UNIVERSITÀ DEGLI STUDI DI ANCONA

DIPARTIMENTO DI ECONOMIA

INCONSISTENCY OF NAIVE GMM ESTIMATION FOR QR MODELS WITH ENDOGENOUS REGRESSORS

Riccardo Lucchetti

QUADERNI DI RICERCA n. 140

Luglio 2000

Abstract

A naive GMM approach to estimating QR (logit and probit) models with endogenous explanatory variables can lead to inconsistent estimators. This result was previously shown by Dagenais via simulation. In this paper, a special case is presented for which an analytical proof is possible; it turns out that the estimator is indeed inconsistent, but the framework analysed here can be useful for hypothesis testing.

Sintesi

Un approccio GMM 'naive' per la stima di modelli QR con regressori endogeni porta a stimatori inconsistenti. Questo risultato, ottenuto via simulazione da Dagenais, viene qui provato analiticamente in un caso particolare. Si ha che lo stimatore in effetti inconsitente, ma pu essere di una qualche utilit per la prova di ipotesi.

JEL Class.: C25

Inconsistency of naive GMM estimation for QR models with endogenous regressors

Riccardo Lucchetti

1 Introduction

In 1990, Grogger [3] put forward a nonlinear instrumental variables estimator (which will be henceforth called 'naive GMM' estimator) for QR (logit and probit) models with endogenous explanatory variables. The logic behind this estimator can be summarised as follows: assume the DGP is

$$\begin{aligned} y_i &= \mathbb{I}(y_i^* > 0) & (1)\\ y_i^* &= X_i\beta + u_i & (2) \end{aligned}$$

where $\mathbb{I}(A)$ equals 1 if A is true and 0 if A is false.

Assuming the sample observations are independent and identically distributed, the log-likelihood for the i-th observation can be written as

$$\mathscr{L}_i = y_i \ln F(X_i\beta) + (1 - y_i) \ln\left[1 - F(X_i\beta)\right], \tag{3}$$

where $F(\cdot)$ is the distribution function of u_i; the i-th contribution to the score is therefore

$$s_i = \frac{f_i}{F_i(1 - F_i)} \left[(y_i - F_i)X_i\right] = \omega_i e_i X_i, \tag{4}$$

where the shorthand notation $f_i = f(X_i\beta)$ and $F_i = F(X_i\beta)$ is used; ω_i is the "weight" and e_i is the "residual".

If $\mathsf{E}\{s_i|X_i\} = 0$, it follows that $\mathsf{E}\{(y_i - F_i)|X_i\}$ must be 0 as well. The latter expression can be thought of as a moment-generating equation in a GMM strategy[1]. If there is correlation between X_i and u_i, then one may conjecture that an "instrumental variables" version of the orthogonality condition could be used; if some W_i variable — correlated with X_i but not with

[1] In the special case of the logit model, $\frac{f_i}{F_i(1-F_i)} = 1$, so that ML and this particular GMM estimator coincide.

u_i — was available, estimation could proceed from the modified version of the orthogonality condition, namely:

$$\sum_{i=1}^{n} W_i(y_i - F_i) = 0 \tag{5}$$

Consider therefore the estimator

$$\hat{\beta} = \underset{\beta}{\text{Argmin}} \sum_{i=1}^{n} (y_i - F_i)' P_W (y_i - F_i), \tag{6}$$

which is the Grogger (NGMM) estimator. However, this estimator was proven to be inconsistent by Dagenais [1] due to the nonlinearity of $F(\cdot)$. The proof was achieved via simulation. In this paper, a case will be analysed in which an analytical proof is possible; it turns out that the estimator is indeed inconsistent, but the framework analysed here can be useful for hypothesis testing.

2 The DGP

In order to keep matters tidy and concentrate on the final result, the simplest possible model will be analysed here: a probit model with no exogenous regressors, one endogenous regressor and one instrument; both variables will be assumed to be normally distributed.

Assume, therefore, that the sample contains three observable variables: an explanatory variable X_i, a qualitative variable y_i (which can only be 0 or 1) and an instrumental variable W_i; the latter is a zero-mean random variable with variance σ_W^2.

$$\begin{aligned} y_i &= \mathbb{I}(y_i^* > 0) && (7) \\ y_i^* &= X_i\beta + u_i && (8) \\ X_i &= \alpha W_i + \eta_i && (9) \end{aligned}$$

It is also assumed that the two disturbances η_i and u_i are jointly normal, and independent of W_i.

$$\left(\begin{array}{c} \eta_i \\ u_i \end{array} \right) \Bigg| W_i \sim N \left[\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma_\eta^2 & \theta\sigma_\eta \\ \theta\sigma_\eta & 1 \end{array} \right) \right]$$

It is the normality of u_i makes this model a probit model[2]. As customary to achieve identification, the variance of u_i is set to 1. The two disturbances

[2] In principle, the logit case could also be analysed within the same framework as here; however most of the results presented in this paper depend on certain characteristics of the normal distribution which are not easily generalised.

may be correlated (when the parameter θ is not 0). If they are, then X_i becomes "endogenous" for the the parameter of interest β, in the sense that the expected value of y_i^* conditional on X_i is different from $X_i\beta$.

The distribution of X_i and y_i^* conditional on W_i is therefore

$$\left.\begin{pmatrix} X_i \\ y_i^* \end{pmatrix}\right| W_i \sim N\left[\begin{pmatrix} \alpha W_i \\ \alpha\beta W_i \end{pmatrix}\begin{pmatrix} \sigma_\eta^2 & \beta\sigma_\eta(1+\theta) \\ \beta\sigma_\eta(1+\theta) & \beta^2\sigma_\eta^2 + 2\beta\theta\sigma_\eta + 1 \end{pmatrix}\right]$$

3 The main result

The estimator $\hat{\beta}$, defined in equation (6), converges in probability to β if $\mathsf{E}\{W_i e_i\} = 0$. In this case, however, it is not so, as will be shown presently. In fact,

$$\mathsf{E}\{W_i e_i\} = \mathsf{E}\{W_i y_i\} - \mathsf{E}\{W_i \Phi(X_i\beta)\},$$

and it suffices to prove that the two subexpressions to the right of the equal sign differ.

In order to evaluate the relevant expected values, we will use the following lemmas, proofs of which are in the Appendix:

Lemma 1 *Let $X \sim N(0, \sigma^2)$. Then, for any constant a and b, it holds that* $\mathsf{E}\{\Phi(a + bX)\} = \Phi\left(\frac{a}{\sqrt{1+\sigma^2 b^2}}\right)$.

Lemma 2 *Let $X \sim N(0, \sigma^2)$. Then, for any constant λ, it holds that* $\mathsf{E}\{X\Phi(\lambda X)\} = \frac{1}{\sqrt{2\pi}} \frac{\lambda\sigma^2}{\sqrt{1+\lambda^2\sigma^2}}$.

Lemma 3 *Let X and ϵ be jointly distributed as*

$$\begin{pmatrix} X \\ \epsilon \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_\epsilon^2 \end{pmatrix}\right].$$

Then, for any constant k, it holds that $\mathsf{E}\{X\Phi(kX + \epsilon)\} = \frac{1}{\sqrt{2\pi}} \cdot \frac{k\sigma_X^2}{\sqrt{1+k^2\sigma_X^2+\sigma_\epsilon^2}}$.

By use of these lemmas, the two quantities $\mathsf{E}\{W_i y_i\}$ and $\mathsf{E}\{W_i \Phi(X_i\beta)\}$ will be shown to be different from each other. In order to evaluate $\mathsf{E}\{W_i y_i\}$, define ϵ_i as

$$\epsilon_i = y_i^* - \mathsf{E}\{y_i^*|W_i\} = u_i + \beta\eta_i; \tag{10}$$

so that $\sigma_\epsilon^2 = \beta^2\sigma_\eta^2 + 2\beta\theta\sigma_\eta + 1$. The expected value of $W_i y_i$ can therefore be written as

$$\mathsf{E}\{W_i y_i\} = \mathsf{E}_W\{W_i \mathsf{E}\{y_i|W_i\}\} = \mathsf{E}_W\{W_i P[y_i^* > 0|W_i]\} = \mathsf{E}_W\left\{W_i \Phi\left(\frac{\alpha\beta W_i}{\sigma_\epsilon}\right)\right\}$$

where the second equality comes from the fact that y_i is a Bernoulli random variable, and its mean is equal to the probability of it being 1. Lemma 2 makes it possible to evaluate the rightmost expression as

$$\mathsf{E}_W \left\{ W_i \Phi\left(\frac{\alpha\beta}{\sigma_\epsilon} W_i\right)\right\} = \frac{\sigma_W^2 \alpha\beta}{\sqrt{2\pi}\sigma_\epsilon} \cdot \left[1 + \left(\frac{\sigma_W \alpha\beta}{\sigma_\epsilon}\right)^2\right]^{-1/2},$$

and so, after some simplifications, we obtain

$$\mathsf{E}\left\{W_i y_i\right\} = \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma_W^2 \alpha\beta}{\sqrt{\sigma_\epsilon^2 + \sigma_W^2 \alpha^2 \beta^2}}. \tag{11}$$

The expression $\mathsf{E}\left\{W_i \Phi(\beta X_i)\right\}$, on the other hand, can be computed by starting from the equality

$$\mathsf{E}\left\{W_i \Phi(\beta X_i)\right\} = \mathsf{E}\left\{W_i \Phi\left[\beta(\alpha W_i + \eta_i)\right]\right\}.$$

Since W_i and η_i are independent by hypothesis, by applying lemma 3 we get

$$\mathsf{E}\left\{W_i \Phi\left[\beta(\alpha W_i + \eta_i)\right]\right\} = \frac{1}{\sqrt{2\pi}} \cdot \frac{\alpha\beta\sigma_W^2}{\sqrt{1 + \sigma_W^2 \alpha^2 \beta^2 + \beta^2 \sigma_\eta^2}}. \tag{12}$$

The final result is that

$$\mathsf{E}\left\{W_i e_i\right\} = \frac{\sigma_W^2 \alpha\beta}{\sqrt{2\pi}} \left[\frac{1}{\sqrt{\sigma_\epsilon^2 + \sigma_W^2 \alpha^2 \beta^2}} - \frac{1}{\sqrt{1 + \sigma_W^2 \alpha^2 \beta^2 + \beta^2 \sigma_\eta^2}}\right]$$

Given the definition of ϵ_i (see equation (10)), $\mathsf{E}\left\{W_i e_i\right\}$ is null if $\beta\theta\sigma_\eta = 0$, i. e. if at least one of the three quantities involved is 0. Let us examine the three cases separately:

$\theta = 0$ In this case there is no endogeneity, since X_i is independent from u_i; the whole point of using an instrumental variable therefore collapses.

$\sigma_\eta = 0$ In this case there is no endogeneity either, because X_i is a simple multiple of W_i, which is independent from u_i by hypothesis.

$\beta = 0$ This case is the most interesting, since in this case $\hat{\beta}$ converges to its true value of 0; moreover, all conditions for asymptotic normality are satisfied (see eg [2]), so that it is possible to test the hypothesis $\beta = 0$ by considering $\hat{\beta}$ suitably standardised. If this is all that is needed, then the use of the NGMM estimator is perfectly acceptable.

4 Conclusions

The use of instrumental variables in QR models should be handled with care. In fact, the intrinsic nonlinearity of such models makes the analogy with the familiar linear models misleading, so that even a 'plain vanilla' GIVE estimator can be shown to be inconsistent. However, such a method could have some limited usefulness as long as the researcher's aim is not to estimate a parameter, but only to test if its true value is 0.

References

[1] Dagenais, Marcel G., (1999) "Inconsistency of a proposed nonlinear instrumental variables estimator for probit and logit models with endogenous regressors", *Economics Letters*, **63**, p. 19-21

[2] Gourieroux, Christian and Monfort, Alain (1989), *Statistique et modèles économétriques*, Économica (English translation: *Statistics and Econometric Models*, (1995), Cambridge University Press)

[3] Grogger, Jeffrey, (1990) "A simple test for exogeneity in probit, logit, and poisson regression models", *Economics Letters*, **33**, p. 329-332

A Proofs

Proof of Lemma 1: Consider first the case where $\sigma^2 = 1$. Define $E[\Phi(a + bX)] = h(a)$, as a function of a. Since $h(a)$ is everywhere continuous and differentiable, it holds that

$$h'(a) = E[\phi(a + bX)] = \int_{\mathbb{R}} \phi(a + bX)\phi(X)dX,$$

where $\phi(\cdot)$ is the density of a standard normal r.v. It is easily shown that

$$\phi(a + bX)\phi(X) = \phi\left(\frac{a}{\sqrt{1 + b^2}}\right)\phi\left(\sqrt{1 + b^2}X + \frac{ab}{\sqrt{1 + b^2}}\right).$$

By a change of variable $Z = \sqrt{1 + b^2}X + \frac{ab}{\sqrt{1+b^2}}$, it is trivial to show that

$$\int_{\mathbb{R}} \phi\left(\sqrt{1 + b^2}X + \frac{ab}{\sqrt{1 + b^2}}\right) dX = \frac{1}{\sqrt{1 + b^2}},$$

so that

$$h'(a) = \frac{1}{\sqrt{1 + b^2}}\phi\left(\frac{a}{\sqrt{1 + b^2}}\right).$$

Integrating the previous expression yields

$$h(a) = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right) + K,$$

where K is a constant to be determined. Note, however that $h(0) = 1/2$ for any b (this follows easily from the fact that the function $[\Phi(\cdot) - 1/2]$ is bounded and hemisymmetric for any b), so K turns out to be 0. The proof for the general case where $\sigma^2 \neq 1$ follows trivially by setting $X = \sigma X^*$ (so that $X^* \sim N(0,1)$) and $b^* = \sigma b$.

Proof of Lemma 2: Consider first the case where $\sigma^2 = 1$. Since

$$\frac{\mathrm{d}\,\phi(X)}{\mathrm{d}\,X} = -X\phi(X),$$

it is possible to write

$$E[X\Phi(\lambda X)] = -\int_{\mathbb{R}} \Phi(\lambda X)\frac{\mathrm{d}\,\phi(X)}{\mathrm{d}\,X}dX,$$

which can be easily integrated by parts. In order to do so, notice that

$$\phi(\lambda X)\phi(X) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(1+\lambda^2)X^2\right\}$$

and therefore, by a suitable change of variable,

$$\int_{\mathbb{R}} \phi(\lambda X)\phi(X)dX = \frac{\lambda}{\sqrt{(1+\lambda^2)2\pi}}.$$

Integration by parts gives the desired result after noting that

$$\lim_{X\to\pm\infty} \Phi(\lambda X)\phi(X) = 0$$

for any value of λ, due to the boundedness of $\Phi(\cdot)$. The proof for the general case where $\sigma^2 \neq 1$ follows trivially by setting $X = \sigma X^*$ (so that $X^* \sim N(0,1)$) and $\lambda^* = \sigma\lambda$.

Proof of Lemma 3:

$$\mathsf{E}\left\{X\Phi(kX+\epsilon)\right\} = \mathsf{E}_X\left\{X\,\mathsf{E}\left\{\Phi(kX+\epsilon)|X\right\}\right\}$$

Conditional on X, the expression $\mathsf{E}\left\{\Phi(kX+\epsilon)|X\right\}$ can be evaluated via lemma 1as

$$\mathsf{E}\left\{\Phi(kX+\epsilon)|X\right\} = \Phi\left(\frac{kX}{\sqrt{1+\sigma_\epsilon^2}}\right).$$

Therefore,

$$\mathsf{E}\left\{X\Phi(kX+\epsilon)\right\} = \mathsf{E}_X\left\{X\,\Phi\left(\frac{kX}{\sqrt{1+\sigma_\epsilon^2}}\right)\right\}$$

which, by lemma 2, equals

$$\frac{1}{\sqrt{2\pi}}\cdot\frac{k\sigma_X^2}{\sqrt{1+k^2\sigma_X^2+\sigma_\epsilon^2}},$$

as claimed.