Weak identification in probit models with endogenous covariates[1]

Jean-Marie Dufour[2]

Department of Economics, McGill University, Canada

Joachim Wilde[3]

Department of Business Administration and Economics, Osnabrueck University, Germany

August 2017

Abstract

Weak identification is a well-known issue in the context of linear structural models. However, little is known on this problem for probit models with endogenous explanatory variables. In this paper, we study by simulation the behavior of the usual z-test and the LR-test in the presence of weak identification. We find that the usual asymptotic z-test exhibits large size distortions (overrejections under the null hypothesis). The magnitude of the size distortions depends heavily on the parameter value tested. In contrast, asymptotic LR-tests do not over-reject and appear to be robust to weak identification.

Keywords: probit model, weak identification

JEL classification: C 35

1. Introduction

Probit models are by now widely used in applied econometrics (in recent years, e.g., Abramitzky/ Lavy 2014, Beck/ Lin/ Ma 2014, Bijsterbosch/ Dahlhaus 2015, Bouoiyour/ Miftah/ Mouhoud 2016, Cornelli/ Kominek/ Ljungqvist 2013, Croushore/ Marsten 2016, Engelhardt et.al. 2010, Esaka 2010, Fitzenberger/ Kohn/ Wang 2011, Haider / Jahangir 2017, Hao/ Ng 2011, Hlaing/ Pourjalali 2012, Horvath/ Katuscakova 2016, Khanna/ Kim/ Lu 2015, Litchfield/ Reilly/ Veneziani 2012, Massa/ Zhang 2013, Wen/ Gordon 2014). As in linear models, one or more explaining variables might be endogenous. This problem can be solved by using instrumental variables (cf. Wilde 2008 for a comparison of different estimation methods using instrumental variables). The resulting estimates can be used to calculate test statistics for the parameters of the model.

---

[1] The authors thank Leandro Magnusson for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

[2] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ).
Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800;
e-mail: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com.

[3] Corresponding author. Mailing address: Fachbereich Wirtschaftswissenschaften, Rolandstr. 8, 49069 Osnabrueck, Germany; Tel: (49) 541 969 2746; e-mail: Joachim.Wilde@uni-osnabrueck.de. Web page: https://www.wiwi.uni-osnabrueck.de/en/departments_and_institutes/econometrics_and_statistics_prof_wilde.html

However, in linear models it is well known that weak instruments may cause considerable size distortions (cf. Dufour 2003 for an overview). Wald-type tests like the usual t-tests and F-tests are especially vulnerable to this problem (cf. Dufour 1997). In probit models a single parameter hypothesis is usually tested by the so-called z-test, i.e. the ratio of a consistent estimate and its asymptotical standard error. This is a Wald-type test. Therefore, big size distortions can be expected. Nevertheless, the topic seems to be largely a white spot in the literature. Exceptions are the recent theoretical papers of Andrews/ Cheng (2013) and Andrews/ Cheng (2014), who address the probit model as an example. However, Andrews/ Cheng (2013) restrict their numerical analysis to a probit model with a nonlinear regression function and without endogeneity, and Andrews/ Cheng (2014) don't analyze the probit model numerically.[4]

The paper contributes different new aspects: First, large size distortions in probit models with endogeneity are demonstrated with a simulation study. Second, a new puzzle is shown: Whereas size distortions testing the null of a parameter being equal to zero are of medium size, testing other values of the parameter gives large size distortions like those in linear models. Third, the behaviour of the classical likelihood ratio statistic in that case is analysed. Given the simulation design no size distortions are found. However, undersizing is observed so that the power of the test might be low. Fourth, as a by-product some insights concerning the estimation of probit models with endogenous covariates are given.

Section 2 shows the econometric model and the test statistics. Section 3 describes the simulation design and the estimators that are used. Since a probit equation is part of the model, some formulas become more complicated than in the linear case. They are described detailedly because textbook descriptions are missing so far. Section 4 presents the results of the simulation study, and section 5 concludes. For ease of exposition always the binary probit model is considered.

2. Model and classical tests

We study a structural probit model, where one of the explaining variables is endogenous, and a reduced form equation for this variable can be specified. The specific model considered is:

$$\begin{aligned} y_{1i}^* &= \gamma_1 y_{2i} + \beta_1 x_{1i} + u_{1i} \\ y_{2i} &= \pi_{21} x_{1i} + \pi_{22} x_{2i} + v_{2i} \end{aligned} \qquad y_{1i} = \begin{cases} 1, & y_{1i}^* > 0 \\ 0, & \text{else} \end{cases}, \quad i = 1, \ldots, N, \tag{2.1}$$

where y_{1i}^* is a latent variable, y_{1i} is its observable indicator, y_{2i} is an endogenous (observable) variable, x_{1i} and x_{2i} are $(K_1 \times 1)$ and $(K_2 \times 1)$ vectors of exogenous variables, γ_1, β_1, π_{21}, π_{22} are unknown parameter vectors of dimensions 1, $(1 \times K_1)$, $(1 \times K_1)$ and $(1 \times K_2)$ respectively, and u_{1i} and v_{2i} are error terms with mean zero, variance $\sigma_{u_1}^2$ and $\sigma_{v_2}^2$ respectively, and $\mathrm{Cov}(u_{1i}, u_{1j}) = \mathrm{Cov}(v_{2i}, v_{2j}) = \mathrm{Cov}(u_{1i}, v_{2j}) = 0$ for $i \neq j$. The probit model assumes that the u_{1i}'s are normally distributed. Whether the distribution of v_{2i} must be also specified or not depends on the estimation method. If a distribution is needed a joint normal distribution of u_{1i} and v_{2i} is assumed, i.e.

$$\left(u_{1i}, v_{2i}\right) \middle| x_{1i}, x_{2i} \overset{iid}{\sim} N\left(0, \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1 v_2} \\ \sigma_{u_1 v_2} & \sigma_{v_2}^2 \end{bmatrix}\right).$$

---

[4] A further exception is Magnusson (2007), who considered in an early version of his paper the probit model with endogenous covariates as an example and found medium size distortions. However, in later versions of the working paper and in the published version (Magnusson 2010) the probit example was deleted.

The parameter of special interest is γ_1. It is not identified if π_{22} is equal to zero. Therefore, we say γ_1 is weakly identified if π_{22} is close to zero. Sometimes weak identification is quantified by the so-called concentration parameter (cf. Stock/ Wright/ Yogo 2002, p. 519). However, this parameter grows with N, and hence it suggests that the problem of weak identification is reduced by enlarging the sample size. This is misleading, and therefore we don't use it as guideline in our study.

Testing the significance of γ_1 in empirical studies is usually done by the z-statistic (implemented in almost all econometric software packages):

$$z = \frac{\hat{\gamma}_1}{\sqrt{\hat{V}(\hat{\gamma}_1)}}. \tag{2.2}$$

where $\hat{V}(\hat{\gamma}_1)$ is a consistent estimate (assuming identification) of the asymptotic variance of $\sqrt{N}(\hat{\gamma}_1 - \gamma_1)$. If $\hat{\gamma}_1$ is consistent with an asymptotic normal distribution, z is asymptotically standard normal distributed under the assumption of strong identification. The parameter γ_1 can be estimated via two step methods (cf. Blundell/ Smith 1993 for an overview) or via joint GMM or joint Maximum Likelihood (ML) estimation of both equations (cf. Wilde 2008 for a comparison).

The z-test is a Wald-type test. The classical alternatives, the Likelihood Ratio (LR) and Lagrange Multiplier (LM) test, are based on the ML estimation of the parameters. In linear models they are less susceptible to weak identification than Wald-type tests. Here, we focus on the LR test. Given the estimates, the LR-statistic is calculated easily, whereas for the LM test an estimation of the complicated information matrix is needed, and the results may depend on which estimation procedure was chosen. The loglikelihood function of model (2.1) under the standard assumptions above is:

$$\begin{aligned}
\ln l(\theta) = \sum_{i=1}^{N} \Bigg[& -0.5 \ln\left(2\Pi\sigma_{v_2}{}^2\right) - 0.5\left(\frac{y_{2i} - \pi_2 x_i}{\sigma_{v_2}}\right)^2 \\
& + y_{1i} \ln \Phi\left(\frac{1}{\sqrt{1-\rho_v{}^2}}\left(\frac{(\gamma_1\pi_{21} + \beta_1)x_{1i} + \gamma_1\pi_{22}x_{2i}}{\sigma_{v_1}} + \rho_v\left(\frac{y_{2i} - \pi_2 x_i}{\sigma_{v_2}}\right)\right)\right) \\
& + (1 - y_{1i}) \ln\left(1 - \Phi\left(\frac{1}{\sqrt{1-\rho_v{}^2}}\left(\frac{(\gamma_1\pi_{21} + \beta_1)x_{1i} + \gamma_1\pi_{22}x_{2i}}{\sigma_{v_1}} + \rho_v\left(\frac{y_{2i} - \pi_2 x_i}{\sigma_{v_2}}\right)\right)\right)\right)\Bigg],
\end{aligned} \tag{2.3}$$

where $x_i = (x_{1i}', x_{2i}')'$, $v_{1i} = u_{1i} + \gamma_1 v_{2i}$, $\theta = (\gamma_1, \beta_1, \pi_2, \sigma_{v_2}, \rho_v)'$, $\pi_2 = (\pi_{21}, \pi_{22})$, $\sigma_{v_1} = \sqrt{Var(v_{1i})}$, and $\rho_v = Corr(v_{1i}, v_{2i})$ (see Wilde 2008, appendix 2). Since the structural parameters enter the likelihood only through ratios with a standard deviation and the latter does not appear separately, only these ratios are identifiable. Therefore, in our simulation study σ_{v_1} is taken as known.

We consider the problem of testing

$$H_0: \gamma_1 = \tilde{\gamma}_1 \text{ vs. } H_1: \gamma_1 \neq \tilde{\gamma}_1.$$

We denote by $\hat{\theta}_{ML}$ the unrestricted ML estimator of θ (based on (2.3)) and by $\hat{\theta}_{RML}$ the restricted ML estimator under the null hypothesis. The LR statistic takes the form

$$LR = 2\left(\ln l\left(\hat{\theta}_{ML}\right) - \ln l\left(\hat{\theta}_{RML}\right)\right).$$

Under the usual assumptions (including strong identification) LR is asymptotically $\chi^2(1)$ distributed.

3. Design of the simulation

We choose the simulation design as simple as possible to avoid arbitrary choices of unnecessary nuisance parameters. However, ML estimation of this simple model causes numerical problems, i.e. many replications ended with the message "maximum iterations reached" even after 200 and more iterations. Therefore, we used GMM estimation for the z-test. To analyse the behaviour of the LR-test we used a slight generalization of the basic model which avoided the numerical problems on the one hand and did not change the results concerning the weak identification on the other hand (details see below).

In (2.1), the second equation is a reduced form equation. Endogeneity of y_2 can be caused by two reasons: correlation between the error terms of the structural equations and/ or simultaneity between y_1 and y_2. Both cause a correlation between u_1 and v_2. In our simulation design we focus on simultaneity because it can be interpreted more easily. Nevertheless, all results can be reproduced by assuming correlation between the error terms of the structural equations. For simplicity, in the first design $\beta_1 = \pi_{21} = 0$ and $K_2 = 1$. Therefore, our basic data generating model is:

$$\begin{aligned} y_{1i}^* &= \gamma_1 y_{2i} + u_{1i} \\ y_{2i} &= \gamma_2 y_{1i}^* + \beta_2 x_i + u_{2i} \end{aligned} \qquad y_{1i} = \begin{cases} 1, & y_{1i}^* > 0 \\ 0, & \text{elsewhere} \end{cases} \qquad i = 1, ..., N \tag{3.1}$$

The residuals u_{1i} and u_{2i} were drawn independently from a N(0, 16) distribution, i.e. the residual variances are equal for both equations. Alternatively, we also draw independently from a N(0, 1) distribution. The exogenous variable x_i was drawn from a N(0.5, 16) distribution so that the expected number of ones for y_{1i} differs from the expected number of zeros. Weak identification is equivalent to β_2 close to zero. In our simulation, we choose $\beta_2 = 0.01$. Smaller values of β_2 did not sharpen the results any more. The case of strong identification was simulated by $\beta_2 = 1$. The simulations were done for the sample sizes N = 400 (medium size) and N = 2000 (large size) and were replicated 5000 times.

The estimated model is

$$\begin{aligned} y_{1i}^* &= \gamma_1 y_{2i} + u_{1i} \\ y_{2i} &= \pi_2 x_i + v_{2i} \end{aligned} \qquad y_{1i} \text{ as defined in (3.1)}, \quad i = 1, ..., N \tag{3.2}$$

where $\pi_2 = \dfrac{\beta_2}{1-\gamma_1\gamma_2}$ and $v_{2i} = \dfrac{u_{2i} + \gamma_2 u_{1i}}{1-\gamma_1\gamma_2}$. Both parameters in (3.2) are exactly identified.

(3.2) is estimated via GMM using the “natural” moment conditions (see Wilde 2008)

$$E\begin{bmatrix} x_i\left(y_{1i} - \Phi\left(\gamma_1\pi_2 x_i / \sigma_{v_1}\right)\right) \\ x_i\left(y_{2i} - \pi_2 x_i\right) \end{bmatrix} = 0 .$$

Denoting

$$\theta = \begin{pmatrix} \gamma_1 \\ \pi_2 \end{pmatrix},\ m_i(\theta) = \begin{pmatrix} x_i\left(y_{1i} - \Phi\left(\gamma_1\pi_2 x_i / \sigma_{v_1}\right)\right) \\ x_i\left(y_{2i} - \pi_2 x_i\right) \end{pmatrix}, \text{ and } \bar{m}(\theta) = \frac{1}{N}\sum_{i=1}^{N} m_i(\theta)$$

we calculated

$$\hat{\theta}_{GMM} = \arg\min_{\theta}\left\{\bar{m}(\theta)' W_N \bar{m}(\theta)\right\}$$, W_N a weighting matrix,

with regard to θ.[5] Since the number of moment conditions is equal to the number of parameters the weighting matrix in the criterion function of the GMM estimator does not matter theoretically and the same asymptotic covariance matrix of the estimator can be used for all choices of W_N (Harris/ Mátyás 1999, p. 22). To be more precise, the asymptotic covariance matrix is (see Greene 2008, p. 445):

$$\text{asyVar}\left(\hat{\theta}\right) = \frac{1}{N}\left[G'\Psi^{-1}G\right]^{-1},$$

$$\Psi = \text{asyVar}\left(\sqrt{N}\bar{m}\right),$$

$$G = \frac{\partial \bar{m}}{\partial \theta'} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\frac{-x_i^{\,2}\pi_2}{\sigma_{v_1}}\varphi\left(\frac{\gamma_1\pi_2 x_i}{\sigma_{v_1}}\right) & \frac{1}{N}\sum_{i=1}^{N}\frac{-x_i^{\,2}\gamma_1}{\sigma_{v_1}}\varphi\left(\frac{\gamma_1\pi_2 x_i}{\sigma_{v_1}}\right) \\ 0 & \frac{1}{N}\sum_{i=1}^{N} -x_i^{\,2} \end{pmatrix},$$

Given the assumptions above this matrix can be estimated consistently via

$$\text{est asyVar}\left(\hat{\theta}\right) = \frac{1}{N}\left[\hat{G}'\hat{\Psi}^{-1}\hat{G}\right]^{-1},$$

$$\hat{\Psi} = \frac{1}{N}\sum_{i=1}^{N} m_i\left(\hat{\theta}\right) m_i\left(\hat{\theta}\right)',\ \hat{G} = G \text{ after substituting } \hat{\theta} \text{ for } \theta.$$

The square root of the first diagonal element is the denominator of the z-statistic (2.2). The nominator is calculated using the identity matrix and the estimator $\hat{\Psi}^{-1}$ of the optimal weighting matrix. The results were often the same. However, sometimes the identity matrix produces numerically more stable results, because no inverse needs to be calculated. Therefore, the reported results are based on the identity matrix.

Concerning the ML estimation of the basic model everything works fine for strong identification, whereas for weak identification the algorithm did not find the maximum for nearly half of the replications. This is caused by the following: In the basic model, for instance the second summand of (2.3) reduces to

$$y_{1i}\ln\Phi\left(\frac{1}{\sqrt{1-\rho_v^{\,2}}}\left(\frac{\gamma_1\pi_2 x_i}{\sigma_{v_1}} + \rho_v\left(\frac{y_{2i}-\pi_2 x_i}{\sigma_{v_2}}\right)\right)\right).$$

Thus, the parameter γ_1 is part of the log likelihood function only in the product $\gamma_1\pi_2$. However, in case of weak identification π_2 is close to zero, so that it is very difficult to find the "true" value of γ_1. The log likelihood function is rather flat concerning γ_1. Therefore, a constant is added in both equations and the following data generating process is considered:

$$\begin{aligned} y_{1i}^{*} &= \gamma_1 y_{2i} + \beta_{10} + u_{1i} \\ y_{2i} &= \gamma_2 y_{1i}^{*} + \beta_{20} + \beta_{21}x_i + u_{2i} \end{aligned} \qquad y_{1i} = \begin{cases} 1, & y_{1i}^{*} > 0 \\ 0, & \text{else} \end{cases}$$

where $\beta_{10} = 0.5$, $\beta_{20} = 0.25$, $x_i \sim N(0,16)$ for all replications, and u_1 and u_2 are as above. The estimated model is now

---

[5] Again, σ_{v_1} is not identified. It is replaced by the standard deviation of v_1 in each replication.

$$y_{1i}^{*} = \gamma_1 y_{2i} + \beta_{10} + u_{1i} \qquad y_{1i} = \begin{cases} 1, & y_{1i}^{*} > 0 \\ 0, & \text{else} \end{cases}$$
$$y_{2i} = \tilde{\pi}_{20} + \tilde{\pi}_{21} x_i + v_{2i}$$

with $\tilde{\pi}_{20} = \frac{\beta_{20} + \gamma_2 \beta_{10}}{1 - \gamma_1 \gamma_2}$, $\tilde{\pi}_{21} = \frac{\beta_{21}}{1 - \gamma_1 \gamma_2}$, $v_{2i} = \frac{u_{2i} + \gamma_2 u_{1i}}{1 - \gamma_1 \gamma_2}$. As expected, this avoids the numerical problems discussed above without changing the results concerning weak identification. The latter aspect was confirmed for those parameter values for which the optimum was found in the basic model.[6]

4. Simulation results

We distinguish two cases: First, we test the null hypothesis “$\gamma_1 = 0$”, i.e. the significance of y_2. Second, we test the null hypothesis “$\gamma_1 = c$”, c a constant different from zero, and present the results for c = 2. In both cases, the simulations were done for different values of γ_2. Small values of γ_2 correspond with a “small” problem of simultaneity. We choose $\gamma_2 = \pm 0.5$. A “medium” problem of simultaneity is represented by $\gamma_2 = \pm 1.5$, and a “large” problem of simultaneity is represented by $\gamma_2 = \pm 3$ and $\gamma_2 = \pm 6$. Since the differences between a medium and a large sample size were only moderate, we present the results only for N = 2000. Furthermore, we get similar results for $\sigma_{u_1} = \sigma_{u_2} = 1$ and $\sigma_{u_1} = \sigma_{u_2} = 4$ and present results for the latter.

4.1 Results for the z-test

In case of testing significance and under strong identification, no size problems can be observed (see table 1). In case of weak identification the picture is mixed.[7] If there is only “weak“ simultaneity, an extreme undersizing is observed. However, “strong” simultaneity causes medium size distortions (see figure 1). Nevertheless, the size distortions are smaller than those in linear simultaneous equations. This is surprising because weak identification should cause similar problems in probit models.

However, the results change by simulating data using $\gamma_1 = 2$ and testing H_0: $\gamma_1 = 2$. Again, in case of strong identification, no size distortion appears (see table 2). However, in case of weak identification and strong simultaneity, the size distortions became very large ($\gamma_2 = 0.5$ has to be omitted here because γ_1 is no longer identified in that case); see figure 2. The empirical size becomes more than tenfold as high as the nominal size. Thus, in probit models similar size distortions as in linear simultaneous equations models can be observed.

Table 1: Rejection frequencies of the z-test under strong identification, H_0: $\gamma_1 = 0$.

	nominal size		
γ_2	1 %	5 %	10 %
–3	0.0120	0.0470	0.0982
–0.5	0.0098	0,0502	0.1002
3	0.0094	0.0492	0.0964

---

[6] All simulations were done using R. The R codes are available on request.

[7] This result is similar to that of Magnusson (2007).

Figure 1: Rejection frequencies of the z-test under weak identification, H_0: $\gamma_1 = 0$, nominal size 5%.

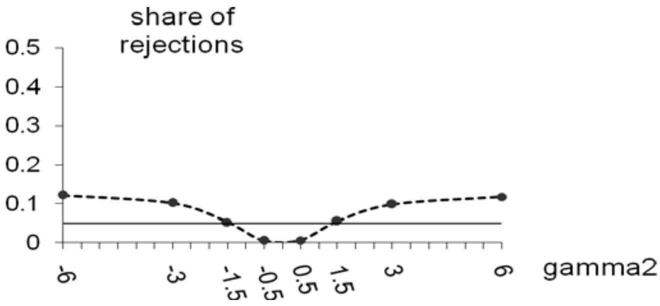
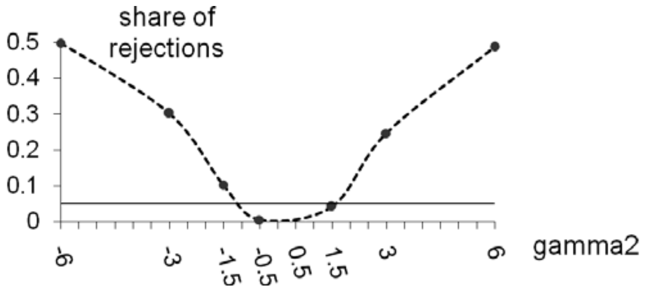


Table 2: Rejection frequencies of the z-test under strong identification, H_0: $\gamma_1 = 2$

	nominal size		
γ_2	1 %	5 %	10 %
–3	0.0110	0.0470	0.0910
–0.5	0.0056	0,0356	0.0688
3	0.0086	0.0456	0.0896

Figure 2: Rejection frequencies of the z-test under weak identification, H_0: $\gamma_1 = 2$, nominal size 5%



This demonstrates a new puzzle: In probit models the size distortion depends heavily on the parameter value tested. This puzzle cannot be solved by the concentration parameter. Following the definition of Stock/ Wright/ Yogo (2002, p. 519), in our model the concentration parameter μ^2 is:

$$\mu^2 = \left(\sum_{i=1}^{N} x_i^2\pi_2^2\right)\Big/\sigma_{v_2}^2 = \left(\sum_{i=1}^{N} x_i^2\left(\frac{\beta_2}{1-\gamma_1\gamma_2}\right)^2\right)\Bigg/\left(\frac{\sigma_{u_2}^2+\gamma_2^2\sigma_{u_1}^2}{(1-\gamma_1\gamma_2)^2}\right) = \left(\sum_{i=1}^{N} x_i^2\beta_2^2\right)\Big/\left(\sigma_{u_2}^2+\gamma_2^2\sigma_{u_1}^2\right).$$

The concentration parameter does not depend on γ_1. Thus, concerning the concentration parameter the problem of weak identification should not depend on the value of γ_1.

4.2 Results for the LR-Test

In case of testing significance and under strong identification, again no size problems can be observed (see table 3). Simulating the case of weak identification, $\beta_2 = 0.0001$ is used, i.e. a smaller value of β_2 is needed to get the strongest results. They differ substantially from those for the z-test (see figure 3). If simultaneity is weak, the observed share of rejections is near to the true size, if it is medium or strong, *under*sizing is observed. Results for $\gamma_2 = \pm 6$ are missing, because the program stopped with an error message for some replications. This message was caused by the following: Consider for instance $\gamma_1 = 0$ and $\gamma_2 = 6$. This implies $\rho_v = 0.9864$, i.e. the bivariate normal distribution of v_1 and v_2 is near to singularity. Furthermore, in (2.3) $1/\sqrt{1-\rho_v^2} = 6.08$, $\Phi(\approx 6) = 1$, $1-\Phi(\approx 6) = 0$, and $\ln(1-\Phi(\approx 6))$ is not defined. Therefore, ML estimation is less robust against a high correlation of the reduced form errors than GMM estimation.

The results for the LR-test do not change if γ_1=2 is tested. In that case γ_2=6 is possible (see figure 4). Again, in case of weak identification and strong simultaneity the empirical size is about half of the nominal size.

Table 3: Rejection frequencies of the LR-test under strong identification, H_0: $\gamma_1 = 0$

	nominal size		
γ_2	1 %	5 %	10 %
–3	0.01	0.0494	0.101
–0.5	0.012	0,0488	0.0986
3	0.0106	0.0548	0.0998

Figure 3: Rejection frequencies of the LR-test under weak identification, H_0: $\gamma_1 = 0$, nominal size 5%.

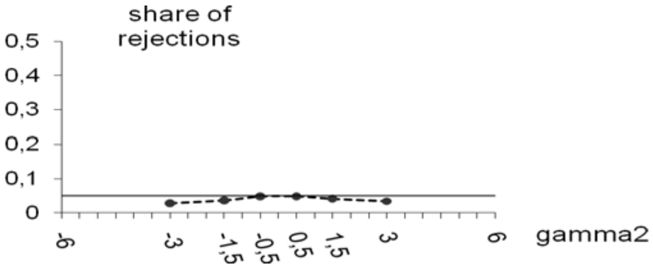
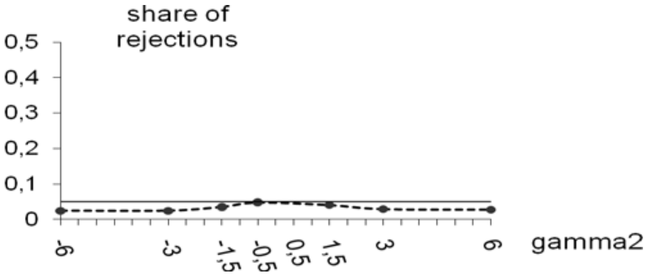


Figure 4: Rejection frequencies of the LR-test under weak identification, H_0: $\gamma_1 = 2$, nominal size 5%.



5. Conclusion

The paper analyses weak identification in probit models with endogenous covariates. It shows remarkable size distortions concerning the usual z-test. However, further research is needed to clarify why the magnitude depends heavily on the parameter value tested. The likelihood ratio statistic seems to be a conservative alternative which is robust to weak identification. Further research is useful to clarify how advanced methods like those of Andrews/Cheng (2014), Dufour (2006) or Kleibergen (2005) will work for probit models with endogenous covariates.

References

Abramitzky, R. / Lavy, V. (2014), How responsive is investment in schooling to changes in redistributive policies and in returns?, Econometrica 82, 1241–1272.

Andrews, D.W.K./ Cheng, X. (2013), Maximum likelihood estimation and uniform inference with sporadic identification failure, Journal of Econometrics 173, 36-56.

Andrews, D.W.K./ Cheng, X. (2014), GMM estimation and uniform subvector inference with possible identification failure, Econometric Theory 30, 287-333.

Beck, T./ Lin, C./ Ma, Y. (2014), Why Do Firms Evade Taxes? The Role of Information Sharing and Financial Sector Outreach, Journal of Finance 69, 763-817.

Bijsterbosch, M./ Dahlhaus, T. (2015), Key features and determinants of credit-loss recoveries, Empirical Economics 49, 1245-1269.

Bouoiyour, J./ Miftah, A./ Mouhoud, E.M. (2016), Education, Male Gender Preference and Migrants' Remittances: Interactions in Rural Morocco, Economic Modelling 57, 324-331.

Cornelli, F./ Kominek, Z./ Ljungqvist, A. (2013), Monitoring Managers: Does It Matter?**,** Journal of Finance 68, 431-481.

Croushore, D./ Marsten, K. (2016), Reassessing the Relative Power of the Yield Spread in Forecasting Recessions, Journal of Applied Econometrics 31, 1183-1191.

Dufour, J.M. (1997), Some impossibility theorems in econometrics with applications to structural and dynamic models, Econometrica 65, 1365-1387.

Dufour, J.M. (2003), Identification, weak instruments, and statistical inference in econometrics, Canadian Journal of Economics 36, 767-808.

Dufour, J.M. (2006), Monte carlo tests with nuisance parameters: A general approach to finite sample inference and nonstandard asymptotics, Journal of Econometrics 133, 443-477.

Engelhardt, G.V./ Eriksen, M.D./ Gale, W.G./ Mills, G.B. (2010), What are the social benefits of homeownership? Empirical evidence for low-income households, Journal of Urban Economics 67, 249-258.

Esaka, T. (2010), De facto exchange rate regimes and currency crises: Are pegged regimes with capital account liberalization really more prone to speculative attacks?, Journal of Banking and Finance 34, 1109-1128.

Fitzenberger, B./ Kohn, K./ Wang, Q. (2011), The erosion of union membership in Germany: determinants, densities, decompositions, Journal of Population Economics 24, 141-165.

Greene, W.H. (2008), Econometric analysis 6th ed., Prentice Hall, Upper Saddle River.

Haider, A./ Jahangir, A. (2017), La Familia - How Trust towards Family Decreases Female Labor Force Participation, Journal of Labor Research 38, 122-144.

Hao, L./ Ng, E.C.Y. (2011), Predicting Canadian recessions using dynamic probit modelling approaches, Canadian Journal of Economics 44, 1297-1330.

Harris, D./ Mátyás, L. (1999), Introduction to the generalized method of moments estimation, in: Mátyás, L. (ed.), Generalized method of moments estimation, Cambridge University Press, Cambridge, 3-30.

Hlaing, K.P./ Pourjalali, H. (2012), Economic Reasons for Reporting Property, Plant, and Equipment at Fair Market Value by Foreign Cross-Listed Firms in the United States, Journal of Accounting, Auditing & Finance 27, 557–576.

Horvath, R./ Katuscakova, D. (2016), Transparency and Trust: The Case of the European Central Bank, Applied Economics 48, 5625-5638.

Khanna, V./ Kim, E.H./ Lu, Y. (2015), CEO connectedness and corporate fraud, Journal of Finance 70, 1203-1252.

Kleibergen, F. (2005), Testing parameters in GMM without assuming that they are identified, Econometrica 73, 1103-1123.

Litchfield, J./ Reilly, B./ Veneziani, M. (2012), An analysis of life satisfaction in Albania: An heteroscedastic ordered probit model approach, Journal of Economic Behavior & Organization 81, 731-741.

Magnusson, L.M. (2007), Weak instruments robust tests for limited dependent variable models, Working Paper, Brown University (RI).

Magnusson, L.M. (2010), Inference in limited dependent variable models robust to weak identification, Econometrics Journal 13, S56-S79.

Massa, M./ Zhang, L. (2013), Monetary policy and regional availability of debt financing, Journal of Monetary Economics 60, 439–458.

Stock, J.H./ Wright, J.H./ Yogo, M. (2002), A survey of weak instruments and weak identification in Generalized Method of Moments, Journal of Business and Economic Statistics 20, 518-529

Wen, J.F./ Gordon, D.V. (2014), An empirical model of tax convexity and self-employment, Review of Economics and Statistics 96, 471-482.

Wilde, J. (2008), A note on GMM estimation of probit models with endogenous regressors, Statistical Papers 49, 471-484.