	0.0494	0.0946	0.0086	0.0472	0.0926	0.0058	0.047	0.0936
		0.80	0.90	0.014	0.0492	0.0956	0.01	0.0508	0.096	0.0072	0.0422	0.0884
		0.10	0.20	0.0136	0.0484	0.093	0.0124	0.0574	0.1076	0.008	0.0496	0.0964
400	N(0,1)	0.45	0.55	0.0168	0.0436	0.0798	0.0078	0.0444	0.091	0.009	0.037	0.0738
		0.10	0.90	0.017	0.0478	0.0824	0.0068	0.0426	0.0918	0.0034	0.0318	0.076
		0.80	0.90	0.0154	0.0448	0.085	0.0144	0.0556	0.107	0.0054	0.0364	0.0904
		0.10	0.20	0.016	0.0488	0.0836	0.0172	0.0542	0.099	0.007	0.0404	0.0856

table 5: share of rejections, uneven distributions of y_i, q1 = quantile determining μ_1, q2 = quantile determining μ_2

Furthermore, we vary the variance of u_i using σ^2 = 4, 10, 0.25, and 0.1 (see table 6). For the LM statistic most variations lead to similar results as in the baseline scenario, whereas some results for a very small variance of u_i met the asymptotic distribution less well – especially for N=400: Here some size distortion is observed for the 1 % level, whereas the shares of rejections are lower for the 10 % level. For the t_{c1} statistic even for N=2000 moderate size distortions are observed if the variance of u_i is very small, whereas in this case for the t_{c2} statistic for N=400 the share of rejections is much lower than the nominal size. The same results can be observed if the variance ratio is changed by the variance of x instead of the variance of u. E.g., choosing $Var(x_i)=0.25$ and $Var(u_i)=1$ leads to nearly the same results as $Var(x_i)=1$ and $Var(u_i)=4$. This is due to the normalization during the ML estimation of the probit model which compensates for any multiplication of the model equation with a constant factor.

N	u_i ~	q1	q2	LM 1%	LM 5%	LM 10%	t_{c1} 1%	t_{c1} 5%	t_{c1} 10%	t_{c2} 1%	t_{c2} 5%	t_{c2} 10%
2000	N(0,4)	0.33	0.67	0.0102	0.0516	0.101	0.0114	0.0508	0.099	0.0092	0.0504	0.0992
	N(0,10)			0.0094	0.0482	0.103	0.0102	0.0522	0.1028	0.0102	0.0476	0.096
	N(0,0.25)			0.0144	0.0476	0.0916	0.0102	0.0544	0.1046	0.0076	0.0476	0.1004
	N(0,0.1)			0.016	0.0478	0.0868	0.0162	0.0696	0.129	0.009	0.0522	0.1094
400	N(0,4)	0.33	0.67	0.0112	0.0506	0.0998	0.0084	0.0476	0.0938	0.0118	0.0536	0.1008
	N(0,10)			0.0082	0.0474	0.0948	0.0078	0.0456	0.089	0.0078	0.0464	0.1002
	N(0,0.25)			0.0188	0.0408	0.066	0.0054	0.0426	0.0982	0.0032	0.0312	0.0798
	N(0,0.1)			0.0222	0.043	0.067	0.0066	0.0514	0.1142	0.001	0.0188	0.0626

table 6: share of rejections, different variances of x_i and u_i, q1 = quantile determining μ_1, q2 = quantile determining μ_2

Summing up, both types of variation influence the results in some way, and the results for the t-statistics are slightly worse than for the LM-statistics. Therefore, we use both types of variation also for analyzing the power under alternative distributions.

4.2. Power under alternative distributions

a) Testing against fat tails

We present the results for t distributions with 3 and 9 degrees of freedom (df). The results for 5 and 7 df were between the results presented here and are therefore omitted. Since the degree of freedom fixes the variance of the error term, we vary the variance ratio by changing the variance of x. Since the variance of u is equal to df/(df–2) (Müller 1991, p. 554), the scenarios with a standard normal distribution of x_i now represent a variance ratio unequal to 1. We use the theoretical variance of a t(3) and a t(9) distribution to reproduce the same variance ratios as in ch. 4.1, e.g. for a t(3) distribution a variance of x_i equal to 3 now corresponds to a variance ratio of 1, or a variance of x_i equal to 30 now corresponds to a variance ratio of 10. The results are summarized in table 7 (t(3) distribution) and table 8 (t(9) distribution).

As expected, the greater the kurtosis and the larger the sample size are, the greater is the power of the LM test. This applies to all designs. However, the magnitude of the power varies. Concerning the distribution of y_i the power is reduced considerably if the middle category is only weakly filled. Concerning the variance ratio, a small variance of x_i reduces the power considerably. For the t(9) distribution with a sample size of N=400, the test is even close to be biased. Both weaknesses are plausible. However, if the (empirical) kurtosis and the sample size are large, the LM test shows a very large power for most designs, i.e. the asymptotic properties are excellent for most designs.

For the t_{c2} statistic the results are more heterogenous. As with the LM test the power increases for most designs with increasing (empirical) kurtosis and increasing sample size. Whereas the magnitude of the power is similar to the power of the LM test for some distributions of y_i, it is smaller if the middle category *and* an outer category are only weakly filled. Furthermore, the power of the t_{c2} test is only similar to that of the LM statistic if the variances of x_i and u_i differ not too much, whereas it is smaller if the differences between the variances are large. Thus, the t_{c2} statistic detects the deviation from a normal distribution less powerful than the LM statistic for some designs. Nevertheless, if the (empirical) kurtosis and the sample size are large, the t_{c2} statistic shows a large power, and for all designs rejecting its null hypothesis is a valid indicator for fat tails.

Unfortunately, for the t_{c1}-statistic size distortions appear for many designs and both sample sizes. They are larger for a large (empirical) kurtosis, i.e. there seems to be some co-movement of the two t-statistics. Thus, the t_{c1}-statistic is not a valid indicator for a missing skewness, and both t-statistics together cannot be used directly to decide whether (only) fat tails occur. The size distortions of the t_{c1}-statistic can be reduced by using White or HC3 standard errors. However, the robust standard errors also reduce the power of the t_{c2}-statistic so that this test is even biased for the t(9) distribution and N=400 for some designs. E.g., if $x_i \sim N(0,12.9)$, the shares of rejection using HC3 standard errors are 0.0016, 0.0268, and 0.0776 for the 1%, 5%, and 10% level. Thus, the problem of size distortion of the t_{c1} test cannot be solved by using robust standard errors due to a resulting bias for the t_{c2}-statistic for some cases.

N	x_i ~	q1	q2	LM 1%	LM 5%	LM 10%	t_{c1} 1%	t_{c1} 5%	t_{c1} 10%	t_{c2} 1%	t_{c2} 5%	t_{c2} 10%
2000	N(0,1)	0.33	0.67	0.8278	0.9192	0.9526	0.0326	0.1116	0.1766	0.8686	0.9502	0.9722
		0.45	0.55	0.4724	0.6524	0.7392	0.0298	0.098	0.1684	0.5288	0.7216	0.802
		0.10	0.90	0.864	0.9414	0.968	0.0306	0.0904	0.1574	0.8934	0.9632	0.9844
		0.80	0.90	0.9568	0.9852	0.995	0.0228	0.0856	0.1506	0.2864	0.4628	0.559
		0.10	0.20	0.9546	0.9826	0.9924	0.0164	0.0668	0.1264	0.3096	0.4992	0.601
	N(0,3)	0.33	0.67	0.996	0.9986	0.9994	0.0368	0.1004	0.1698	0.9464	0.9688	0.979
	N(0,0.75)			0.6772	0.8256	0.8862	0.0286	0.0904	0.1526	0.7382	0.8788	0.9284
	N(0,0.3)			0.1926	0.3694	0.4902	0.013	0.0628	0.1204	0.2426	0.4474	0.5638
	N(0,12)			0.999	0.9994	0.9998	0.005	0.0208	0.0368	0.4518	0.545	0.5942
	N(0,30)			0.9998	1	1	0.0034	0.011	0.0218	0.2514	0.3262	0.3596
400	N(0,1)	0.33	0.67	0.2386	0.3746	0.4638	0.0316	0.1034	0.167	0.2294	0.3928	0.4792
		0.45	0.55	0.1328	0.2346	0.3118	0.0276	0.096	0.1626	0.127	0.2374	0.3128
		0.10	0.90	0.2472	0.3912	0.4824	0.0254	0.0904	0.1568	0.2138	0.391	0.4944
		0.80	0.90	0.3156	0.475	0.5758	0.0206	0.0798	0.1442	0.0718	0.154	0.2266
		0.10	0.20	0.3114	0.4704	0.5682	0.0096	0.0582	0.1098	0.0644	0.157	0.2304
	N(0,3)	0.33	0.67	0.5326	0.6564	0.7162	0.0414	0.1222	0.1968	0.4218	0.619	0.709
	N(0,0.75)			0.1696	0.3012	0.391	0.0262	0.0942	0.1572	0.1682	0.3178	0.4082
	N(0,0.3)			0.0448	0.1198	0.1932	0.0172	0.0658	0.125	0.0474	0.1302	0.2056
	N(0,12)			0.7284	0.8026	0.8472	0.0308	0.101	0.1654	0.3454	0.5674	0.6736
	N(0,30)			0.7432	0.7972	0.8314	0.033	0.1106	0.176	0.2666	0.4856	0.5904

table 7: share of rejections, t(3) distribution, q1 = quantile determining μ_1, q2 = quantile determining μ_2

N	x_i ~	q1	q2	LM 1%	LM 5%	LM 10%	t_{c1} 1%	t_{c1} 5%	t_{c1} 10%	t_{c2} 1%	t_{c2} 5%	t_{c2} 10%
2000	N(0,1)	0.33	0.67	0.1768	0.3232	0.4198	0.0178	0.0652	0.1264	0.1802	0.3522	0.4666
		0.45	0.55	0.096	0.1978	0.2738	0.0174	0.0702	0.1246	0.0996	0.2106	0.2976
		0.10	0.90	0.2786	0.4382	0.5424	0.0176	0.0718	0.128	0.2234	0.4326	0.5554
		0.80	0.90	0.2602	0.4318	0.5324	0.0084	0.0462	0.096	0.0604	0.1652	0.2468
		0.10	0.20	0.2706	0.4268	0.5348	0.01	0.0426	0.098	0.0666	0.176	0.2634
	N(0,1.29)	0.33	0.67	0.2354	0.3922	0.48	0.02	0.0746	0.126	0.2346	0.4208	0.5184
	N(0,0.32)			0.0438	0.1182	0.1826	0.0124	0.058	0.1086	0.0494	0.1248	0.2002
	N(0,0.13)			0.0138	0.062	0.1156	0.0096	0.046	0.0936	0.0174	0.0688	0.1278
	N(0,5.14)			0.4898	0.6278	0.7034	0.017	0.069	0.1264	0.3068	0.5326	0.6486
	N(0,12.9)			0.4926	0.6208	0.69	0.0224	0.075	0.144	0.2628	0.4744	0.594
400	N(0,1)	0.33	0.67	0.0736	0.1376	0.1866	0.0158	0.07	0.1252	0.0508	0.1172	0.1748
		0.45	0.55	0.05	0.1022	0.155	0.0144	0.0652	0.1186	0.0416	0.0964	0.1444
		0.10	0.90	0.0964	0.1754	0.2374	0.0126	0.0642	0.1258	0.033	0.1144	0.1928
		0.80	0.90	0.0768	0.143	0.1992	0.0074	0.0452	0.0972	0.0134	0.0664	0.128
		0.10	0.20	0.0882	0.1602	0.2234	0.0082	0.0462	0.0952	0.0192	0.0728	0.1416
	N(0,1.29)	0.33	0.67	0.0848	0.1532	0.2132	0.0174	0.0696	0.13	0.0532	0.1312	0.1954
	N(0,0.32)			0.0204	0.0684	0.1182	0.0118	0.0526	0.105	0.021	0.0644	0.1202
	N(0,0.13)			0.0084	0.0522	0.1026	0.011	0.052	0.0976	0.0078	0.0506	0.1018
	N(0,5.14)			0.155	0.232	0.294	0.0158	0.0804	0.1444	0.04	0.1374	0.218
	N(0,12.9)			0.185	0.2536	0.3008	0.0216	0.092	0.1758	0.029	0.1248	0.2178

table 8: share of rejections, t(9) distribution, q1 = quantile determining μ_1, q2 = quantile determining μ_2

b) Testing against skewness

We present the results for Pearson 1 (P1) distributions with a large skewness of 1 and a moderate skewness of 0.5. The results for −1 and −0.5 are quite similar (with the results for 0.80/0.90 now being similar to the results for 0.10/0.20 and vice versa) and are therefore omitted. The variance of u is normalized to 1 so that the ratio of explained and unexplained variance is equal to one the baseline scenario, and can be varied using the same factors as in chapter 4.1.

The power is very large for both, the LM and the t_{c1}-statistic. It is equal to one if skewness and sample size are large. For a moderate skewness or a medium sample size the power is still large, but with certain differences between the designs. It is substantially lower if the middle category *and* the outer category on the skewed tail are only weakly filled. Nevertheless, both tests are good indicators for skewness with the t_{c1}-statistic being even slightly better.

Concerning the t_{c2}-statistic there are notable size distortions so that the t_{c2}-statistic cannot be used to detect missing fat tails. Again, there seems to be some co-movement between the t-statistics because the size distortions for the kurtosis increase if skewness increases. With a skewness of 1 the shares of rejections for designs in which the middle category *and* an outer category of y_i are only weakly filled are even close to or equal to 1. Again, this problem cannot be solved by using robust standard errors. E.g., for thresholds of 0.8 and 0.9 the shares of rejection at the 5% and the 10% level are still 1 for both types of robust standard errors. Thus, the robust standard errors are useless in this case.

N	**x_i ~**	**q1**	**q2**	**LM 1%**	**LM 5%**	**LM 10%**	**t_{c1} 1%**	**t_{c1} 5%**	**t_{c1} 10%**	**t_{c2} 1%**	**t_{c2} 5%**	**t_{c2} 10%**
2000	N(0,1)	0.33	0.67	1	1	1	1	1	1	0.0752	0.183	0.2682
		0.45	0.55	1	1	1	1	1	1	0.0532	0.1296	0.197
		0.10	0.90	1	1	1	1	1	1	0.0164	0.0758	0.1406
		0.80	0.90	0.9944	0.9998	1	0.9982	1	1	0.9986	1	1
		0.10	0.20	1	1	1	1	1	1	0.9946	0.9978	0.9992
	N(0,0.25)	0.33	0.67	1	1	1	1	1	1	0.265	0.4768	0.5904
	N(0,0.1)			1	1	1	1	1	1	0.8532	0.9496	0.9736
	N(0,4)			1	1	1	1	1	1	0.014	0.063	0.1226
	N(0,10)			1	1	1	1	1	1	0.0078	0.0478	0.0986
400	N(0,1)	0.33	0.67	0.9998	1	1	1	1	1	0.022	0.0816	0.1412
		0.45	0.55	0.986	0.9998	1	0.9906	0.9982	0.9996	0.0276	0.082	0.1408
		0.10	0.90	0.8348	0.9866	0.9978	0.9586	0.9938	0.9984	0.0044	0.045	0.0946
		0.80	0.90	0.1192	0.3634	0.5352	0.268	0.5676	0.7102	0.2516	0.5806	0.7312
		0.10	0.20	0.9798	0.999	0.9996	0.9754	0.9894	0.9948	0.594	0.6938	0.7468
	N(0,0.25)	0.33	0.67	1	1	1	0.9996	0.9998	1	0.05	0.1622	0.26
	N(0,0.1)			0.9982	1	1	0.9858	0.9972	0.9982	0.176	0.3788	0.5092
	N(0,4)			0.9132	0.993	0.9994	0.9826	0.9984	0.9996	0.0052	0.0374	0.0872
	N(0,10)			0.3572	0.7388	0.9038	0.8358	0.9646	0.9862	0.003	0.0256	0.0706

table 9: share of rejections, P1 distribution, s = 1, q1 = quantile determining μ_1, q2 = quantile determining μ_2

N	$x_i \sim$	q1	q2	LM 1%	LM 5%	LM 10%	t_{c1} 1%	t_{c1} 5%	t_{c1} 10%	t_{c2} 1%	t_{c2} 5%	t_{c2} 10%
2000	N(0,1)	0.33	0.67	0.944	0.9918	0.9976	0.9596	0.9922	0.997	0.021	0.0972	0.1722
		0.45	0.55	0.652	0.8668	0.932	0.7258	0.8936	0.9486	0.0158	0.0918	0.1704
		0.10	0.90	0.9656	0.996	0.9988	0.9792	0.997	0.9984	0.0122	0.0694	0.1346
		0.80	0.90	0.092	0.2558	0.3726	0.08	0.2342	0.3524	0.0272	0.1048	0.1812
		0.10	0.20	0.881	0.9764	0.9918	0.4922	0.6738	0.7568	0.035	0.1016	0.1684
	N(0,0.25)	0.33	0.67	0.5292	0.767	0.8554	0.6084	0.8172	0.888	0.0232	0.0974	0.1732
	N(0,0.1)			0.1714	0.3806	0.5116	0.229	0.456	0.587	0.0156	0.067	0.1254
	N(0,4)			0.9394	0.9936	0.999	0.966	0.9928	0.9976	0.0138	0.0648	0.1208
	N(0,10)			0.7416	0.9338	0.9744	0.8944	0.9678	0.985	0.015	0.0732	0.15
400	N(0,1)	0.33	0.67	0.144	0.3336	0.4768	0.2104	0.4598	0.5986	0.0076	0.0474	0.1102
		0.45	0.55	0.0734	0.1804	0.284	0.0962	0.2592	0.3838	0.0122	0.0484	0.099
		0.10	0.90	0.1324	0.303	0.4418	0.189	0.4582	0.6124	0.0048	0.0494	0.1136
		0.80	0.90	0.0224	0.0696	0.122	0.0108	0.0592	0.133	0.0048	0.0414	0.0964
		0.10	0.20	0.1148	0.2526	0.3692	0.1574	0.2696	0.3348	0.0356	0.084	0.1358
	N(0,0.25)	0.33	0.67	0.055	0.1806	0.2922	0.0834	0.2372	0.3548	0.0088	0.0506	0.1094
	N(0,0.1)			0.0284	0.1072	0.1854	0.0354	0.1308	0.2178	0.011	0.0554	0.1082
	N(0,4)			0.1382	0.3002	0.4368	0.2148	0.491	0.6404	0.0042	0.0408	0.1028
	N(0,10)			0.0868	0.175	0.2458	0.117	0.338	0.483	0.0018	0.0288	0.0914

table 10: share of rejections, P1 distribution, s = 0.5, q1 = quantile determining μ_1, q2 = quantile determining μ_2

Summing up, the results for skewness are stronger than those for fat tails: The power of the LM test is much higher. The power of the corresponding t-statistic is also much higher, and for some designs it is even a stronger indicator than the LM test. The problem of size distortion for the "wrong" t-statistic is also much higher, and the robust standard errors are less effective in reducing the size distortion. Therefore, the picture is mixed: On the one hand, skewness and fat tails are detected well by the corresponding t-statistic, on the other hand, there are size distortions for the non-corresponding t-statistic. Thus, statistical significance of the t-statistics cannot be used as the sole criterion for determining skewness or fat tails. It must be supplemented by a second criterion.

c) Testing against a skew distribution with fat tails

We present the results for a $\Gamma(4)$ distribution, i.e. a distribution with skewness 1 and kurtosis 4.5 (cf. table 3). The results for this distribution can be directly compared with the results presented above, i.e. they can be compared with a P(1) distribution with skewness 1 and a t(9) distribution, the kurtosis of which is close to 4.5 (cf. table 1). Thus, it can be seen whether the appearance of a second problem sharpens or weakens the results of the different test statistics. The results are shown in table 11.

The power of the LM test is very high for a large sample size. The reduction of the power due to a lower sample size varies between the different designs. The highest reduction is observed for an uneven distribution of y_i with the thresholds 0.8 and 0.9, and for a very small variance of x_i. The results for the t_{c1}-statistic are similar except a stronger power reduction for the 0.8/ 0.9 case. The power of the t_{c2}-statistic is much lower than that of the two other statistics, i.e. it is more difficult to detect the fat tails of the $\Gamma(4)$ distribution than its skewness.

N	$x_i \sim$	q1	q2	LM 1%	LM 5%	LM 10%	t_{c1} 1%	t_{c1} 5%	t_{c1} 10%	t_{c2} 1%	t_{c2} 5%	t_{c2} 10%
2000	N(0,1)	0.33	0.67	1	1	1	0.9984	1	1	0.1314	0.2738	0.3702
		0.45	0.55	0.9978	1	1	0.999	0.9998	1	0.0762	0.1642	0.2278
		0.10	0.90	1	1	1	0.9796	0.9894	0.9928	0.2416	0.464	0.5928
		0.80	0.90	0.8572	0.9548	0.9786	0.2372	0.4644	0.5844	0.024	0.093	0.1672
		0.10	0.20	0.9966	0.9998	1	0.9224	0.964	0.979	0.2384	0.3614	0.4358
	N(0,0.25)	0.33	0.67	0.9934	0.9992	1	0.9966	0.9996	1	0.0202	0.0748	0.1406
	N(0,0.1)			0.75	0.8954	0.9456	0.8166	0.9356	0.966	0.0208	0.077	0.1418
	N(0,4)			1	1	1	0.9006	0.9308	0.9422	0.3364	0.5684	0.682
	N(0,10)			0.9996	1	1	0.7888	0.8392	0.8672	0.3182	0.5326	0.6396
400	N(0,1)	0.33	0.67	0.6018	0.8182	0.9038	0.711	0.901	0.9484	0.0456	0.1208	0.1824
		0.45	0.55	0.3094	0.5358	0.6718	0.4076	0.6662	0.7876	0.0442	0.1006	0.152
		0.10	0.90	0.5944	0.8064	0.8952	0.6624	0.885	0.9414	0.0292	0.1242	0.2138
		0.80	0.90	0.1614	0.3252	0.4406	0.0334	0.1406	0.2346	0.0128	0.0684	0.135
		0.10	0.20	0.3098	0.4978	0.621	0.379	0.5154	0.5896	0.1132	0.1852	0.245
	N(0,0.25)	0.33	0.67	0.274	0.5216	0.6546	0.3696	0.622	0.7408	0.0108	0.0592	0.1136
	N(0,0.1)			0.1108	0.2676	0.386	0.144	0.3378	0.4572	0.0124	0.0538	0.109
	N(0,4)			0.6294	0.802	0.8826	0.6388	0.8506	0.9092	0.0468	0.1588	0.2584
	N(0,10)			0.4856	0.6364	0.72	0.4362	0.7014	0.8028	0.0486	0.1626	0.2778

table 11: share of rejections, $\Gamma(4)$ distribution, q1 = quantile determining μ_1, q2 = quantile determining μ_2

Comparing the results for the $\Gamma(4)$ distribution with those of a P(1) distribution with the same skewness (cf. table 9) shows a weaker power of the LM statistic and the t_{c1}-statistic for the most cases, i.e. the additional fat tails weaken the power of the two tests. If the results are compared with those of a t(9) distribution with a similar kurtosis (cf. table 8) the power of the LM statistic is much higher, whereas the t_{c2} statistic shows a mixed picture. I.e. the additional skewness does not systematically change the results for the t_{c2}-statistic. For some designs its power is still weak. However, the test always remains at least unbiased.

Summing up, additional skewness increases the power of the LM statistic a lot, whereas additional fat tails can reduce it. Furthermore, additional skewness does not systematically change the results for the t_{c2} statistic, whereas the results for the t_{c1}-statistic are weakened by additional fat tails.

4.3. Double indicators for testing against skewness and fat tails

Ch. 4.2 shows that the single use of the statistics t_{c1} and t_{c2} lead to size-distorted test decisions. However, additional calculations (not reported here) indicate, that in case of pure skewness or pure fat tails the test-statistic measuring the "true" problem is often larger than the test-statistic measuring the "wrong" problem. Thus, we use the ratio of the test statistics in absolute values as an indicator of which of the two problems is relevant. We find that a ratio between 1.5 and 2 is often enough to avoid wrong decisions for more than α% of the simulations, α being the nominal size of the tests. However, in case of normally distributed errors using only the ratio leads to too much "rejections" of symmetry or a kurtosis of 3. Therefore, we combine both to a double indicator, i.e. we check whether the test-statistic rejects the null at the nominal size of 5% *and* is more than one and a half times the other. To be more precise: If the indicator SI = "$|t_{c1}| \geq 1.96$ AND $|t_{c1}| \geq 1.5|t_{c2}|$" is fulfilled we conclude that the error terms are skew distributed, and if the indicator FTI = "$|t_{c2}| \geq 1.96$ AND $|t_{c2}| \geq 1.5|t_{c1}|$" is fulfilled we conclude that the error terms are distributed with fat tails. Since both results are mutually exclusive, we can only conclude that skewness or fat tails matters, we cannot conclude that only skewness or only fat tails matters. In our simulation we check whether using the double indicator avoids the size distortions observed in ch. 4.2 and how the power of t_{c1} and t_{c2} is influenced by using the additional ratio. In addition to the baseline scenario, we are focusing on those designs that have shown problems in some of the previous simulations. We compare the share of rejections for the t-statistics at the 5% level with the share of rejections using the double indicators. According to the formulas for the double indicators the second share is always smaller than the first one. Ideally, using the double indicators reduces the power only slightly, while the size distortions are eliminated completely.

If the error terms are normally distributed the second indicator reduces the power only slightly except the case of thresholds 0.8 and 0.9. I.e. if both outer categories are sufficiently filled, the results of the double indicator are nearly as good as the single test decisions. Size distortions cannot appear since the double indicator can only reduce the share of rejections, and even for the single test decisions no size distortions were observed for these designs.

N	$x_i \sim$	$u_i \sim$	q1	q2	t_{c1} 5%	SI 5%	t_{c2} 5%	FTI 5%
2000	N(0,1)	N(0,1)	0.33	0.67	0.049	0.043	0.0444	0.0382
			0.45	0.55	0.0488	0.0424	0.0502	0.0444
			0.80	0.90	0.0508	0.0130	0.0422	0.0074
	N(0,0.1)		0.33	0.67	0.052	0.045	0.048	0.0416
400	N(0,1)	N(0,1)	0.33	0.67	0.045	0.0388	0.038	0.0304
			0.45	0.55	0.0444	0.0346	0.037	0.0272
			0.80	0.90	0.0556	0.0170	0.0364	0.0042
	N(0,0.1)		0.33	0.67	0.0468	0.036	0.046	0.0378

table 12: share of rejections (t_{c1}, t_{c2}), share of ones (SI, FTI), normally distributed error terms, q1 = quantile determining μ_1, q2 = quantile determining μ_2

If the error terms are t-distributed, the double indicator eliminates all size distortions of the t_{c1}-statistic, i.e. the problem of the single test-statistic is solved. Furthermore, in most designs the power of the t_{c2} decision reduces only slightly. For a medium sample size and thresholds 0.8/ 0.9 the reduction is somewhat greater.

N	$x_i \sim$	$u_i \sim$	q1	q2	t_{c1} 5%	SI 5%	t_{c2} 5%	FTI 5%
2000	N(0,1)	t(3)	0.33	0.67	0.1116	0.0018	0.9502	0.9052
			0.45	0.55	0.098	0.0116	0.7216	0.6528
			0.80	0.90	0.0856	0.038	0.4628	0.4588
	N(0,0.3)		0.33	0.67	0.0628	0.0232	0.4474	0.4012
	N(0,1)	t(9)	0.33	0.67	0.0652	0.0264	0.3522	0.3082
			0.45	0.55	0.0702	0.0376	0.2106	0.1778
			0.80	0.90	0.0462	0.0130	0.1652	0.1244
	N(0,0.13)		0.33	0.67	0.046	0.0382	0.0688	0.0604
400	N(0,1)	t(3)	0.33	0.67	0.1034	0.0324	0.3928	0.3262
			0.45	0.55	0.096	0.0434	0.2374	0.1842
			0.80	0.90	0.0798	0.0168	0.154	0.0978
	N(0,0.3)		0.33	0.67	0.0658	0.045	0.1302	0.1084
	N(0,1)	t(9)	0.33	0.67	0.07	0.0446	0.1172	0.09
			0.45	0.55	0.0652	0.0410	0.0964	0.0716
			0.80	0.90	0.0452	0.0080	0.0664	0.0276
	N(0,0.13)		0.33	0.67	0.052	0.0428	0.0506	0.041

table 13: share of rejections (t_{c1}, t_{c2}), share of ones (SI, FTI), t-distributed error terms, q1 = quantile determining μ_1, q2 = quantile determining μ_2

If the error terms are P1-distributed, the double indicator eliminates all size distortions of the t_{c2} statistic, i.e. again the problem of the single test-statistic is solved. Furthermore, in three of the four designs the power of the t_{c1} decision reduces only slightly. For the thresholds 0.8/ 0.9, however, the power of the t_{c1} decision collapses so that the decision is mostly not even unbiased. A closer look reveals that for the thresholds 0.8/ 0.9 there is a strong co-movement of the two test statistics so that the second condition of our double indicator is no longer fulfilled for many replications. Further research is needed to clarify the reasons for this strong co-movement.

N	$x_i \sim$	$u_i \sim$	q1	q2	t_{c1} 5%	SI 5%	t_{c2} 5%	FTI 5%
2000	N(0,1)	P1, s=1	0.33	0.67	1	1	0.183	0
			0.45	0.55	1	1	0.1296	0
			0.80	0.90	1	0	1	0
	N(0,0.1)		0.33	0.67	1	0.9978	0.9496	0
	N(0,1)	P1, s=0.5	0.33	0.67	0.9922	0.9356	0.0972	0.0014
			0.45	0.55	0.8936	0.8164	0.0918	0.0124
			0.80	0.90	0.2342	0.08	0.1048	0.002
	N(0,0.1)		0.33	0.67	0.456	0.4096	0.067	0.0276
400	N(0,1)	P1, s=1	0.33	0.67	1	0.9974	0.0816	0
			0.45	0.55	0.9982	0.9782	0.082	0
			0.80	0.90	0.5676	0.0026	0.5806	0.0022
	N(0,0.1)		0.33	0.67	0.9972	0.8942	0.3788	0.0042
	N(0,1)	P1, s=0.5	0.33	0.67	0.4598	0.4216	0.0474	0.0172
			0.45	0.55	0.2592	0.2244	0.0484	0.0212
			0.80	0.90	0.0592	0.0094	0.0414	0.0058
	N(0,0.1)		0.33	0.67	0.1308	0.1116	0.0554	0.0386

table 14: share of rejections (t_{c1}, t_{c2}), share of ones (SI, FTI), P1 distributed error terms, q1 = quantile determining μ_1, q2 = quantile determining μ_2

If the error terms are $\Gamma(4)$ distributed, the power of the t_{c1} decision reduces only slightly for most cases. Again, for a medium sample size the thresholds 0.8/ 0.9 are an exception. In contrast to this, the power of the t_{c2} decision is equal or near to zero. This is not surprising since the second conditions of the double indicators are mutually exclusive, i.e. if the second condition is fulfilled for one indicator, it cannot be fulfilled for the other.

N	x_i ~	u_i ~	q1	q2	t_{c1} 5%	SI 5%	t_{c2} 5%	FTI 5%
2000	N(0,1)	Γ(4)	0.33	0.67	1	0.995	0.2738	0
			0.45	0.55	0.9998	0.9918	0.1642	0
			0.80	0.90	0.4644	0.3942	0.093	0.0260
	N(0,0.1)		0.33	0.67	0.9356	0.894	0.077	0.0024
400	N(0,1)	Γ(4)	0.33	0.67	0.901	0.8146	0.1208	0.0052
			0.45	0.55	0.6662	0.5750	0.1006	0.0094
			0.80	0.90	0.1406	0.0354	0.0684	0.0184
	N(0,0.1)		0.33	0.67	0.3378	0.2802	0.0538	0.0286

table 15: share of rejections (t_{c1}, t_{c2}), share of ones (SI, FTI), Γ(4) distributed error terms, q1 = quantile determining μ_1, q2 = quantile determining μ_2

Summing up, the double indicators remove size distortion completely. If only one of the two problems occurs, the power reduction compared to the single test decision is only moderate provided that both outer categories are sufficiently filled. If skewness and fat tails occur, again the power reduction is small for the more severe problem whereas the other is no longer detected due to the mutual exclusivity of the second part of the indicator.

4.4. Selected results for four categories

Using four categories instead of three implies that more "information" is available. Thus, the power of the tests should increase compared to similar scenarios in case of three categories. However, several questions remain: How large is this increase of power? Does more information reduce the observed size distortions of the single t-tests? Does the double indicator still work fine in most scenarios, i.e. can the results be generalized to more than three categories? Finally, are the problems of the indicator in the extreme scenario reduced?

To answer these questions, we consider the ordered probit model (3.1) with four observed categories We consider two scenarios. The first scenario corresponds to our baseline scenario and assumes equally distributed y_i's, i.e. μ_1/ μ_2/ μ_3 are chosen as 25%/ 50%/ 75% quantile of the distribution of y_i^*. The second scenario corresponds to the only scenario that caused severe problems for the double indicator, i.e. the scenario with thresholds 0.8/ 0.9, where many observations are concentrated in an outer category. The threshold values μ_1/ μ_2/ μ_3 are now chosen as 70%/ 80%/ 90% quantile of the distribution of y_i^*. Thus, on the one hand it is still a rather asymmetric distribution, on the other hand there is "more" information due to an additional category. In both scenarios x_i is assumed to be standard normally distributed, i.e. the column concerning the distribution of x_i is dropped in the tables. For the u_i's the same distributions and sample sizes as in ch. 4.1-4.3 are chosen so that the results can be compared directly. Table 16 contains the results for the different test statistics, table 17 the results for the double indicator.

Comparing table 16 with the results in the tables 7-11 shows that the additional information increases the power of both the LM test and the relevant t test across all distributions and both sample sizes – in some cases considerably. For the Γ(4) distribution concerning the t tests the increase is focused on the power of the t_{c1} test, whereas for the t_{c2} test only slight changes are observed. In contrast to the better results for the power, the size distortion is *not* reduced, i.e. this problem cannot be solved by increasing the number of categories. However, the double indicators still work reliably, i.e. the size distortion of the single test statistics is eliminated by using the double indicators. Furthermore, for most cases the power is only reduced slightly, and the relative loss of power is smaller than for the model with 3 categories. The only exception remains for the P1 distribution with high skewness, where for the design with many observations in an outer category the power is still close to 0, i.e. this problem is not solved by additional information. However, this is still an extreme scenario so that for more common data sets the indicator should work well.

N	**u_i ~**	**q1**	**q2**	**q3**	**LM 1%**	**LM 5%**	**LM 10%**	**t_{c1} 1%**	**t_{c1} 5%**	**t_{c1} 10%**	**t_{c2} 1%**	**t_{c2} 5%**	**t_{c2} 10%**
2000	N(0,1)	0.25	0.50	0.75	0.0102	0.0500	0.0948	0.0106	0.0502	0.0994	0.0086	0.0442	0.0952
		0.70	0.80	0.90	0.012	0.0512	0.0968	0.0112	0.0526	0.1024	0.011	0.0462	0.0942
	t(3)	0.25	0.50	0.75	0.981	0.9946	0.9968	0.0362	0.1152	0.1872	0.9894	0.997	0.9988
		0.70	0.80	0.90	0.992	0.9976	0.999	0.0136	0.0708	0.1306	0.673	0.833	0.8918
	t(9)	0.25	0.50	0.75	0.3116	0.483	0.5794	0.0198	0.076	0.1358	0.3268	0.528	0.6332
		0.70	0.80	0.90	0.3442	0.5248	0.6298	0.0082	0.0436	0.0906	0.119	0.269	0.3846
	P1, s=1	0.25	0.50	0.75	1	1	1	1	1	1	0.0918	0.2138	0.3068
		0.70	0.80	0.90	1	1	1	1	1	1	1	1	1
	P1, s=0.5	0.25	0.50	0.75	0.9978	1	1	0.998	1	1	0.0358	0.1284	0.2122
		0.70	0.80	0.90	0.307	0.5784	0.7024	0.2742	0.5508	0.682	0.0472	0.165	0.2696
	Γ(4)	0.25	0.50	0.75	1	1	1	0.9978	0.9986	0.9994	0.1924	0.3596	0.4644
		0.70	0.80	0.90	0.9938	0.9998	0.9998	0.5834	0.7974	0.8702	0.021	0.0898	0.1612
400	N(0,1)	0.25	0.50	0.75	0.0178	0.0510	0.0894	0.0084	0.0452	0.0936	0.0076	0.0434	0.087
		0.70	0.80	0.90	0.0164	0.0466	0.0862	0.0122	0.0554	0.1002	0.0078	0.0438	0.0918
	t(3)	0.25	0.50	0.75	0.4008	0.5726	0.654	0.0376	0.112	0.1904	0.4122	0.5934	0.69
		0.70	0.80	0.90	0.4358	0.605	0.6872	0.0098	0.0576	0.1138	0.1284	0.2628	0.355
	t(9)	0.25	0.50	0.75	0.0936	0.1828	0.2506	0.0166	0.0716	0.1314	0.065	0.1642	0.2426
		0.70	0.80	0.90	0.0878	0.1648	0.23	0.0094	0.042	0.0878	0.0232	0.0854	0.1494
	P1, s=1	0.25	0.50	0.75	1	1	1	1	1	1	0.0218	0.077	0.1382
		0.70	0.80	0.90	0.562	0.8408	0.9222	0.7618	0.9404	0.9752	0.5594	0.8454	0.9256
	P1, s=0.5	0.25	0.50	0.75	0.2378	0.5154	0.6808	0.36	0.643	0.7574	0.0102	0.055	0.1188
		0.70	0.80	0.90	0.0376	0.1146	0.1916	0.0212	0.1052	0.2046	0.0088	0.0546	0.1188
	Γ(4)	0.25	0.50	0.75	0.8262	0.9502	0.9806	0.8944	0.9684	0.985	0.0478	0.1344	0.2166
		0.70	0.80	0.90	0.2894	0.5116	0.6316	0.0704	0.2284	0.3498	0.013	0.074	0.1466

table 16: 4 categories, share of rejections, q1/ q2/ q3 = quantiles determining μ_1/ μ_2/ μ_3

N	u_i ~	q1	q2	q3	t_{c1} 5%	SI 5%	t_{c2} 5%	FTI 5%
2000	N(0,1)	0.25	0.50	0.75	0.0502	0.0444	0.0442	0.0390
		0.70	0.80	0.90	0.0526	0.0180	0.0462	0.0138
	t(3)	0.25	0.50	0.75	0.1152	0.0002	0.9970	0.9860
		0.70	0.80	0.90	0.0708	0.0166	0.8330	0.8304
	t(9)	0.25	0.50	0.75	0.0760	0.0198	0.5280	0.4702
		0.70	0.80	0.90	0.0436	0.0158	0.2690	0.2418
	P1, s=1	0.25	0.50	0.75	1	1	0.2138	0
		0.70	0.80	0.90	1	0.0042	1	0
	P1, s=0.5	0.25	0.50	0.75	1	0.9722	0.1284	0.0002
		0.70	0.80	0.90	0.5508	0.3486	0.1650	0.0024
	Γ(4)	0.25	0.50	0.75	0.9986	0.9982	0.3596	0
		0.70	0.80	0.90	0.7974	0.7818	0.0898	0.0292
400	N(0,1)	0.25	0.50	0.75	0.0452	0.0370	0.0434	0.0366
		0.70	0.80	0.90	0.0554	0.0212	0.0438	0.0112
	t(3)	0.25	0.50	0.75	0.112	0.0202	0.5934	0.5122
		0.70	0.80	0.90	0.0576	0.0208	0.2628	0.2422
	t(9)	0.25	0.50	0.75	0.0716	0.0446	0.1642	0.1318
		0.70	0.80	0.90	0.042	0.0118	0.0854	0.0528
	P1, s=1	0.25	0.50	0.75	1	0.9996	0.0770	0
		0.70	0.80	0.90	0.9404	0.0624	0.8454	0.0008
	P1, s=0.5	0.25	0.50	0.75	0.643	0.5975	0.055	0.0214
		0.70	0.80	0.90	0.1052	0.0376	0.0546	0.0074
	Γ(4)	0.25	0.50	0.75	0.9684	0.9080	0.1344	0.0018
		0.70	0.80	0.90	0.2284	0.1434	0.074	0.0262

table 17: 4 categories, share of rejections (t_{c1}, t_{c2}), share of ones (SI, FTI), q1/ q2/ q3 = quantiles determining μ_1/ μ_2/ μ_3

5. Conclusion

The paper derives an artificial regression version of the LM test for testing the distribution assumption for an ordered probit model with more than two categories, where normality is tested against the class of Pearson distributions as it is done by the Bera-Jarque-Lee test for binary probit models. The artificial regression for the ordered model is also only based on standard algebraic transformations and, therefore, simplifies the calculation of the test-statistic. However, compared to the binary model an additional step is needed that is known from operating with panel data, i.e. the cost reduction is smaller in the ordered model compared to the binary model.

The artificial regression again contains a regressor that is related to skewness and a regressor that is related to fat tails. An extensive simulations study for 3 categories, supplemented by selected simulations for 4 categories, shows that the power of detecting skewness or fat tails by the relevant t-statistic is high for most of the designs, being higher for skewness than for fat tails. I.e. the t-statistics contain additional information compared to the LM statistic. However, there are also noteworthy size distortions, i.e. the “wrong” t-statistic often also indicates a violation of symmetry or a violation of a kurtosis of 3 although there is none.

Therefore, the single test statistics can be misleading for concluding skewness or fat tails as the main reason for violating normality. We introduce two new double indicators. The indicator equals 1 if the test-statistic rejects H_0 at the 5% level *and* the relevant t-statistic is 1.5 times larger than the other t-statistic in absolute value. If the indicator equals 1 we conclude that a problem of skewness or a problem of fat tails is observed. The simulation study shows that these double indicators work well: The size distortion is eliminated completely, and for most designs the power is only reduced slightly compared to a decision only based on the test-statistic. Only if the distribution of the categorical variable is highly asymmetrical with a very high percentage of observations in a marginal category on the “wrong” side, severe problems with the power are observed. However, if the distribution of y_i is less extreme the double indicators work fine for 3 categories as well as for 4 categories. If skewness and fat tails appear simultaneously, mostly the indicator for skewness is activated, i.e. fat tails are not detected in this case.

References

Balakrishnan. N./ Read. C.B./ Vidakovic. B. (eds.) (2006). Encyclopedia of statistical sciences. vol. 13. 2nd ed. Hoboken (NJ): Wiley.

Bera. A.K./ Jarque. C.M./ Lee. L.F. (1984). Testing the normality assumption in limited dependent variable models. International Economic Review 25. pp. 563-578.

Davidson. R./ MacKinnon. J.G. (1993). Estimation and inference in econometrics. New York: Oxford University Press.

Glewwe. P. (1997). A test of the normality assumption in the ordered probit model. Econometric Reviews 16. 1-19.

Greene. W.H.. Hensher. D.A. (2010). Modelling ordered choices: A primer. Cambridge: Cambridge University Press.

Johnson. N.L./ Kotz. S./ Balakrishnan. N. (1995). Continuous univariate distributions vol. 2. 2nd ed.. New York: Wiley.

Johnson. P.A. (1996). A test of the normality assumption in the ordered probit model. Metron – international journal of statistics 54. 213-221.

Müller. P.H. (ed.) (1991). Lexikon der Stochastik. 5. Aufl.. Berlin: Akademie Verlag.

Murphy. A. (1994). Artificial regression based misspecification tests for discrete choice models. The Economics and Social Review 26. 69-74.

Nagahara. Y. (2004). A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation. Computational Statistics and Data Analysis 47. pp. 1-29.

Ord. J.K. (1972). Families of frequency distributions. London: Griffin.

Wilde. J. (2008). A simple representation of the Bera-Jarque-Lee test for probit models. Economics Letters 101. pp. 119-121.

Appendix 1: Generalization of the formulas to more than 3 categories and an arbitrary number of exogenous variables

We generalize the model and formulas to any number J of categories and any number of . exogenous variables. Thus, x_i is a $(K\times1)$ vector of exogenous variables (with possibly different distributions) and $\beta = (\beta_1. \ldots. \beta_K)'$ the corresponding $(K\times1)$ vector of parameters:

$$y_i^* = \beta' x_i + u_i. \quad i = 1. \ldots. N.$$

$$y_i = \begin{cases} 1, & \text{if } y_i^* \le \mu_1 \\ j, & \text{if } \mu_{j-1} < y_i^* \le \mu_j, \ j = 2,\ldots,J-1 \\ J, & \text{if } \mu_{J-1} < y_i^* \end{cases} \tag{A1.1}$$

The loglikelihood function slightly changes to:

$$\ln L(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} \ln\left(p_{ij}(\theta)\right). \ y_{ij} = \begin{cases} 1, & \text{if } y_i = j \\ 0, & \text{otherwise} \end{cases}. \ j = 1. \ldots. J. \ p_{ij}(\theta) = P\left(y_{ij} = 1 \mid \theta,\, x_i\right).$$

$$\theta = \left(\beta',\, \mu_1, \ldots,\, \mu_{J-1}\right)' \text{ the vector of unknown parameters.} \tag{A1.2}$$

$$p_{i1}(\theta) = \Phi(\mu_1 - \beta' x_i). \ p_{ij}(\theta) = \Phi\left(\mu_j - \beta' x_i\right) - \Phi\left(\mu_{j-1} - \beta' x_i\right). \ j = 2. \ldots. J-1.$$

$$p_{iJ}(\theta) = 1 - \Phi\left(\mu_{J-1} - \beta' x_i\right). \tag{A1.3}$$

Under the alternative now $\theta = \left(\beta',\, \mu_1, \ldots,\, \mu_{J-1},\, c_1,\, c_2\right)'$. The structure of the LM statistic remains unchanged. i.e.

$$LM = \left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial\theta}\right)' \left[E_{H_0}\left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial\theta}\right)\left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial\theta}\right)'\right]^{-1} \left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial\theta}\right). \tag{A1.4}$$

lnL again the loglikelihood function under H_1. $\hat{\theta}_r = \left(\hat{\beta}',\, \hat{\mu}_1, \ldots,\, \hat{\mu}_{J-1},\, 0,\, 0\right)'$. $\hat{\beta}'$. $\hat{\mu}_1$. …. $\hat{\mu}_{J-1}$ the ML-estimates of an ordinal probit.

Let

$$\hat{\mu}_{im} = \hat{\mu}_m - \hat{\beta}' x_i. \ \hat{p}_{i1} = \Phi\left(\hat{\mu}_{i1}\right). \ \hat{p}_{ij} = \Phi\left(\hat{\mu}_{ij}\right) - \Phi\left(\hat{\mu}_{i,j-1}\right). \ j = 2. \ldots. J-1.$$

$$\hat{p}_{iJ} = 1 - \Phi\left(\hat{\mu}_{i,J-1}\right). \text{ and } \hat{\varphi}_{im} = \varphi\left(\hat{\mu}_{im}\right). \ m = 1. \ldots. J-1 \tag{A1.5}$$

Then. the gradient vector equals (cf. Johnson 1996. pp. 215-216):

$$\frac{\partial \ln L}{\partial \mu_j}\left(\hat{\theta}_r\right) = \sum_{i=1}^{N}\left[\frac{y_{ij}}{\hat{p}_{ij}} - \frac{y_{i,j+1}}{\hat{p}_{i.j+1}}\right]\hat{\varphi}_{ij}. \quad j = 1. \ldots. J-1 \tag{A1.6a}$$

$$\frac{\partial \ln L}{\partial \beta_k}\left(\hat{\theta}_r\right)=\sum_{i=1}^{N}\left[y_{i1}\frac{-\hat{\varphi}_{i1}}{\hat{p}_{i1}}+\sum_{j=2}^{J-1}y_{ij}\frac{\hat{\varphi}_{i,j-1}-\hat{\varphi}_{ij}}{\hat{p}_{ij}}+y_{iJ}\frac{\hat{\varphi}_{i,J-1}}{\hat{p}_{iJ}}\right]x_{ik}\,.\quad k=1.\;\ldots.\;K \tag{A1.6b}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial c_1}\left(\hat{\theta}_r\right)=-\frac{1}{3}\sum_{i=1}^{N}\Bigg[&y_{i1}\frac{-\left(\hat{\mu}_{i1}^2-1\right)\hat{\varphi}_{i1}}{\hat{p}_{i1}}+\sum_{j=2}^{J-1}y_{ij}\frac{\left(\hat{\mu}_{i,j-1}^2-1\right)\hat{\varphi}_{i,j-1}-\left(\hat{\mu}_{ij}^2-1\right)\hat{\varphi}_{ij}}{\hat{p}_{ij}}\\&+y_{iJ}\frac{\left(\hat{\mu}_{i,J-1}^2-1\right)\hat{\varphi}_{i,J-1}}{\hat{p}_{iJ}}\Bigg]\end{aligned} \tag{A1.6c}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial c_2}\left(\hat{\theta}_r\right)=\frac{1}{4}\sum_{i=1}^{N}\Bigg[&y_{i1}\frac{-\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}}{\hat{p}_{i1}}+\sum_{j=2}^{J-1}y_{ij}\frac{\hat{\mu}_{i,j-1}\left(3+\hat{\mu}_{i,j-1}^2\right)\hat{\varphi}_{i,j-1}-\hat{\mu}_{ij}\left(3+\hat{\mu}_{ij}^2\right)\hat{\varphi}_{ij}}{\hat{p}_{ij}}\\&+y_{iJ}\frac{\hat{\mu}_{i,J-1}\left(3+\hat{\mu}_{i,J-1}^2\right)\hat{\varphi}_{i,J-1}}{\hat{p}_{iJ}}\Bigg]\end{aligned} \tag{A1.6d}$$

Denoting $\mu = (\mu_1.\;\ldots.\;\mu_{J-1})'$. the information matrix can be partitioned as follows:

$$E_{H_0}\left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial \theta}\right)\left(\frac{\partial \ln L\left(\hat{\theta}_r\right)}{\partial \theta}\right)'$$

$$=E_{H_0}\begin{pmatrix}
\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\\
\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\\
\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\\
\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \beta}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\left(\frac{\partial \ln L(\hat{\theta}_r)}{\partial \mu}\right)' & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_1} & \frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}\frac{\partial \ln L(\hat{\theta}_r)}{\partial c_2}
\end{pmatrix}$$

$$=:\begin{pmatrix}
J_{\beta\beta} & J_{\beta\mu} & J_{\beta c_1} & J_{\beta c_2}\\
J_{\beta\mu}{}' & J_{\mu\mu} & J_{\mu c_1} & J_{\mu c_2}\\
J_{\beta c_1}{}' & J_{\mu c_1}{}' & J_{c_1c_1} & J_{c_1c_2}\\
J_{\beta c_2}{}' & J_{\mu c_2}{}' & J_{c_1c_2} & J_{c_2c_2}
\end{pmatrix} \tag{A1.7}$$

Following Glewwe (1997. pp. 5-6) the partial matrices are:

$$J_{\beta\beta}=\sum_{i=1}^{N}x_i x_i{}'\left(\frac{\hat{\varphi}_{i1}^2}{\hat{p}_{i1}}+\sum_{j=2}^{J-1}\frac{\left(\hat{\varphi}_{i,j-1}-\hat{\varphi}_{ij}\right)^2}{\hat{p}_{ij}}+\frac{\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}}\right) \tag{A1.8a}$$

$$\begin{aligned} J_{\beta\mu} = \Bigg(& \sum_{i=1}^{N} x_i \left(\frac{-\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} - \frac{\left(\hat{\varphi}_{i1} - \hat{\varphi}_{i2}\right)\hat{\varphi}_{i1}}{\hat{p}_{i2}} \right) \cdots \sum_{i=1}^{N} x_i \left(\frac{\left(\hat{\varphi}_{i,j-1} - \hat{\varphi}_{ij}\right)\hat{\varphi}_{ij}}{\hat{p}_{ij}} - \frac{\left(\hat{\varphi}_{ij} - \hat{\varphi}_{i,j+1}\right)\hat{\varphi}_{ij}}{\hat{p}_{i,j+1}} \right) \\ & \cdots \sum_{i=1}^{N} x_i \left(\frac{\left(\hat{\varphi}_{i,J-2} - \hat{\varphi}_{i,J-1}\right)\hat{\varphi}_{i,J-1}}{\hat{p}_{i,J-1}} - \frac{\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}} \right) \Bigg) \end{aligned} \tag{A1.8b}$$

$$J_{\beta c_1} = -\frac{1}{3} \sum_{i=1}^{N} x_i \left(\frac{\left(\hat{\mu}_{i1}^2 - 1\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \sum_{j=2}^{J-1} \frac{\left(\hat{\varphi}_{i,j-1} - \hat{\varphi}_{ij}\right)\left(\left(\hat{\mu}_{i,j-1}^2 - 1\right)\hat{\varphi}_{i,j-1} - \left(\hat{\mu}_{ij}^2 - 1\right)\hat{\varphi}_{ij}\right)}{\hat{p}_{ij}} + \frac{\left(\hat{\mu}_{i,J-1}^2 - 1\right)\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}} \right) \tag{A1.8c}$$

$$\begin{aligned} J_{\beta c_2} = \frac{1}{4} \sum_{i=1}^{N} x_i \Bigg(& \frac{\hat{\mu}_{i1}\left(3 + \hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \sum_{j=2}^{J-1} \frac{\left(\hat{\varphi}_{i,j-1} - \hat{\varphi}_{ij}\right)\left(\hat{\mu}_{i,j-1}\left(3 + \hat{\mu}_{i,j-1}^2\right)\hat{\varphi}_{i,j-1} - \hat{\mu}_{ij}\left(3 + \hat{\mu}_{ij}^2\right)\hat{\varphi}_{ij}\right)}{\hat{p}_{ij}} \\ & + \frac{\hat{\mu}_{i,J-1}\left(3 + \hat{\mu}_{i,J-1}^2\right)\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}} \Bigg) \end{aligned} \tag{A1.8d}$$

$$J_{\mu\mu} = \begin{pmatrix} \sum_{i=1}^{N} \left(\frac{\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \frac{\hat{\varphi}_{i1}^2}{\hat{p}_{i2}} \right) & -\sum_{i=1}^{N} \frac{\hat{\varphi}_{i1}\hat{\varphi}_{i2}}{\hat{p}_{i2}} & 0 & & \\ & & \vdots & & \\ 0 & -\sum_{i=1}^{N} \frac{\hat{\varphi}_{i,j-1}\hat{\varphi}_{ij}}{\hat{p}_{ij}} & \sum_{i=1}^{N} \left(\frac{\hat{\varphi}_{ij}^2}{\hat{p}_{ij}} + \frac{\hat{\varphi}_{ij}^2}{\hat{p}_{i,j+1}} \right) & -\sum_{i=1}^{N} \frac{\hat{\varphi}_{ij}\hat{\varphi}_{i,j+1}}{\hat{p}_{i,j+1}} & 0 \\ & & \vdots & & \\ & & 0 & -\sum_{i=1}^{N} \frac{\hat{\varphi}_{i,J-2}\hat{\varphi}_{i,J-1}}{\hat{p}_{i,J-1}} & \sum_{i=1}^{N} \left(\frac{\hat{\varphi}_{i,J-1}^2}{\hat{p}_{i,J-1}} + \frac{\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}} \right) \end{pmatrix} \tag{A1.8e}$$

$$J_{\mu c_1} = \begin{pmatrix} -\frac{1}{3} \sum_{i=1}^{N} \left(\frac{-\left(\hat{\mu}_{i1}^2 - 1\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} - \frac{\hat{\varphi}_{i1}\left(\left(\hat{\mu}_{i1}^2 - 1\right)\hat{\varphi}_{i1} - \left(\hat{\mu}_{i2}^2 - 1\right)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}} \right) \\ \vdots \\ -\frac{1}{3} \sum_{i=1}^{N} \left(\frac{\hat{\varphi}_{ij}\left(\left(\hat{\mu}_{i,j-1}^2 - 1\right)\hat{\varphi}_{i,j-1} - \left(\hat{\mu}_{ij}^2 - 1\right)\hat{\varphi}_{ij}\right)}{\hat{p}_{ij}} - \frac{\hat{\varphi}_{ij}\left(\left(\hat{\mu}_{ij}^2 - 1\right)\hat{\varphi}_{ij} - \left(\hat{\mu}_{i,j+1}^2 - 1\right)\hat{\varphi}_{i,j+1}\right)}{\hat{p}_{i,j+1}} \right) \\ \vdots \\ -\frac{1}{3} \sum_{i=1}^{N} \left(\frac{\hat{\varphi}_{i,J-1}\left(\left(\hat{\mu}_{i,J-2}^2 - 1\right)\hat{\varphi}_{i,J-2} - \left(\hat{\mu}_{i,J-1}^2 - 1\right)\hat{\varphi}_{i,J-1}\right)}{\hat{p}_{i,J-1}} - \frac{\left(\hat{\mu}_{i,J-1}^2 - 1\right)\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}} \right) \end{pmatrix} \tag{A1.8f}$$

$$J_{\mu c_2} = \begin{pmatrix} \frac{1}{4}\sum_{i=1}^{N}\left(\frac{-\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} - \frac{\hat{\varphi}_{i1}\left(\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1} - \hat{\mu}_{i2}\left(3+\hat{\mu}_{i2}^2\right)\hat{\varphi}_{i2}\right)}{\hat{p}_{i2}}\right) \\ \vdots \\ \frac{1}{4}\sum_{i=1}^{N}\left(\frac{\hat{\varphi}_{ij}\left(\hat{\mu}_{i,j-1}\left(3+\hat{\mu}_{i,j-1}^2\right)\hat{\varphi}_{i,j-1} - \hat{\mu}_{ij}\left(3+\hat{\mu}_{ij}^2\right)\hat{\varphi}_{ij}\right)}{\hat{p}_{ij}} - \frac{\hat{\varphi}_{ij}\left(\hat{\mu}_{ij}\left(3+\hat{\mu}_{ij}^2\right)\hat{\varphi}_{ij} - \hat{\mu}_{i,j+1}\left(3+\hat{\mu}_{i,j+1}^2\right)\hat{\varphi}_{i,j+1}\right)}{\hat{p}_{i,j+1}}\right) \\ \vdots \\ \frac{1}{4}\sum_{i=1}^{N}\left(\frac{\hat{\varphi}_{i,J-1}\left(\hat{\mu}_{i,J-2}\left(3+\hat{\mu}_{i,J-2}^2\right)\hat{\varphi}_{i,J-2} - \hat{\mu}_{i,J-1}\left(3+\hat{\mu}_{i,J-1}^2\right)\hat{\varphi}_{i,J-1}\right)}{\hat{p}_{i,J-1}} - \frac{\hat{\mu}_{i,J-1}\left(3+\hat{\mu}_{i,J-1}^2\right)\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}}\right) \end{pmatrix} \tag{A1.8g}$$

$$J_{c_1c_1} = \frac{1}{9}\sum_{i=1}^{N}\left(\frac{\left(\hat{\mu}_{i1}^2-1\right)^2\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \sum_{j=2}^{J-1}\frac{\left(\left(\hat{\mu}_{i,j-1}^2-1\right)\hat{\varphi}_{i,j-1} - \left(\hat{\mu}_{ij}^2-1\right)\hat{\varphi}_{ij}\right)^2}{\hat{p}_{ij}} + \frac{\left(\hat{\mu}_{i,J-1}^2-1\right)^2\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}}\right) \tag{A1.8h}$$

$$\begin{aligned} J_{c_1c_2} = -\frac{1}{12}\sum_{i=1}^{N}\Bigg(&\frac{\left(\hat{\mu}_{i1}^2-1\right)\hat{\mu}_{i1}\left(3+\hat{\mu}_{i1}^2\right)\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} \\ &+\sum_{j=2}^{J-1}\frac{\left(\left(\hat{\mu}_{i,j-1}^2-1\right)\hat{\varphi}_{i,j-1} - \left(\hat{\mu}_{ij}^2-1\right)\hat{\varphi}_{ij}\right)\left(\hat{\mu}_{i,j-1}\left(3+\hat{\mu}_{i,j-1}^2\right)\hat{\varphi}_{i,j-1} - \hat{\mu}_{ij}\left(3+\hat{\mu}_{ij}^2\right)\hat{\varphi}_{ij}\right)}{\hat{p}_{ij}} + \frac{\left(\hat{\mu}_{i,J-1}^2-1\right)\hat{\mu}_{i,J-1}\left(3+\hat{\mu}_{i,J-1}^2\right)\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}}\Bigg) \end{aligned} \tag{A1.8i}$$

$$J_{c_2c_2} = \frac{1}{16}\sum_{i=1}^{N}\left(\frac{\hat{\mu}_{i1}^2\left(3+\hat{\mu}_{i1}^2\right)^2\hat{\varphi}_{i1}^2}{\hat{p}_{i1}} + \sum_{j=2}^{J-1}\frac{\left(\hat{\mu}_{i,j-1}\left(3+\hat{\mu}_{i,j-1}^2\right)\hat{\varphi}_{i,j-1} - \hat{\mu}_{ij}\left(3+\hat{\mu}_{ij}^2\right)\hat{\varphi}_{ij}\right)^2}{\hat{p}_{ij}} + \frac{\hat{\mu}_{i,J-1}^2\left(3+\hat{\mu}_{i,J-1}^2\right)^2\hat{\varphi}_{i,J-1}^2}{\hat{p}_{iJ}}\right) \tag{A1.8j}$$

These expressions are again used for an artificial regression as in (2.9) and the calculation of the test-statistic according to (2.10). Denote with r_i and R_i the now J rows of r and R for the i-th observation. Then.

$$r_i = \begin{pmatrix} \frac{y_{i1}}{\sqrt{\hat{p}_{i1}}} \\ \vdots \\ \frac{y_{iJ}}{\sqrt{\hat{p}_{iJ}}} \end{pmatrix}.$$

$$
R_i = \left(\begin{matrix}
\dfrac{-\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} x_{i1} & & \dfrac{-\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} x_{iK} & \dfrac{\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} & 0 & & \\
\vdots & & \vdots & \dfrac{-\hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i2}}} & \dfrac{\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i2}}} & & \\
\dfrac{\hat{\varphi}_{i,j-1} - \hat{\varphi}_{ij}}{\sqrt{\hat{p}_{ij}}} x_{i1} & \cdots & \dfrac{\hat{\varphi}_{i,j-1} - \hat{\varphi}_{ij}}{\sqrt{\hat{p}_{ij}}} x_{iK} & 0 & \dfrac{-\hat{\varphi}_{i2}}{\sqrt{\hat{p}_{i3}}} & \cdots & 0 \\
\vdots & & \vdots & & 0 & & \dfrac{\hat{\varphi}_{i,J-1}}{\sqrt{\hat{p}_{i,J-1}}} \\
\dfrac{\hat{\varphi}_{i,J-1}}{\sqrt{\hat{p}_{iJ}}} x_{i1} & & \dfrac{\hat{\varphi}_{i,J-1}}{\sqrt{\hat{p}_{iJ}}} x_{iK} & & & & \dfrac{-\hat{\varphi}_{i,J-1}}{\sqrt{\hat{p}_{iJ}}}
\end{matrix} \right.
$$

$$
\left. \begin{matrix}
\dfrac{1}{3} \dfrac{\left(\hat{\mu}_{i1}^2 - 1\right) \hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} & -\dfrac{1}{4} \dfrac{\hat{\mu}_{i1} \left(3 + \hat{\mu}_{i1}^2\right) \hat{\varphi}_{i1}}{\sqrt{\hat{p}_{i1}}} \\
\vdots & \vdots \\
-\dfrac{1}{3} \dfrac{\left(\hat{\mu}_{i,j-1}^2 - 1\right) \hat{\varphi}_{i,j-1} - \left(\hat{\mu}_{ij}^2 - 1\right) \hat{\varphi}_{ij}}{\sqrt{\hat{p}_{ij}}} & \dfrac{1}{4} \dfrac{\hat{\mu}_{i,j-1} \left(3 + \hat{\mu}_{i,j-1}^2\right) \hat{\varphi}_{i,j-1} - \hat{\mu}_{ij} \left(3 + \hat{\mu}_{ij}^2\right) \hat{\varphi}_{ij}}{\sqrt{\hat{p}_{ij}}} \\
\vdots & \vdots \\
-\dfrac{1}{3} \dfrac{\left(\hat{\mu}_{i,J-1}^2 - 1\right) \hat{\varphi}_{i,J-1}}{\sqrt{\hat{p}_{iJ}}} & \dfrac{1}{4} \dfrac{\hat{\mu}_{i,J-1} \left(3 + \hat{\mu}_{i,J-1}^2\right) \hat{\varphi}_{i,J-1}}{\sqrt{\hat{p}_{iJ}}}
\end{matrix} \right)
$$

(A1.11a)

$$
r = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}. \quad R = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{pmatrix}.
$$

(A1.11b)

Appendix 2: Threshold values of the different designs depending on the quantiles used

a) Model with 3 categories

x_i ~	u_i ~	Quantile 1	Quantile 2	μ_1	μ_2
N(0,1)	N(0,1)	0.33	0.67	–0.62	0.62
		0.45	0.55	–0.18	0.18
		0.10	0.90	–1.81	1.81
		0.80	0.90	1.19	1.81
		0.10	0.20	–1.81	–1.19
	N(0,4)	0.33	0.67	–0,98	0,98
	N(0,10)			–1,46	1,46
	N(0,0.25)			–0,49	0,49
	N(0,0.1)			–0,46	0,46
	t(3)	0.33	0.67	–0.69	0.70
		0.45	0.55	–0.2	0.2
		0.10	0.90	–2.13	2.14
		0.80	0.90	1.36	2.14
		0.10	0.20	–2.13	–1.35
N(0,3)		0.33	0.67	–0.95	0.96
N(0,0.75)				–0.65	0.65
N(0,0.3)				–0.56	0.57
N(0,12)				–1.65	1.66
N(0,30)				–2.50	2.51
N(0,1)	t(9)	0.33	0.67	–0.64	0.65
		0.45	0.55	–0.18	0.19
		0.10	0.90	–1.90	1.90
		0.80	0.90	1.24	1.90
		0.10	0.20	–1.90	–1.24
N(0,1.29)		0.33	0.67	–0.69	0.69
N(0,0.32)				–0.52	0.53
N(0,0.13)				–0.48	0.49
N(0,5.14)				–1.11	1.11
N(0,12.9)				–1.65	1.66
N(0,1)	P1, s=1	0.33	0.67	–0.70	0.53
		0.45	0.55	–0.29	0.06
		0.10	0.90	–1.72	1.92
		0.80	0.90	1.18	1.92
		0.10	0.20	–1.72	-1.20
N(0,0.25)		0.33	0.67	–0.64	0.32
N(0,0.1)				–0.66	0.27
N(0,4)				–1.01	0.96
N(0,10)				–1.47	1.45
N(0,1)	P1, s=0.5	0.33	0.67	–0.66	0.59
		0.45	0.55	–0.22	0.13
		0.10	0.90	–1.78	1.85
		0.80	0.90	1.18	1.85
		0.10	0.20	–1.78	–1.20
N(0,0.25)		0.33	0.67	–0.56	0.44
N(0,0.1)				–0.54	0.40
N(0,4)				–1.00	0.97
N(0,10)				–1.47	1.45
N(0,1)	Γ(4)	0.33	0.67	–0.67	0.54
		0.45	0.55	–0.25	0.09
		0.10	0.90	–1.74	1.83

		0.80	0.90	1.14	1.83
		0.10	0.20	–1.74	–1.19
N(0,0.25)		0.33	0.67	–0.57	0.36
N(0,0.1)				–0.56	0.32
N(0,4)				–1.00	0.95
N(0,10)				–1.47	1.44

b) Model with 4 categories

x_i ~	**u_i ~**	**Quantile 1**	**Quantile 2**	**Quantile 3**	μ_1	μ_2	μ_3
N(0,1)	N(0,1)	0.25	0.50	0.75	-0.95	0	0.95
		0.70	0.80	0.90	0.74	1.19	1.81
	t(3)	0.25	0.50	0.75	-1.07	0	1.08
		0.70	0.80	0.90	0.83	1.36	2.14
	t(9)	0.25	0.50	0.75	-0.99	0	0.99
		0.70	0.80	0.90	0.77	1.24	1.90
	P1, s=1	0.25	0.50	0.75	-1.00	-0.11	0.90
		0.70	0.80	0.90	0.66	1.18	1.92
	P1, s=0.5	0.25	0.50	0.75	-0.98	-0.05	0.93
		0.70	0.80	0.90	0.71	1.18	1.85
	Γ(4)	0.25	0.50	0.75	-0.98	-0.08	0.88
		0.70	0.80	0.90	0.66	1.14	1.83

Appendix 3: R program for the simulation study

The following code shows the simulation against the Pearson 1 distribution as an example.

```
library(MASS)
library(AER) # package is needed for robust standard errors

bjl_oprobit <- function(replikat,n,beta,c1,c2, robust=c("NONE", "HC3", "White")) {
K=1 # K number of exogenous variables

# matrices for simulation results
testerg=matrix(nrow=replikat+1,ncol=30)
testerg[1,]=c("LM 1%", "LM 5%", "LM 10%", "c1 1%", "c1 5%", "c1 10%", "c2 1%", "c2 5%",
"c2 10%", "tc1-tc2", "LM 1%", "LM 5%", "LM 10%", "c1 1%", "c1 5%", "c1 10%", "c2 1%",
"c2 5%", "c2 10%", "tc1-tc2","LM 1%", "LM 5%", "LM 10%", "c1 1%", "c1 5%", "c1 10%",
"c2 1%", "c2 5%", "c2 10%", "tc1-tc2")
Erg=matrix(nrow=replikat+1, ncol=4)
Erg[1,]=c("tc1Wahr2000", "tc1Wahr400","tc2Wahr2000", "tc2Wahr400")

# calculating parameters for Pearson 1 distribution
schiefe=1
r=2*(2-schiefe^2)/schiefe^2
r3=0.5*r+0.5*r*(r+2)*sqrt(schiefe^2/(schiefe^2*(r+2)^2+16*(r+1)))
r4=0.5*r-0.5*r*(r+2)*sqrt(schiefe^2/(schiefe^2*(r+2)^2+16*(r+1)))
if(schiefe>0)
{ q=max(r3,r4)
  p=min(r3,r4)} else{
  q=min(r3,r4)
  p=max(r3,r4)
}
b=(p+q)*sqrt((p+q+1)/(p*q))
a=-b*p/(p+q)

# drawing random numbers for x (once) and u (each replication)
m=1 # m loop index for sample size, starting with largest sample size
while(m<=2) {
set.seed(210465)
x=rnorm(n,mean=0,sd=sqrt(1))

set.seed(310465)

i=1
while(i<=replikat) {
u1=rgamma(n,p)
u2=rgamma(n,q)
u=b*u1/(u1+u2)+a
y_st=beta*x+u
y=as.numeric(cut(y_st,c(-1000,c1,c2,1000)))
yi0=as.numeric(y==1)
yi1=as.numeric(y==2)
yi2=as.numeric(y==3)
```

```
# estimation of ordered probit model and saving results
oprobit = polr(factor(y) ~ x, start=c(0,-0.5,0.5), method="probit")
beta_hat=oprobit$coefficients
y_st_hat=beta_hat*x
a0_hat=oprobit$zeta[1]
a1_hat=oprobit$zeta[2]

# calculation of auxiliary variables for calculating the variables of the auxiliary regression
phi_i0=dnorm(a0_hat-beta_hat*x)
phi_i1=dnorm(a1_hat-beta_hat*x)
Phi_i0=pnorm(a0_hat-beta_hat*x)
Phi_i1=pnorm(a1_hat-beta_hat*x)

# calculation of the variables of the auxiliary regression
indiv=1:n
ri.1=yi0/sqrt(Phi_i0)
ri.2=yi1/sqrt(Phi_i1-Phi_i0)
ri.3=yi2/sqrt(1-Phi_i1)
Rbi.1=-phi_i0*x/sqrt(Phi_i0)
Rbi.2=(phi_i0-phi_i1)*x/sqrt(Phi_i1-Phi_i0)
Rbi.3=phi_i1*x/sqrt(1-Phi_i1)
Ra0i.1=phi_i0/sqrt(Phi_i0)
Ra0i.2=-phi_i0/sqrt(Phi_i1-Phi_i0)
Ra0i.3=indiv*0
Ra1i.1=indiv*0
Ra1i.2=phi_i1/sqrt(Phi_i1-Phi_i0)
Ra1i.3=-phi_i1/sqrt(1-Phi_i1)
Rc1i.1=(1/3)*((a0_hat-beta_hat*x)^2-1)*phi_i0/sqrt(Phi_i0)
Rc1i.2=-(1/3)*(((a0_hat-beta_hat*x)^2-1)*phi_i0-((a1_hat-beta_hat*x)^2-
1)*phi_i1)/sqrt(Phi_i1-Phi_i0)
Rc1i.3=-(1/3)*(((a1_hat-beta_hat*x)^2-1)*phi_i1)/sqrt(1-Phi_i1)
Rc2i.1=-(1/4)*(a0_hat-beta_hat*x)*(3+(a0_hat-beta_hat*x)^2)*phi_i0/sqrt(Phi_i0)
Rc2i.2=(1/4)*((a0_hat-beta_hat*x)*(3+(a0_hat-beta_hat*x)^2)*phi_i0-(a1_hat-
beta_hat*x)*(3+(a1_hat-beta_hat*x)^2)*phi_i1)/sqrt(Phi_i1-Phi_i0)
Rc2i.3=(1/4)*(a1_hat-beta_hat*x)*(3+(a1_hat-beta_hat*x)^2)*phi_i1/sqrt(1-Phi_i1)

data_wide=data.frame(indiv,ri.1,ri.2,ri.3,Rbi.1,Rbi.2,Rbi.3,Ra0i.1,Ra0i.2,Ra0i.3,Ra1i.1,Ra1i.2,R
a1i.3,Rc1i.1,Rc1i.2,Rc1i.3,Rc2i.1,Rc2i.2,Rc2i.3)
data_long=reshape(data_wide,idvar="indiv",varying=2:19,sep=".",direction="long")
rij=data_long$ri
Rbij=data_long$Rbi
Ra0ij=data_long$Ra0i
Ra1ij=data_long$Ra1i
Rc1ij=data_long$Rc1i
Rc2ij=data_long$Rc2i

# estimation of the auxiliary regression
hilfsregr = lm(rij ~ 0 + Rbij + Ra0ij + Ra1ij + Rc1ij + Rc2ij)
rij_hat=fitted(hilfsregr)
```

```
# test decision LM test
LM = sum(rij_hat*rij_hat)
LM_1=as.numeric(LM>=qchisq(0.99, df=2))
LM_5=as.numeric(LM>=qchisq(0.95, df=2))
LM_10=as.numeric(LM>=qchisq(0.90, df=2))

# test decisions t-statistics and double indicators depending on choice of standard errors
if (robust=="NONE")
{
 pvalue_c1=summary(hilfsregr)$coefficients[K+3,4]
 pvalue_c2=summary(hilfsregr)$coefficients[K+4,4]
 tc1=summary(hilfsregr)$coefficients[K+3,3]
 tc2=summary(hilfsregr)$coefficients[K+4,3]
}

if (robust=="HC3"){

 robusteSE=coeftest(hilfsregr, vcov=vcovHC(hilfsregr, type=c("HC3"))) #Ergebnis ist Matrix
mit Spalten estimate, SE, t-Stat., p-Wert
 pvalue_c1=robusteSE[K+3,4]
 pvalue_c2=robusteSE[K+4,4]
 tc1=robusteSE[K+3,3]
 tc2=robusteSE[K+4,3]
}

if (robust=="White"){

 robusteSE=coeftest(hilfsregr, vcov=vcovHC(hilfsregr, type=c("HC"))) #Ergebnis ist Matrix mit
Spalten estimate, SE, t-Stat., p-Wert
 pvalue_c1=robusteSE[K+3,4]
 pvalue_c2=robusteSE[K+4,4]
 tc1=robusteSE[K+3,3]
 tc2=robusteSE[K+4,3]
}

tc1_1=as.numeric(pvalue_c1<=0.01)
tc1_5=as.numeric(pvalue_c1<=0.05)
tc1_10=as.numeric(pvalue_c1<=0.1)
tc2_1=as.numeric(pvalue_c2<=0.01)
tc2_5=as.numeric(pvalue_c2<=0.05)
tc2_10=as.numeric(pvalue_c2<=0.1)
d_tc1_tc2=as.numeric(abs(tc1)-abs(tc2)>=0)
Ergebnis=as.numeric(abs(tc1) >= 1.96 & abs(tc1)>=1.5*abs(tc2))
Ergebnis2=as.numeric(abs(tc2) >= 1.96 & abs(tc2)>=1.5*abs(tc1))

# results written into matrices that can be read out afterwards
testerg[i+1,1+(m-1)*10]=LM_1
testerg[i+1,2+(m-1)*10]=LM_5
testerg[i+1,3+(m-1)*10]=LM_10
testerg[i+1,4+(m-1)*10]=tc1_1
testerg[i+1,5+(m-1)*10]=tc1_5
```

```
testerg[i+1,6+(m-1)*10]=tc1_10
testerg[i+1,7+(m-1)*10]=tc2_1
testerg[i+1,8+(m-1)*10]=tc2_5
testerg[i+1,9+(m-1)*10]=tc2_10
testerg[i+1,10+(m-1)*10]=d_tc1_tc2
Erg[i+1,1+(m-1)]=Ergebnis
Erg[i+1,3+(m-1)]=Ergebnis2
i=i+1
} # end of the loop for replications

m=m+1
n=n/5
} # end of the loop for sample sizes

write.matrix(testerg,sep="\t",file="local path\\tests.r")
write.matrix(Erg,sep="\t",file="local path\\indicators.r")

} # end of function

# example: results used in table 9, quantiles 0.33/0.67, and table 14, s=1, quantiles 0.33/0.67
bjl_oprobit(5000,2000,1,-0.70,0.53, robust="NONE")
```

Appendix 4: R program for application (3 categories)

```
# The subsequent code can be used to calculate the tests and indicators for an ordered probit
# model with 3 categories and an arbitrary number of exogenous variables.
# The first argument of the function is the name of the dataset. The endogenous categorical
# variable has to be named y, and its categories have to be named 0, 1, and 2.
# The second argument is the function required as an argument for the ordered probit estimation.
# E.g., if the exogenous variables are x1, x2 and x3, the formula would be: factor(y)~x1+x2+x3

lmtestprobit=function(data, formula)
{if (!require("MASS")) install.packages("MASS", dependencies = TRUE)
 library(MASS)
 n=nrow(data)
 yi0=as.numeric(y==0)
 yi1=as.numeric(y==1)
 yi2=as.numeric(y==2)

# estimation of ordered probit model and saving results
 oprobit = polr(formula, data=data, method="probit")
  beta_hat = oprobit$coefficients
  a0_hat = oprobit$zeta[1]
  a1_hat = oprobit$zeta[2]
  independent_vars = data[, names(beta_hat), drop = FALSE]
  y_st_hat = as.matrix(independent_vars) %*% beta_hat

# calculation of auxiliary variables for calculating the variables of the auxiliary regression
 phi_i0=dnorm(a0_hat-y_st_hat)
 phi_i1=dnorm(a1_hat-y_st_hat)
 Phi_i0=pnorm(a0_hat-y_st_hat)
 Phi_i1=pnorm(a1_hat-y_st_hat)

# calculation of the variables of the auxiliary regression
 indiv=1:n
 ri.1=yi0/sqrt(Phi_i0)
 ri.2=yi1/sqrt(Phi_i1-Phi_i0)
 ri.3=yi2/sqrt(1-Phi_i1)
 calculate_Rbi = function(x_matrix) {
  k = nrow(x_matrix)
  m = ncol(x_matrix)
  Rb_matrix1 = matrix(0, nrow = k, ncol = m)
  Rb_matrix2 = matrix(0, nrow = k, ncol = m)
```

```
    Rb_matrix3 = matrix(0, nrow = k, ncol = m)
    for (j in 1:m) {
      x_values = x_matrix[, j]  # Werte der j-ten Spalte
      for (i in 1:k) {
        Rb_matrix1[i, j] = -phi_i0[i] * x_values[i] / sqrt(Phi_i0[i])
        Rb_matrix2[i, j] = (phi_i0[i] - phi_i1[i]) * x_values[i] / sqrt(Phi_i1[i] - Phi_i0[i])
        Rb_matrix3[i, j] = phi_i1[i] * x_values[i] / sqrt(1 - Phi_i1[i])
      }
    }
    colnames(Rb_matrix1) = paste0("Rbx", 1:m, ".1")
    colnames(Rb_matrix2) = paste0("Rbx", 1:m, ".2")
    colnames(Rb_matrix3) = paste0("Rbx", 1:m, ".3")

    return(list(Rb_matrix1 = Rb_matrix1, Rb_matrix2 = Rb_matrix2, Rb_matrix3 = Rb_matrix3))
  }
  Rb_matrices = calculate_Rbi(independent_vars)
  Rb_matrices = data.frame(Rb_matrices$Rb_matrix1, Rb_matrices$Rb_matrix2, Rb_matrices$Rb_matrix3)
  Ra0i.1=phi_i0/sqrt(Phi_i0)
  Ra0i.2=-phi_i0/sqrt(Phi_i1-Phi_i0)
  Ra0i.3=indiv*0
  Ra1i.1=indiv*0
  Ra1i.2=phi_i1/sqrt(Phi_i1-Phi_i0)
  Ra1i.3=-phi_i1/sqrt(1-Phi_i1)
  Rc1i.1=(1/3)*((a0_hat-y_st_hat)^2-1)*phi_i0/sqrt(Phi_i0)
  Rc1i.2=-(1/3)*(((a0_hat-y_st_hat)^2-1)*phi_i0-((a1_hat-y_st_hat)^2-1)*phi_i1)/sqrt(Phi_i1-Phi_i0)
  Rc1i.3=-(1/3)*(((a1_hat-y_st_hat)^2-1)*phi_i1)/sqrt(1-Phi_i1)
  Rc2i.1=-(1/4)*(a0_hat-y_st_hat)*(3+(a0_hat-y_st_hat)^2)*phi_i0/sqrt(Phi_i0)
  Rc2i.2=(1/4)*((a0_hat-y_st_hat)*(3+(a0_hat-y_st_hat)^2)*phi_i0-(a1_hat-
y_st_hat)*(3+(a1_hat-y_st_hat)^2)*phi_i1)/sqrt(Phi_i1-Phi_i0)
  Rc2i.3=(1/4)*(a1_hat-y_st_hat)*(3+(a1_hat-y_st_hat)^2)*phi_i1/sqrt(1-Phi_i1)

  data_wide = data.frame(indiv, ri.1, ri.2, ri.3)  # Initialisiere mit den ersten Spalten
  data_wide = cbind(data_wide, Rb_matrices)
  data_wide = cbind(data_wide,
    Ra0i.1, Ra0i.2, Ra0i.3,
    Ra1i.1, Ra1i.2, Ra1i.3,
    Rc1i.1, Rc1i.2, Rc1i.3,
    Rc2i.1, Rc2i.2, Rc2i.3
  )
  data_long=reshape(data_wide,idvar="indiv",varying=2:ncol(data_wide),sep=".",direction="long")
```

```
rij=data_long$ri
 Rbx_columns = grep("^Rbx", colnames(data_long), value = TRUE)
 Rbx_vectors = lapply(Rbx_columns, function(col_name) data_long[[col_name]])
 for (i in 1:length(Rbx_vectors)) {
   assign(paste0("Rbx", i), Rbx_vectors[[i]])
 }
 Ra0ij=data_long$Ra0i
 Ra1ij=data_long$Ra1i
 Rc1ij=data_long$Rc1i
 Rc2ij=data_long$Rc2i
 index_i=data_long$indiv
 index_j=data_long$time

 # estimation of the auxiliary regression and test decisions
 model_formula <- paste("rij ~ 0 +", paste(Rbx_columns, collapse = " + "), "+ Ra0ij + Ra1ij + Rc1ij + Rc2ij")
 cat("formula for auxiliary regression: ", model_formula, "\n")
 hilfsregr <- lm(as.formula(model_formula), data = data_long)
 K=ncol(independent_vars)
 rij_hat=fitted(hilfsregr)
 LM = sum(rij_hat*rij_hat)
 pvalue_c1=summary(hilfsregr)$coefficients[K+3,4]
 pvalue_c2=summary(hilfsregr)$coefficients[K+4,4]
 tc1=summary(hilfsregr)$coefficients[K+3,3]
 tc2=summary(hilfsregr)$coefficients[K+4,3]
 cat("The LM test statistic is", LM, "\n","\n")
 cat("The test statistic and p-value for skewness are t=", tc1, "and p=",pvalue_c1, "\n")
 cat("The test statistic and p-value for kurtosis are t=", tc2, "and p=",pvalue_c2, "\n", "\n" )
 if(abs(tc1) < 1.96 & abs(tc1)<1.5*abs(tc2)&abs(tc2) < 1.96 & abs(tc2)<1.5*abs(tc1))
   {
   cat("The empirical measure does not indicate a skewness or kurtosis problem", "\n")
 }
 if(abs(tc1) < 1.96 & abs(tc1)<1.5*abs(tc2)&abs(tc2) >= 1.96 & abs(tc2)>=1.5*abs(tc1))
 {
   cat("The empirical measure indicates a kurtosis problem", "\n")
 }
 if(abs(tc1) >= 1.96 & abs(tc1)>=1.5*abs(tc2)&abs(tc2) < 1.96 & abs(tc2)<1.5*abs(tc1))
 {
   cat("The empirical measure indicates a skewness problem", "\n")
 }

}
```