A note on GMM-estimation of probit models with endogenous regressors

Joachim Wilde

September 2005   No. 4

A note on GMM-estimation of probit models
with endogenous regressors*

Abstract

Dagenais (1999) and Lucchetti (2002) have demonstrated that the naive GMM estimator of Grogger (1990) for the probit model with an endogenous regressor is not consistent. This paper completes their discussion by explaining the reason for the inconsistency and presenting a natural solution. Furthermore, the resulting GMM estimator is analyzed in a Monte-Carlo simulation and compared with alternative estimators.

Keywords: generalized method of moments, probit model, endogenous regressor

JEL classification: C25, C35

1. Introduction

GMM estimation is useful for estimating many types of econometric models. It has also been applied to probit models ever since the pioneering work of Avery, Hansen and Hotz (1983). However, their approach assumes that all explanatory variables are exogenous. This assumption might be violated in some applications. Therefore, Grogger (1990, p. 330) considers the following model:

$$y_1^* = y_2\gamma + x_1\beta + u$$
$$y_1 = \begin{cases} 1, & y_1^* > 0 \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where y_1^* is a latent variable, y_2 is a $(1{\times}G)$ vector of endogenous and observable metric variables, x_1 is a $(1{\times}K_1)$ vector of exogenous regressors, γ and β are vectors of unknown parameters, and u is a stochastic error term with mean zero and variance σ^2, which may be normalized to one. In addition, define $z = [y_2\ x_1]$, $\delta = [\gamma'\ \beta']'$, and $w = (1{\times}(G{+}K_1))$ vector of instrumental variables which are correlated with y_2 and uncorrelated with u. The dimension $G{+}K_1$ is just a lower bound. If available, more instruments may be used.

Grogger proposed a GMM estimator using the residuals

$$\varepsilon = y_1 - F(z\delta), \tag{2}$$

F being the distribution function either of a standard normal distribution (probit model) or of a logistic distribution (logit model). However, Grogger's residuals violate the crucial assumption

$$E(w'\varepsilon) = 0. \tag{3}$$

This has been demonstrated by Dagenais (1999) in a simulation experiment. In addition, Lucchetti (2002) proved the inconsistency of Grogger's GMM estimator analytically for the special case $G{=}1$ and $K_1{=}0$.

Section 2 explains why assumption (3) is violated in the presence of endogenous variables. Based on this result a natural and consistent GMM estimator is presented in section 3. The article focuses on probit models, i.e. $F(...) = \Phi(...)$ being the distribution function of the standard normal distribution. Probit models are more flexible than logit models, and a natural GMM estimator can be derived more easily. A simulation study in section 4 compares this estimator with Grogger's GMM and standard Maximum Likelihood for different sample sizes. Section 5 explains extensions of the estimator to more complicated models and discusses the advantages in comparison with traditional two-step estimators and a recent proposal of Kawaguchi and Naito (2005). Section 6 concludes with a summary of the main results.

2. The cause of inconsistency

Avery, Hansen and Hotz (1983) assume that all regressors are exogenous, i.e. $\gamma = 0$. In that case (2) reduces to $\varepsilon = y_1 - \Phi(x_1\beta)$. Since all variables are exogenous, the second right-hand term is equal to the conditional expectation of the endogenous dummy variable: $\Phi(x_1\beta) = E(y_1|x_1)$. Therefore, $E(\varepsilon|x_1) = 0$. Using the law of iterated expectations gives:

$$E\left(x_1{}'\varepsilon\right) = E_{x_1}\left(E_{\varepsilon|x_1}\left(x_1{}'\varepsilon \mid x_1\right)\right) = E_{x_1}\left(x_1{}'E_{\varepsilon|x_1}\left(\varepsilon \mid x_1\right)\right) = 0.$$

Thus, the crucial assumption holds by definition if the residuals are the difference between the endogenous dummy variable and its conditional expectation given the instrumental variables. In other words: Choosing instrumental variables w and residuals ε with

$$\varepsilon = y_1 - E(y_1|w) \tag{4}$$

is sufficient to ensure non-correlation of ε and w and thereby consistency of the GMM estimator.

However, Grogger's residuals (2) are not equal to those in (4) even if all exogenous variables are part of w, what will be assumed in the following. The conditional expectation of y_1 is

$$E\left(y_1 \mid w\right) = P\left(u + y_2\gamma > -x_1\beta \mid w\right) \neq \Phi\left(z\delta\right). \tag{5}$$

The endogenous variable y_2 is *not* a constant given x_1 and other instrumental variables. Thus, it cannot be merged with $-x_1\beta$ as Grogger did. Instead the conditional distribution of $u + y_2\gamma$ must be determined to derive the expectation.

3. A natural GMM estimator

To simplify notation, G is set to one. The conditional distribution of $u+y_2\gamma$ may become very complicated depending on the conditional distribution of y_2. However, choosing the normal distribution leads to a feasible solution. Assume that a reduced form equation for y_2 can be specified:

$$y_2 = x\pi + v, \quad x = \left(x_1 \quad x_2\right), \quad v \mid x \quad N\left(0, \sigma_v{}^2\right), \tag{6}$$

v being an error term, π a vector of unknown parameters, and x_2 a ($1 \times K_2$) vector of exogenous variables which are uncorrelated with v and u. Furthermore, it is assumed that the conditional distribution of u does not depend on x_2, i.e.

$$u \mid x \quad N\left(0, \sigma^2\right). \tag{7}$$

For instance, this is valid if the x_2 variables are stochastically independent of u. Using (6) and (7) gives

$$u + y_2\gamma = u + v\gamma + x\pi\gamma \qquad \text{and}$$

$$\left(u + y_2\gamma\right) \mid x \quad N\left(x\pi\gamma, \sigma_{sum}^{\ 2}\right), \quad \sigma_{sum}^{\ 2} = \sigma^2 + \gamma^2\sigma_v^{\ 2} + 2\gamma Cov\left(u, v\right).$$

Following (5)

$$E\left(y_1 \mid x\right) = \Phi\left(\frac{x\pi\gamma + x_1\beta}{\sigma_{sum}}\right). \tag{8}$$

Thus, normalizing σ_{sum} to one (instead of σ) and choosing

$$\varepsilon = y_1 - \Phi\left(x\pi\gamma + x_1\beta\right) \tag{9}$$

leads to residuals which are uncorrelated with all components of x, i.e. $E(x'\varepsilon) = 0$. Thus, the crucial assumption (3) holds for $w = x$ and for $w =$ every subvector of x. In addition, further instrumental variables may be included in w that do not change the conditional expectation of y_1, i.e. $E(y_1 \mid x, \text{further variables}) = \Phi(x\pi\gamma + x_1\beta)$.

The residuals (9) contain not only the parameters of interest but also the unknown π. A possible solution is a two-step estimator: The first step is to estimate the reduced form equation (6). The second step is to calculate a GMM estimator using (9) after replacing π by its first step estimate. However, the standard errors of the second step estimator need be adjusted in order to account for the presence of the estimate of π. This problem can be avoided by a joint GMM estimation of $(\gamma, \beta')'$ and π. A joint estimation leads automatically to correct standard errors. If the parameters are just identified, i.e. the number of moment conditions is equal to the number of parameters, the results of the joint estimation are equal to the results of a two step approach with adjusted standard errors (Newey/ McFadden 1994, ch. 6.1).

To estimate π the usual moment conditions are feasible because the regressand in (6) is observable:

$$E\left(x'v\right) = E\left(x'\left(y_2 - x\pi\right)\right) = 0 \tag{10}$$

Define i = observation index (i = 1, ..., N), w_i, x_i, y_{1i}, y_{2i}, and ε_i the values of w, x, y_1, y_2, and ε for the i-th observation, $W = (w_1' \ w_2' \ \cdots \ w_N')'$, $X = (x_1' \ x_2' \ \cdots \ x_N')'$, $y_1 = (y_{11}, ..., y_{1N})'$, $y_2 = (y_{21}, ..., y_{2N})'$, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)'$. A consistent joint GMM estimator of $(\gamma, \beta')'$ and π is given by minimizing the function

$$q = \begin{bmatrix} W'\varepsilon/N \\ X'(y_2 - X\pi)/N \end{bmatrix}' A \begin{bmatrix} W'\varepsilon/N \\ X'(y_2 - X\pi)/N \end{bmatrix} \tag{11}$$

with respect to $(\gamma, \beta')'$ and π. A is a weighting matrix which may be chosen as usual to minimize the asymptotic covariance matrix of the estimator (e.g. Harris/ Mátyás 1999, p.21, Greene 2003, ch. 18.3). If the parameters are just identified, the result does not depend on the choice of A.

4. A simulation study

The properties of the natural GMM estimator and Grogger's GMM are compared in a Monte-Carlo simulation. It is based on the design of Dagenais (1999), who calculates only correlations and no estimates. He uses the following model:

$$\begin{aligned} y_{1i}^* &= y_{2i} + 4 + u_i \\ y_{2i} &= -4 + 4w_i + v_i \\ u_i &= 2v_i + v_{2i}, \end{aligned} \tag{12}$$

v_i and v_{2i} being independently drawn from a N(0, 16) distribution, and w_i is generated from a uniform distribution on [−2, 2].

Since the parameters of probit models are identified only up to a positive factor (Maddala 1983, p. 94), suitable normalizations are necessary to analyze the bias of the estimators. Concerning the natural GMM estimator the factor is equal to $1/\sigma_{sum}$ (cf. (8)). Therefore, the simulation study is based on the residuals

$$\varepsilon_{new,i} = y_{1i} - \Phi\left(\frac{4w_i}{\hat{\sigma}_{sum}}\right), \quad \hat{\sigma}_{sum} = \sqrt{\hat{\sigma}^2 + \hat{\sigma}_v^2 + 2\hat{\sigma}_{uv}}, \tag{13}$$

where $\hat{\sigma}$, $\hat{\sigma}_v$ and $\hat{\sigma}_{uv}$ are the empirical variances and the empirical covariance of the u_i's and the v_i's. Regarding Grogger's GMM estimator this aspect was not discussed before. However, his idea implicitly assumes that the factor is the same as the one for the probit model without endogenous regressors. Hence, the residuals

$$\varepsilon_{\text{grog},i} = y_{1i} - \Phi\left(\frac{y_{2i}+4}{\hat{\sigma}}\right) \tag{14}$$

are used. The instrumental variables are 1 and w_i for both estimators. Thus, there are two moment conditions for Grogger's GMM and four moment conditions for the natural GMM.

Due to the fact that the good properties of the GMM estimator are valid only asymptotically the behaviour of the estimator in small and medium samples is also analyzed. Therefore, all simulations were done for three sample sizes: N=80 (small sample), N=400 (medium sample) and N=2000 (large sample. They are replicated 1000 times. More details can be found in the LIMDEP code in the Appendix.

Table 1 contains the relevant correlations. As the means are very similar for all sample sizes, only the values for N=2000 are presented.

Table 1:
Correlations between different variables (1000 replications)

		Corr(variable 1, variable 2)	
variable 1	variable 2	mean	standarddev.
y_2	u	0.5631	0.0142
w	$y_{1i} - \Phi(y_{2i}+4)$	−0.4636	0.0132
w	$\varepsilon_{\text{grog}}$	−0.1009	0.0205
w	ε_{new}	0.0001	0.0213
w	y_2	0.7765	0.0075

The mean correlation between w and Groggers original residuals (2) is in the interval which is reported by Dagenais (1999, p. 20). It is similar to the correlation between the endogenous regressor and u. The comparison of the correlation between the original residuals and w on the one hand and the correlation between (14) and w on the other hand indicates that the correlation depends on the normalization of the explanatory variables. Finally, the correlation between the instrumental variable and the endogenous regressor is rather high. In other words, w is not a weak instrument.

Table 2 contains the estimation results. In addition to the two GMM estimators the standard probit maximum likelihood estimator (MLE) is presented, which assumes exogeneity of y_2.

Both the standard MLE and Grogger's GMME of γ and β_0 are highly biased, whereas the natural GMME behaves well for medium and large sample sizes. The bias of the estimators influences marginal effects and test statistics. For instance, for all 1000 replications the t-ratios belonging to Grogger's GMME are smaller than those of the natural GMME (N=400 and N=2000).

The last part of the table shows that the problem of bias is much smaller for the ratio of the two parameters. Further work is needed to elucidate whether this result is caused by the simulation design or by systematical reasons.

Table 2:
Means of estimations (1000 replications)

estimate	N	probit MLE (root(MSE))	naive GMME (root(MSE))	natural GMME (root(MSE))
$\hat{\gamma}$ (γ=1)	80	2.096 (1.164)	0.751 (0.370)	0.993 (0.329)
	400	2.042 (1.056)	0.743 (0.281)	0.995 (0.090)
	2000	2.040 (1.042)	0.746 (0.259)	1.001 (0.032)
$\hat{\beta}_0$ (β_0=4)	80	8.822 (5.408)	2.956 (1.983)	3.905 (2.551)
	400	8.617 (4.726)	2.956 (1.258)	3.956 (0.573)
	2000	8.622 (4.645)	2.979 (1.070)	3.994 (0.252)
$\hat{\pi}_0$ (π_0=-4)	80	-	-	-4.005 (0.459)
	400	-	-	-3.999 (0.201)
	2000	-	-	-4.001 (0.091)
$\hat{\pi}_1$ (π_1=4)	80	-	-	3.993 (0.372)
	400	-	-	3.994 (0.161)
	2000	-	-	3.999 (0.073)
$\hat{\beta}_0/\hat{\gamma}$	80	4.196 (0.796)	3.940 (4.994)	3.940 (5.013)
	400	4.218 (0.399)	3.970 (0.692)	3.969 (0.698)
	2000	4.227 (0.272)	3.990 (0.309)	3.989 (0.312)

5. Extensions and discussion

Expression (11) can be generalized easily to more than one endogenous regressor in equation (1) by adding empirical moments according to moment conditions like (10) for each additional endogenous variable. Furthermore, the GMM estimator will also be feasible if some of the endogenous variables are latent. For instance, in case binary outcomes are observed, residuals $\varepsilon_{gi} = y_{gi} - \Phi(x_i \pi_g)$ may be used to define empirical moments (g = index of the endogenous variables). Thus, the proposed GMM estimator can be applied to a wide class of probit models with endogenous regressors.

In a recent discussion paper Kawaguchi and Naito (2005) propose an alternative GMM estimator for probit models with an endogenous regressor. It is based on the conditional expectation of y_{1i} given x_i and v_{2i}. However, they consider only the case of two equations, and it is more difficult to generalize their approach to an arbitrary number of equations. Moreover, their estimator is restricted to observable endogenous regressors. Hence, it is less flexible than the proposed estimator above.

More traditional approaches are two step estimators based on the Maximum Likelihood principle. The first step estimates the reduced form equations, the second step estimates either (1) after replacing y_2 by its reduced form forecast or a conditional likelihood function (cf. Rivers/ Vuong 1988 and Blundell/ Smith 1993 for an overview and discussion). However, the standard errors of the second step need to be corrected. The procedure is well known but involves cumbersome computations (cf. Murphy/ Topel 1985 or Greene 2003, ch. 17.7). This problem is avoided by using the natural GMM estimator.

All estimators are less efficient than a joint Maximum Likelihood estimation of (1) and (6). However, for multiple equation hybrid models the implementation of a joint MLE is more complicated than that of the natural GMME. Moreover, in case three or more endogenous regressors are latent and binary observable, joint ML estimation becomes infeasible. The estimator involves the integration over the density of a multivariate normal distribution; the dimension of the integral is equal to the number of discrete endogenous variables. Such high dimensional integrals cannot be solved numerically. Simulation techniques are needed, which are computationally expensive.

Furthermore, a GMM approach is rather flexible for panel data, whereas adequate ML estimation requires tight restrictions on the error terms (cf. Lechner/ Breitung 1996, p. 583).

A crucial assumption for the natural GMME as well as for the conditional GMME of Kawaguchi and Naito is that the error terms are homoscedastic and normally distributed. Concerning heteroscedasticity a flexible GMM-based estimator has been proposed by Wilde (2005, ch. 4.4). If the normality assumption is violated, semiparametric approaches are necessary (cf. Blundell/ Powell 2004 for an overview and discussion).

6. Conclusion

The naive GMM estimator of Grogger is not consistent because the residuals are not equal to the difference between the endogenous dummy variable and its conditional expectation (given the instrumental variables). A natural GMM estimator is based on this difference. In case the error term of the original equation and the disturbances of the reduced form equations of the endogenous regressors are normally distributed, calculating the conditional expectation is straightforward. A joint GMM estimation of all equations leads not only to consistent estimators but also to correct standard errors. The approach is applicable to an arbitrary number of equations and can also be applied to latent endogenous regressors. A simulation study points out that the estimator behaves well for medium and high sample sizes, whereas the mean square error is rather high for small sample sizes.

Appendix: LIMDEP Code of the simulation study

```
sample      ;1-80$
calc        ;Ran(210465) $
create      ;wi=Rnu(-2,2) $
matrix      ;wi_80=wi $
sample      ;1-2000 $
matrix      ;wi_400=[wi_80/wi_80/wi_80/wi_80/wi_80]
            ;wi_2000=[wi_400/wi_400/wi_400/wi_400/wi_400] $
create      ;wi=wi_2000 $
create      ;obs=Trn(1,1) $
calc        ;replikat=1000 $
matrix      ;b_probit=init(Replikat,9,0) ;b_grogg=init(Replikat,9,0)
            ;b_newgmm=init(Replikat,15,0) ;korrelat=init(Replikat,15,0)
            ;t_grogg=init(Replikat,6,0) ;t_newgmm=init(Replikat,6,0) $
calc        ;Ran(310465) $
calc        ;i=1 ;gamma=1$

procedure
create      ;v1i=Rnn(0,4) $
create      ;v2i=Rnn(0,4) $
create      ;ui=2*v1i+v2i$
calc        ;Sti=2000 ;k=1 ;kk=4 ;kkk=7 ;kkkk=10 ;kkkkk=13$

do while    ;loop;Sti>=80 $
calc        ;sigm_ui=sdv(ui) ;sigm_vi=sdv(v1i) ;cov_uv=cov(ui,v1i)
            ;sigm_sum=sqr(sigm_ui^2+gamma*gamma*sigm_vi^2+2*gamma*cov_uv) $
create      ;gross_yi=-4+4*wi+v1i ;yi_stern=4+gross_yi+ui
            ;yi=0 ;if(yi_stern>0)yi=1$
create      ;gryi_nor=gross_yi/sigm_ui ;one_nor=1/sigm_ui
            ;e_grogg=yi-Phi(4+gross_yi) ;e_norm=yi-Phi(4*one_nor+gryi_nor)
            ;epsilon=yi-Phi((4+1*(-4+4*wi))/sigm_sum)$
calc        ;cor_yu=cor(gross_yi,ui) ;cor_weg=cor(wi,e_grogg)
            ;cor_wen=cor(wi,e_norm) ;cor_wep=cor(wi,epsilon)
            ;cor_wy=cor(wi,gross_yi)$
matrix      ;korrelat(i,k)=cor_yu ;korrelat(i,kk)=cor_weg
```

```
              ;korrelat(i,kkk)=cor_wen ;korrelat(i,kkkk)=cor_wep
              ;korrelat(i,kkkkk)=cor_wy$
prob          ;lhs=yi;rhs=one_nor,gryi_nor$
calc          ;b0_b1=b(1)/b(2) $
matrix        ;b_probit(i,k)=b(1) ;b_probit(i,kk)=b(2) ;b_probit(i,kkk)=b0_b1$
gmme          ;fn1=one*(yi-Phi(b0*one_nor+b1*gryi_nor))
              ;fn2=wi*(yi-Phi(b0*one_nor+b1*gryi_nor))
              ;labels=b0,b1 ;start=4,1 ;maxit=200$
calc          ;b0_b1_gr=b(1)/b(2) $
matrix        ;b_grogg(i,k)=b(1) ;b_grogg(i,kk)=b(2) ;b_grogg(i,kkk)=b0_b1_gr $
matrix        ;se_quadr=Vecd(varb) ;se_kehrw=se_quadr!-0.5 ;t_stat=dirp(b,se_kehrw)
              ;t_grogg(i,k)=t_stat(1) ;t_grogg(i,kk)=t_stat(2)$
gmme          ;fn1=one*(yi-Phi((b0+b1*(pi1+pi2*wi))/sigm_sum))
              ;fn2=wi*(yi-Phi((b0+b1*(pi1+pi2*wi))/sigm_sum))
              ;fn3=one*(gross_yi-pi1-pi2*wi)
              ;fn4=wi*(gross_yi-pi1-pi2*wi)
              ;labels=b0,b1,pi1,pi2 ;start=4,1,-4,4 ;maxit=200$
calc          ;b0_b1_ne=b(1)/b(2) $
matrix        ;b_newgmm(i,k)=b(1) ;b_newgmm(i,kk)=b(2)
              ;b_newgmm(i,kkk)=b0_b1_ne
              ;b_newgmm(i,kkkk)=b(3) ;b_newgmm(i,kkkkk)=b(4)$
matrix        ;se_quadr=Vecd(varb) ;se_kehrw=se_quadr!-0.5
              ;t_stat=dirp(b,se_kehrw)
              ;t_newgmm(i,k)=t_stat(1) ;t_newgmm(i,kk)=t_stat(2)$
calc          ;Sti=Sti/5 ;k=k+1 ;kk=kk+1 ;kkk=kkk+1 ;kkkk=kkkk+1
              ;kkkkk=kkkkk+1$
reject        ;new;obs>Sti $

enddo         ;loop $
sample        ;1-2000 $
calc          ;i=i+1$

endproc
exec          ;n=replikat;silent $
delete        ;logl_obs $
```

References

Avery R. B., Hansen L. P., Hotz V. J. (1983): Multiperiod probit models and orthogonality condition estimation. International Economic Review 24, pp. 21-35.

Blundell R. W., Powell J. L. (2004): Endogeneity in semiparametric binary response models. Review of Economic Studies 71, pp. 655-679.

Blundell R. W., Smith R. J. (1993): Simultaneous microeconometric models with censored or qualitative dependent variables, in: Maddala G. S., Rao C. R., Vinod H. D. (eds), Handbook of statistics vol. 11. North Holland, Amsterdam.

Dagenais M. G. (1999): Inconsistency of a proposed nonlinear instrumental variables estimator for probit and logit models with endogenous regressors. Economics Letters 63, pp. 19-21.

Greene W. (2003): Econometric analysis 5th ed. Prentice Hall, Upper Saddle River

Grogger J. (1990): A simple test for exogeneity in probit, logit and Poisson regression models. Economics Letters 33, pp. 329-332.

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 (UK), pp. 3-30.

Kawaguchi D., Naito H. (2005): The efficient moment estimation of the probit model with an endogenous continuous regressor, Hi-Stat Discussion Paper Series No. 106, Hitotsubashi University Tokyo.

Lechner M., Breitung J. (1996): Some GMM estimation methods and specification tests for nonlinear models, in: Mátyás L. Sevestre P. (eds), The econometrics of panel data – Handbook of the theory with applications 2nd ed. Kluwer, Dordrecht, pp. 583-612.

Lucchetti R. (2002): Inconsistency of naive GMM estimation for QR models with endogenous regressors. Economics Letters 75, pp. 179-185.

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

Murphy K. M., Topel R. H. (1985): Estimation and inference in two-step econometric models. Journal of Business & Economic Statistics 3, pp. 88-97.

Newey W. K., McFadden D. L. (1994): Large sample estimation and hypothesis testing, in: Engle R. F., McFadden D. L. (eds), Handbook of Econometrics Vol. IV. North Holland, Amsterdam, pp. 2111-2245.

Rivers D., Vuong Q. H. (1988): Limited information estimators and exogeneity tests for simultaneous probit models. Journal of Econometrics 39, pp. 347-366.

Wilde J. (2005): Determinanten des Sozialhilfebezugs – Ökonometrische Studien zu Anreizethik und Simultanität in der sozialen Grundsicherung. Habilitationsschrift, Martin-Luther-Universität Halle-Wittenberg.