Estimating variances and covariances in a censored regression model

Giorgio Calzolari and Gabriele Fiorentini

Universita' di Firenze

1993

ESTIMATING VARIANCES AND COVARIANCES IN A CENSORED REGRESSION MODEL

Giorgio Calzolari, Gabriele Fiorentini

1. INTRODUCTION

In empirical work we often encounter models where the dependent variables are limited in their range. Although their use is perhaps more frequent in the microeconometric analysis of survey data, also time series models provide examples of this kind. As well pointed out by Maddala in the introductory notes to his book (1983, p. 1), it is not always necessary to introduce the complications implied by this type of models. "For instance, if we believe that prices are necessarily positive, we might postulate that they have a log-normal distribution rather than the normal. On the other hand, in the limited-dependent-variable models discussed in this book, the variables are limited to their range because of some underlying stochastic choice mechanism".

The case considered in this paper is the so called *censored regression model*.

The first application of this model to economic problems was proposed in a pioneering work by Tobin (1958). Analysing demand for durable goods, he observed a concentration of observations around zero, as most households report zero expenditure on automobiles or other durable goods during any year. The linearity assumption underlying the regression model was clearly inappropriate, and some suitable form of discontinuity had to be introduced. This led to a censored regression model like

$$y_t^* = x_t'\beta + u_t \qquad \begin{cases} y_t = y_t^* & \text{if RHS} > 0 \\ y_t = 0 & \text{if RHS} \leq 0 \end{cases} \tag{1}$$

where y_t^* is a nonobservable random variable, x_t is the vector of exogenous explanatory variables at time t, β is the vector of unknown coefficients, y_t is the observed *censored* value of the dependent variable. Given the strict connection with the literature on probit's models, to synthetyse in one word the concept "Tobin's probit", Goldberger (1964) introduced the term *Tobit*.

Financial support from MURST 40% funds is also gratefully acknowledged.

To distinguish it from its various generalisations, model (1) is usually referred to as the *standard* Tobit model. Here the threshold is zero, but we may easily generalise the model to accomplish for nonzero thresholds and even thresholds varying from individual to individual. It is easy to imagine a model where the threshold acts on the opposite side, like an upper bound rather than a lower bound. For example (Green, 1985), if we investigate the relationship between the sale of tickets and the price of tickets for a match in the stadium, we should expect a decreasing function with an upper threshold given by the maximum capacity of the stadium.

Other applications of the censored regression model to economic problems are illustrated in Maddala (1983, pp. 4-6). Consider for example "a change in a household's holding of liquid assets during a year. The variable... cannot be smaller than the negative of the household's assets at the beginning of the year, because one cannot liquate more assets than one owns". Again, other examples may be related to the labour market. For instance, a married woman participates in the labour force if the market wage is greater than her valuation of time in the household (*i.e.* the reservation wage), otherwise she does not participate.

Ordinary least squares estimation method is clearly inappropriate for these types of models; it produces in fact inconsistent estimates of the parameters β and σ^2 (see, for instance, Dhrymes, 1986).

A number of alternative estimators for β and σ^2 exist, and their properties have been investigated under a number of alternative conditions. Heckman (1976) proposed a *two-stage* estimation method. Powell (1983) introduced a *least absolute deviation* estimator. Fishe, Maddala and Trost (1979), Arabmazar and Schmidt (1981) investigated the problem of heteroskedasticity, Robinson (1982) analysed problems connected with serial correlation, while Arabmazar and Schmidt (1982), Nelson (1981), Newey (1987), Ruud (1986) and Smith (1987) discussed problems connected with the non-normality of the error terms. Also the problem of limited dependent variables in a system of simultaneous equations have been treated in the literature; see, for example, Amemiya (1983), Flood and Tasiran (1989), Nelson and Olson (1978), Sickles and Schmidt (1978).

Since our study has a very specific purpose, we shall focus on one estimation method (maximum likelihood) and one type of model, the standard Tobit model (1), under classical assumptions.

2. THE LIKELIHOOD FUNCTION

Let

Φ_t = cumulated distribution of a standard normal evaluated at $\dfrac{x_t'\beta}{\sigma}$

ϕ_t = probability density of a standard normal evaluated at $\dfrac{x_t'\beta}{\sigma}$

Σ_0 = summation referring to the zero observations
Σ_1 = summation referring to non-zero observations
T_1 = number of nonzero observations.

Assuming independent identically distributed normal error terms u_t, the log-likelihood function is given by

$$\log L(\beta, \sigma) = \sum_0 \log(1-\Phi_t) - \frac{T_1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_1 (y_t - x_t'\beta)^2 . \tag{2}$$

Unlike other types of models whose likelihood function may present multiple maxima, in our case the likelihood function is globally concave (Olsen, 1978).

Maximisation of the log-likelihood (2) provides an estimate of the unknown parameters $\hat\beta$ and $\hat\sigma^2$ which is consistent and asymptotically efficient under suitable regularity conditions.

3. ESTIMATORS OF THE COVARIANCE MATRIX

Several different estimators of the covariance matrix of $\hat\beta$ and $\hat\sigma^2$ can now be used. They are numerically different, but asymptotically equivalent. Equivalence rests upon the property that, under correct specification of the model and suitable regularity conditions, all these covariance estimators asymptotically give the inverse of Fisher's information matrix.

These different estimators are typically – even if not necessarily – associated with different computer algorithms (see Hall, 1984, for a survey on available computer programs).

Let us suppose to use some *Newton-like* method to maximise the likelihood. At each iteration of the process, we compute the vector of first order derivatives of the log-likelihood (gradient) and the matrix of second order derivatives (Hessian). Upon convergence of the iterative maximisation process, the inverse of the Hessian calculated in the last iteration is a suitable estimate of the covariance matrix for $\hat\beta$ and $\hat\sigma^2$.

The first derivatives of the log-likelihoods are (Amemiya, 1984)

$$\frac{\partial \log L}{\partial \beta} = -\frac{1}{\sigma}\sum_0 \frac{\phi_t x_t}{1-\Phi_t} + \frac{1}{\sigma^2}\sum_1 (y_t - x_t'\beta)x_t \tag{3}$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2\sigma^3}\sum_0 \phi_t \frac{x_t'\beta}{1-\Phi_t} - \frac{T_1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_1 (y_t - x_t'\beta)^2 \tag{4}$$

while the blocks of the Hessian matrix are

$$\frac{\partial^2 \log L}{\partial\beta\partial\beta'} = -\frac{1}{\sigma}\sum_0 \frac{\phi_t}{(1-\Phi_t)^2}\left[\frac{\phi_t}{\sigma} - \frac{1}{\sigma^2}(1-\Phi_t)x_t'\beta\right]x_t x_t' - \frac{1}{\sigma^2}\sum_1 x_t x_t' \tag{5}$$

$$\frac{\partial^2 \log L}{\partial\sigma^2\partial\beta'} = -\frac{1}{2\sigma^3}\sum_0 \frac{\phi_t}{(1-\Phi_t)^2}\left[\frac{1}{\sigma^2}(1-\Phi_t)(x_t'\beta)^2 - (1-\Phi_t) - \frac{x_t'\beta\phi_t}{\sigma}\right]x_t'$$
$$- \frac{1}{\sigma^4}\sum_1 (y_t - x_t'\beta)x_t' \tag{6}$$

$$\frac{\partial^2 \log L}{\partial(\sigma^2)^2} = -\frac{1}{4\sigma^5}\sum_0 \frac{\phi_t}{(1-\Phi_t)^2}\left[\frac{1}{\sigma^2}(1-\Phi_t)(x_t'\beta)^3 - 3(1-\Phi_t)x_t'\beta - \frac{(x_t'\beta\phi_t}{\sigma}\right]$$
$$+ \frac{T_1}{2\sigma^4} - \frac{1}{\sigma^6}\sum_1 (y_t - x_t'\beta)^2 . \tag{7}$$

If our program takes advantage of the Berndt, Hall, Hall and Hausman's (1974) suggestions, we can replace the Hessian with a matrix of outer products of the first derivatives of the log-likelihoods, both in the maximisation process and for estimating variances and covariances of $\hat{\beta}$ and $\hat{\sigma}^2$.

With the method of scoring the inverse of the estimated information matrix is used. Its blocks are (Amemiya, 1985, p. 373)

$$E\left(\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right) = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(\phi_t\frac{x_t'\beta}{\sigma} - \frac{\phi_t^2}{(1-\Phi_t)} - \Phi_t\right)x_t x_t' \tag{8}$$

$$E\left(\frac{\partial^2 \log L}{\partial\sigma^2\partial\beta'}\right) = -\frac{1}{2\sigma^3}\sum_{t=1}^{T}\left(\phi_t\frac{(x_t'\beta)^2}{\sigma^2} + \phi_t - \frac{\phi_t^2}{(1-\Phi_t)}\frac{x_t'\beta}{\sigma}\right)x_t' \tag{9}$$

$$E\left(\frac{\partial^2 \log L}{\partial(\sigma^2)^2}\right) = -\frac{1}{4\sigma^4}\sum_{t=1}^{T}\left(\phi_t\frac{(x_t'\beta)^3}{\sigma^3} + \phi_t\frac{x_t'\beta}{\sigma} - \frac{\phi_t^2}{(1-\Phi_t)}\frac{x_t'\beta}{\sigma} - 2\Phi_t\right). \tag{10}$$

Maximisation can also be performed in some other way (*e.g.* Fair, 1977). In this case, estimates of variances and covariances of $\hat{\beta}$ and $\hat{\sigma}^2$ are obtained by calculating one of the above three matrices at the end of the maximisation process.

Also not associated with a particular maximisation algorithm is the robust estimator of the coefficient covariance matrix, whose use has become more and more popular in the last few years. This estimator was discussed in White (1980) for the linear regression model with heteroskedastic errors, then extended in White (1982, 1983) and Gourieroux *et al.* (1984) to cover more general types of models. It gives the covariance matrix of the parameters when the model is not correctly specified (misspecification consistent). Under correct specification, also this estimator is equivalent to the others, as it gives asymptotically the inverse of the Fisher's information matrix.

Let us now suppose to estimate a model in practice by maximum likelihood and to calculate, upon convergence of the maximisation algorithm, the four different estimates of the variance-covariance matrix. Of course we are perfectly aware that these covariance estimators are equivalent only for large

samples. However, as the four matrices are computed at the same parameters values, we would probably expect that even for a small sample the four groups of results had to be sufficiently close to one another, and that if significant differences are found, they had to be interpreted as an indication of misspecification (*e.g.* White 1982). This seems not to be true, unless the sample size is really very large. Even in absence of misspecification, large differences are encountered in small samples, and the sign of the differences is almost systematic.

Similar results have appeared in the literature, related to Probit model (Griffiths *et al.*, 1987), to simultaneous equations (Calzolari and Panattoni, 1988a), and to the linear regression model (Parks and Savin, 1990).

Other recent studies have been focused on the finite sample behaviour of the robust covariance estimator, all evidencing that it *tends to underestimate* the coefficient covariance matrix if the model is correctly specified. Chesher and Jewitt (1987) identify conditions under which this covariance estimator is downward biased in the linear regression model (the *heteroskedasticity consistent* covariance estimator). They show that the bias critically depends on the regression design and can be severe. MacKinnon and White (1985) propose some finite sample corrections for this covariance estimator in linear regressions, while the sampling experiments in Prucha (1984) or in Calzolari and Panattoni (1988b) clearly evidence a similar need for systems of simultaneous equations. Like all these papers, also here we use a Monte Carlo experiment to investigate the finite sample properties of the four covariance matrix estimators in the context of the standard Tobit model.

4. SIMULATION RESULTS

In this section, which is the main part of the paper, we report the results of the simulation experiments. First the design of the Monte Carlo is outlined, then the specification of the models employed in the experiments are described and finally results are discussed.

4.1. *Design of the Monte Carlo Experiments*

To compare the different estimators of the variances and covariances in small-medium sized samples, a wide set of Monte Carlo experiments have been performed on several models. For each model we start from a given vector of *true* parameters held fixed over all replications, we fix a sample period lenght and generate values of the explanatory variables in different ways:

- models 1, 2, 4 and 5: explanatory variables have been kept fixed at their historical values and for longer samples have been repeated consecutively.
- model 3: a multivariate normal generator, with given means and covariance matrix has been used; explanatory variables in different time period were generated independently; dummy variables were randomly generated with 0 or 1 values.

- model 6: the explanatory exogeneous variables are generated with a strong leptokurtic design.

Random values of the disturbance terms, u_t, over the sample period, are then generated independently of the explanatory exogeneous variables with normal distribution, zero mean, the given variance, and independent over time. Values of the endogeneous variable are finally computed with simulation. With the generated data we now estimate parameters by maximum likelihood and we compute, at the values that maximise the likelihood ($\hat{\beta}$ and $\hat{\sigma}^2$), the four estimates of the covariance matrix of $\hat{\beta}$ and $\hat{\sigma}^2$.

For each model, given a sample period lenght, we perform 1000 replications of the Monte Carlo process. A table displays the mean value of each parameter estimates, the mean variance estimated with the four methods, and the mean squared error ratios (ratios between each mean variance estimate and the Monte Carlo M.S.E.).

The last column displays the number of times that we found an outer product estimated variance greater than the corresponding Hessian estimate. In our experiments this is the inequality that occurs more often.

The experiment is then repeated with different sample lenghts. Rather than displaying detailed results for each experiment, we summarize, for groups of parameters, the percentage number of cases in which the variance computed from the matrix of outer products is greater than the corresponding variance computed from the Hessian matrix.

4.2. *Models Specifications*

MODEL 1. From the monthly monetary model of the Bank of Italy (Banca d'Italia, 1988, p. 79), we consider the equation of discount window borrowing.

$$Y_t = 1000c_0 + c_1 Y_{t-1} + 100c_2 X_{1,t} + 10c_3 X_{2,t} + c_4 X_{3,t} + c_5(X_{4,t} - X_{4,t-1}) + u_t \quad (11)$$

Y = Discount window borrowing
X_1 = Overnight rate
X_2 = Discount rate
X_3 = Monetary base of Treasury net of open-market operation (flows)
X_4 = Monetary base abroad

$T = 42$ Truncated observations: 26% 1000 Replications

MODEL 2. A different specification of the previous model has been considered, so model 2 is a simplified version of the equation for demand of discount window borrowing (Banca d'Italia, 1988, p. 79)

$$Y_t = 1000c_0 + 100c_1(X_{1,t} - X_{2,t}) + c_2 X_{3,t} + c_3(X_{4,t} - X_{4,t-1}) + u_t \quad (12)$$

$T = 42$ Truncated observations: 26% 1000 Replications

MODEL 3. This model is a simplified version of the equation proposed by Witte (1980) as an economic model of crime.

$$Y_t = \frac{c_0}{10} + c_1 \frac{X_{1,t}}{10000} + c_2 \frac{X_{2,t}}{100} + c_3 \frac{X_{3,t}}{100} + c_4 \frac{X_{4,t}}{100} + c_5 \frac{X_{5,t}}{10} + c_6 \frac{X_{6,t}}{10} + u_t \quad (13)$$

Y = Number of arrests per month free
X_1 = Accumulated work release funds received
X_2 = Number of months until first job after release
X_3 = Hourly wage after release
X_4 = Age (in years) at release
X_5 = Dummy (1 if serious alcohol or drug problem)
X_6 = Dummy (1 if married)

$T = 150$ Truncated observations: 30% 1000 Replications

MODEL 4. A naive version of Mroz's model for married women labour supply (Mroz, 1987)

$$Y_t = 100c_0 + 100c_1X_{1,t} + c_2X_{2,t} + c_3X_{3,t} + c_4X_{4,t} + c_5X_{5,t} + c_6X_{6,t} + c_7X_{7,t} + c_8X_{8,t} + c_9X_{9,t} + c_{10}\frac{X_{10,t}}{100} + u_t \quad (14)$$

Y = Wife's hours of work
X_1 = Children less than 6 years old in the household
X_2 = Chlidren between ages 6 and 18
X_3 = Wife's age
X_4 = Wife's educational attainment, in years
X_5 = Husband's age
X_6 = Wife's father's educational attaiments
X_7 = Unemployment rate in county of residence
X_8 = Dummy (1 if live in large city)
X_9 = Years of wife's previous labor market experience
X_{10} = Wife's property income

$T = 200$ Truncated observations: 60% 1000 Replications

MODEL 5. A simplified version of Mroz's model for married women labour supply (Mroz, 1987)

$$Y_t = 100c_0 + 100c_1X_{1,t} + c_2X_{2,t} + c_3X_{4,t} + c_4X_{9,t} + c_5\frac{X_{10,t}}{100} + u_t \quad (15)$$

$T = 753$ Truncated observations: 43% 1000 Replications

4.3. *Inequalities on estimated variances*

While for models 4 and 5 only one set of experiments has been performed , for the first three models the experiments have been repeated with longer sample periods. In table 6 we display in a synthetic way the results related to the inequality between the Hessian and the O.P. estimates of the variances for models 1, 2 and 3.

As far as the QML estimator of the variances is concerned we observe in tables 1 to 5, that the QML variance estimate is, on average, always slightly smaller than the corresponding Hessian estimate.

4.4. *Artificial Models 6 and 7*

Two artificial models have been specified and simulated to investigate some behaviors of the alternative estimators under extreme conditions.

Model 6 has 10 explanatory exogeneous variables as well as the constant term (12 parameters, including σ^2). All values of the explanatory variables have been randomly generated with a strong leptokurtic design. These conditions were identified as critical in other type of models (*e.g.* Calzolari and Panattoni, 1988b). Four different sample lenghts have been used for simulation: $T = 100$, $T = 200$, $T = 1000$ and $T = 5000$. For this model, some results are presented in graphical form, as described in the next section.

Model 7 is a very simple one, with a single constant regressor, like one of the models used by Nelson and Savin (1988) to investigate the nonmonotonic behavior of the power in the Wald test for Tobit models.

TABLE 1

Model 1: mean estim. parameters, mean estim. variances, and M.S.E. ratios

par.	*True*	Est.	Inf.	Hes.	O.P.	QML	*I/MSE*	*H/MSE*	*OP/MSE*	*W/MSE*	*OP > H*
c_0	-1.99	-2.000	2.640	2.670	3.450	2.660	0.86	0.87	1.13	0.87	85.6%
c_1	-.048	-0.052	.030	.030	.056	.023	0.83	0.83	1.55	0.65	91.6%
c_2	2.76	2.834	2.073	2.087	3.041	1.987	0.81	0.81	1.19	0.78	86.8%
c_3	-4.89	-5.423	21.789	21.931	29.519	21.744	0.86	0.87	1.17	0.86	92.3%
c_4	-.099	-0.101	.003	.003	.004	.002	0.84	0.85	1.32	0.77	90.0%
c_5	-.318	-0.325	.013	.014	.020	.013	0.85	0.92	1.25	0.81	89.4%
σ^2	152E4	129E4	149E9	153E9	242E9	140E9	0.98	1.01	1.59	0.92	94.0%

TABLE 2

Model 2: mean estim. parameters, mean estim. variances, and M.S.E. ratios

par.	*True*	Est.	Inf.	Hes.	O.P.	QML	*I/MSE*	*H/MSE*	*OP/MSE*	*W/MSE*	*OP > H*
c_0	1.105	1.107	.064	.064	.082	.062	0.87	0.87	1.12	0.85	80.8%
c_1	1.89	1.935	1.356	1.362	1.794	1.309	0.91	0.92	1.20	0.88	79.4%
c_2	-.100	-.101	.002	.002	.003	.002	0.91	0.91	1.26	0.87	80.7%
c_3	-.365	-0.371	.012	.012	.016	.011	0.94	0.95	1.26	0.90	83.0%
σ^2	167E4	116E4	119E9	121E9	173E9	111E9	1.02	1.03	1.48	0.95	87.2%

TABLE 3

Model 3: mean estim. parameters, mean estim. variances, and M.S.E. ratios

par.	*True*	Est.	Inf.	Hes.	O.P.	QML	*I/MSE*	*H/MSE*	*OP/MSE*	*W/MSE*	*OP > H*
c_0	4.59	4.645	.655	.659	.939	.614	0.88	0.88	1.26	0.82	92.2%
c_1	-1.13	-1.300	.562	.582	1.000	.457	0.97	1.00	1.72	0.79	93.4%
c_2	1.76	1.710	.314	.315	.533	.275	0.89	0.89	1.51	0.78	85.3%
c_3	-3.92	-3.940	1.050	1.060	1.520	.979	0.84	0.85	1.21	0.78	87.8%
c_4	-.721	-0.718	.059	.060	.090	.054	0.84	0.85	1.28	0.76	91.7%
c_5	.511	0.499	.127	.128	.164	.126	0.88	0.88	1.13	0.87	85.9%
c_6	-.236	-0.201	.153	.154	.208	.149	0.85	0.86	1.16	0.83	90.0%
σ^2	.012	.012	.70E-9	.70E-5	1.0E-5	.67E-5	0.85	0.86	1.23	0.81	93.1%

TABLE 4

Model 4: mean estim. parameters, mean estim. variances, and M.S.E. ratios

par.	*True*	Est.	Inf.	Hes.	O.P.	QML	*I/MSE*	*H/MSE*	*OP/MSE*	*W/MSE*	*OP > H*
c_0	10.9	12.4	131.1	131.9	156.9	127.1	0.97	0.98	1.16	0.94	89.4%
c_1	-10.5	-11.0	5.91	5.97	7.68	5.52	1.02	1.03	1.33	0.96	90.4%
c_2	36.9	32.4	7398	7444	8862	7186	0.87	0.87	1.04	0.84	88.6%
c_3	-58.7	-60.8	765.9	771.5	947.5	739.1	0.96	0.97	1.19	0.92	84.0%
c_4	74.70	73.92	3514	3547	4368	3368	1.06	1.07	1.32	1.02	88.6%
c_5	-4.3	-4.7	705.7	709.7	870.8	680.3	0.98	0.99	1.21	0.94	81.4%
c_6	35.5	36.3	1156	1166	1423	1107	1.13	1.14	1.39	1.08	90.2%
c_7	-30.6	-32.1	1121	1129	1338	1093	1.01	1.01	1.20	0.98	90.0%
c_8	-310	-274	56378	56696	67453	54869	0.94	0.95	1.13	0.92	94.2%
c_9	92.6	92.6	256.7	258.2	311.7	253.0	0.89	0.89	1.08	0.88	82.2%
c_{10}	-.16	-.35	1.77	1.79	2.34	1.64	0.84	0.85	1.11	0.78	88.2%
σ^2	.14E7	.13E7	.61E11	.62E11	.82E11	.58E11	1.00	1.01	1.35	0.94	92.2%

TABLE 5

Model 5: mean estim. parameters, mean estim. variances, and M.S.E. ratios

par.	*True*	Est.	Inf.	Hes.	O.P.	QML	*I/MSE*	*H/MSE*	*OP/MSE*	*W/MSE*	*OP > H*
c_0	-14.4	-14.5	8.17	8.18	8.39	8.14	0.96	0.96	0.99	0.96	63.3%
c_1	-6.19	-6.17	1.14	1.14	1.18	1.13	1.02	1.02	1.06	1.01	62.2%
c_2	89.7	90.6	1407	1408	1434	1405	1.03	1.03	1.05	1.03	59.8%
c_3	116.6	117.0	495.2	495.9	510.5	492.0	0.96	0.96	0.99	0.95	62.0%
c_4	62.7	62.5	39.5	39.5	40.5	39.3	0.93	0.93	0.95	0.93	61.7%
c_5	-1.58	-1.60	.218	.218	.230	.214	0.97	0.97	1.02	0.95	66.0%
σ^2	.14E7	.14E7	.10E11	.10E11	.11E11	.10E-5	1.12	1.12	1.16	1.11	66.5%

TABLE 6

For the number of parameters indicated on the left, the outer product estimate of the variance is greater than the corresponding Hessian estimate at least in the percentage of cases indicated on the left

	Model 1 (7 parameters)		
$T = 42$	$T = 84$	$T = 210$	$T = 504$
1 param. ≥ 94%	2 param. ≥ 86%	2 param. ≥ 76%	1 param. ≥ 70%
4 param. ≥ 90%	5 param. ≥ 80%	6 param. ≥ 72%	4 param. ≥ 66%
7 param. ≥ 85%	7 param. ≥ 76%	7 param. ≥ 68%	7 param. ≥ 63%
	Model 2 (5 parameters)		
$T = 42$	$T = 84$	$T = 210$	$T = 504$
1 param. ≥ 87%	1 param. ≥ 82%	3 param. ≥ 71%	3 param. ≥ 64%
5 param. ≥ 79%	5 param. ≥ 71%	5 param. ≥ 65%	5 param. ≥ 59%
	Model 3 (8 parameters)		
$T = 150$	$T = 300$	$T = 500$	$T = 1000$
2 param. ≥ 93%	3 param. ≥ 91%	1 param. ≥ 88%	1 param. ≥ 76%
5 param. ≥ 90%	5 param. ≥ 88%	6 param. ≥ 82%	5 param. ≥ 68%
8 param. ≥ 85%	8 param. ≥ 83%	8 param. ≥ 78%	8 param. ≥ 61%

4.5. *Wald statistic: a summary figure*

Let us now combine all the parameters standard errors into a single random variable, like the Wald statistic. Another way to provide a synthetic view of the small sample behavior of the alternative covariance estimators is to combine all the parameters errors into a single random variable, like the Wald statistic. Let θ_0 be the *true* vector of parameters; under the null hypothesis $H_0 : \theta = \theta_0$ the Wald test statistic $(\hat{\theta} - \theta_0)'\Psi^{-1}(\hat{\theta} - \theta_0)$ is asymptotically distributed as a χ^2_k where k is the number of parameters. In each Monte Carlo replication $\hat{\theta} - \theta_0$ is the same and what change are only the different estimates of Ψ. Since is the inverse of the estimated Ψ that enters the Wald statistic, we should expect a value of the outer product based Wald systematically smaller than the corresponding value computed with the Hessian. Therefore, if we display the c.d.f. of these statistics, the curve related to the outer product matrix should be left-shifted with respect to the Hessian.

As far as the distribution of the QML Wald statistic (misspecification consistent) is concerned, we must recall how the covariance estimator $\hat{\Psi}$ is computed in this case

$$\hat{\Psi} = (H^{-1})(OP)(H^{-1}) \qquad (16)$$

Heuristically, if the Hessian estimated covariance matrix (H^{-1}) is smaller than the corresponding outer product estimate (OP^{-1}), the product of matrices resulting from (16) should be even smaller. Therefore its inverse should be larger, and the distribution of the Wald test rightmost shifted ([1]).

These effects are evidenced in figure 1. This figure is related to model 6 (11 coefficients), where the explanatory variables exhibit a very large fourth order moment. We note, of course, that the distance among the curves is larger for short sample periods and becomes negligible when the sample period becomes very long. In fact, consistency of the estimators for the asymptotic covariance matrices ensure that, increasing the sample period lenght, the cumulative distribution of the four Wald statistics collapse over the χ^2_k c.d.f.. It is clear that in this movement toward the χ^2_k curve, the four sampling distribution maintain their relative positions.

The curves related to the Hessian and expected Hessian (estimated information) are practically undistinguishable. Only in the shortest sample case the inequality between these two covariance estimators is observable: the Hessian curve is slightly left-shifted with respect to the other, and therefore slightly closer to the theoretical χ^2_k curve. It is clear, however, that the inequalities with respect to the outer product and the White estimators are much more relevant in practical applications.

4.6. *Powers of the Wald tests*

As far as the size of the test is concerned, little can be said about superiority of the covariance estimators, being only clear that the outer product Walds tend to be too conservative, while the others are prone to overreject the null hypotesis. We see from figure 1 that the χ^2 provides in many cases a bad approximation to the sampling distribution of the Wald statistics (of the QML misspecification consistent Wald, in particular). Therefore to evaluate the powers of the Wald tests based on different covariance estimators, we need to use critical values that are size-corrected by Monte Carlo. For the first experiment we use model 6 with a sample period of 200 observations.

The size corrected 5% and 1% critical values are the 0.95 and 0.99 quantiles of the empirical sampling distributions under the null hypothesis. The 11 coefficients of the model, under the null, have the same values already used to produce the results of figure 1. A sequence of alternatives is generated by multiplying all coefficients (none of which is zero) by a sequence of factors 1.005, 1.010, 1.015, etc.

The size-corrected powers are displayed in table 7. Results related to the estimated information are not displayed, as they are nearly identical to those

(1) Parks and Savin (1990) show in their experiments on a simple linear regression model that the inequality between Hessian and outer products usually holds on the diagonal terms of the covariance matrices, but not on the whole matrices. Therefore our interpretation on the inequality involving the QML matrix is a very rough one and would deserve further investigation.

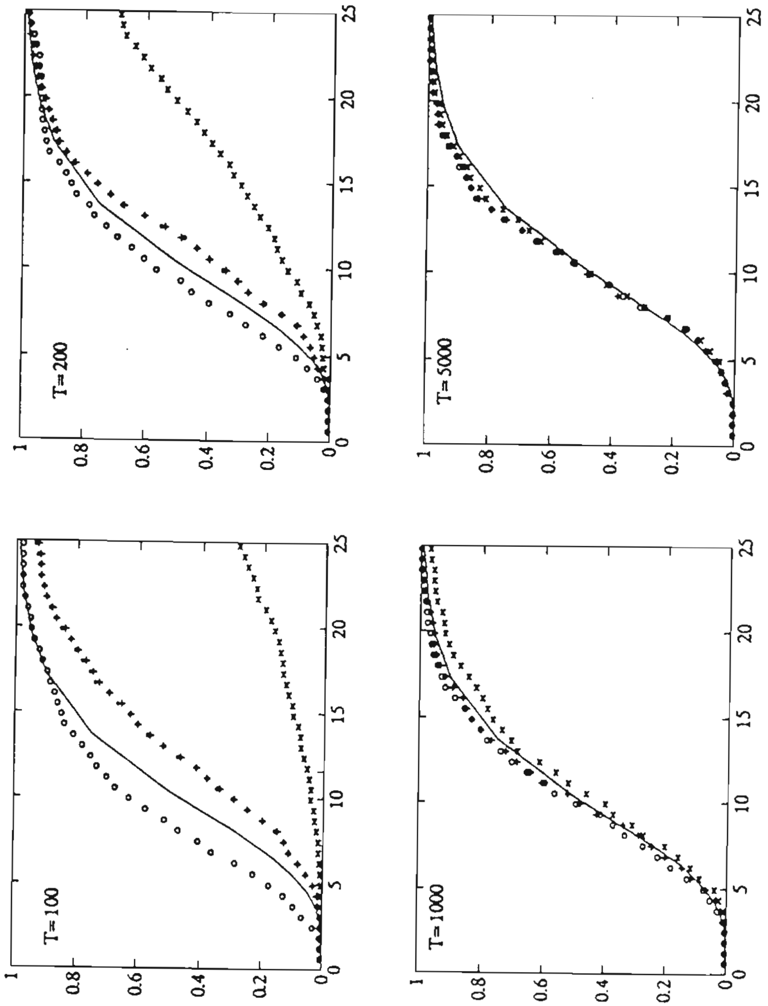

Figure 1 – Empirical sampling distributions of the Wald statistics χ^2_1; key: $-\chi^2_1$; ○ OP; ● INF.; + HES.; x Q.M.L.

obtained from the Hessian matrix. At 5% the powers of the Hessian and outer product are quite similar and outperform the QML Wald test. At 1% the Hessian outperforms the outer product Wald test. The power of the QML Wald test is very low in both cases, suggesting that the leptokurtic design of the explanatory variables is very critical also for the size-corrected power of this test, and not only for the discrepancy between the sampling and the asymptotic distribution.

In table 8 we display the size-corrected powers of the Wald test for model 7. The model has one coefficent and the only regressor is the constant term. For a sample period of 100 observations the sampling distribution of the Wald statistics are all rather close to the asymptotic χ^2_1 distribution. At 5% the power of all tests is nearly the same for local alternatives. When nonlocal alternatives are considered, the outer product version of the test exhibits more than the others the nonmonotonic behaviour discussed by Nelson and Savin (1988). We have not encountered this behavior in model 6, even trying larger departures from coefficients than those displayed in table 7.

At 1% the nonmonotonic behavior is more evident for all versions of the test. The QML Wald test outperforms only slightly the Hessian, while the latter outperforms the outer product version of the test.

It seems very hard to derive some general conclusion, however from these few cases it is possible to identify on average a mild superiority of the Hessian Wald test over the others.

TABLE 7

Model 6: size-corrected powers of the Wald test $H_0: \theta = \theta_0$

$T = 200$		at 5%			at 1%	
$H_1: \theta =$	Hessian	Out. Prod.	QML	Hessian	Out. Prod.	QML
$\theta_0 \times 1.005$	0.06	0.06	0.05	0.02	0.01	0.01
$\theta_0 \times 1.010$	0.13	0.13	0.05	0.03	0.02	0.01
$\theta_0 \times 1.015$	0.25	0.26	0.05	0.08	0.06	0.01
$\theta_0 \times 1.020$	0.46	0.45	0.06	0.20	0.16	0.01
$\theta_0 \times 1.025$	0.70	0.69	0.07	0.39	0.33	0.01
$\theta_0 \times 1.030$	0.89	0.87	0.09	0.65	0.55	0.01
$\theta_0 \times 1.040$	0.99	0.99	0.14	0.97	0.93	0.02
$\theta_0 \times 1.050$	1.00	1.00	0.25	0.99	0.99	0.03
$\theta_0 \times 1.060$	1.00	1.00	0.46	1.00	1.00	0.05
$\theta_0 \times 1.075$	1.00	1.00	0.86	1.00	1.00	0.14
$\theta_0 \times 1.100$	1.00	1.00	1.00	1.00	1.00	0.58

TABLE 8

Model 7: size-corrected powers of the Wald test $H_0 \sigma = 0$

T = 100	at 5%			ıt 1%		
$H_1 : c/\sigma =$	Hessian	Out. Prod.	QML	Hessian	t. Prod.	QML
0.01	0.05	0.05	0.05	.008	.008	.008
0.05	0.06	0.06	0.06	.003	.003	.003
0.10	0.08	0.07	0.07	.003	.003	.005
0.20	0.23	0.22	0.24	.033	.023	.036
0.30	0.52	0.50	0.52	0.15	0.11	0.16
0.50	0.95	0.95	0.95	0.70	0.61	0.72
0.80	1.00	1.00	1.00	0.99	0.93	0.99
1.00	1.00	0.98	1.00	0.99	0.79	0.99
1.20	1.00	0.88	1.00	0.90	0.54	0.95
1.30	0.99	0.77	0.99	0.76	0.40	0.90

5. CONCLUSION

The aim of the present paper was to investigate the sm sample behavior of the alternative covariance matrix estimators of the TobML estimates. We find that on average the Hessian and the estimated infmation matrix give almost identical results, differences being observed only fvery short samples.

Variances estimated from the outer product matrix aremost systematically larger (sometimes three or four times) than variances coiuted with the Hessian or with the estimated information. Therefore thilifference in small samples is not necessarily an indicator of misspecification, it would be argued by the information matrix test (White, 1982). The finitample MSE of the maximum likelihood estimates are usually intermediate bveen these group of estimators.

The systematic inequality clearly observed in the expenents has the same sign as for other type of models already analysed in the erature. Our paper confirms that, for a great variety of models used in ecometric applications, the choice of the covariance estimator is not neutral and hotheses testing may be strongly affected by such a choice.

Dipartimento Statistico
Università di Firenze
GIORGIO CALZOLARI

Department of Economics
European University Institute
ABRIELE FIORENTINI

ACKNOWLEDGMENTS

This paper extends some results that were preliminarily disced at the 6th World Congress of the Econometric Society (Barcelona, 1990) and at : econometrics semiars of Tilburg and Birmingham Universities. C. Dustmann, Gnbens, R.W. Parks, N.E. Savin and partecipants to conference and seminars are gratefully acknowledged for stimulating discussions. Responsability remains fully with the authors.

REFERENCES

T. AMEMIYA (1983), *A comparison of the Amemiya GLS and the Lee-Maddala-Trost G2SLS in a simultaneous equations Tobit model*, "Journal of Econometrics", 23, pp. 295-300.

T. AMEMIYA (1984), *Tobit models: a survey*, "Journal of Econometrics", 24, pp. 3-61.

T. AMEMIYA (1985), *Advanced Econometrics*, Basil Blackwell, Oxford.

A. ARABMAZAR, P. SCHMIDT (1981), *Further evidence on the robustness of the Tobit estimator to heteroscedasticity*, "Journal of Econometrics", 17, pp. 253-258.

A. ARABMAZAR, P. SCHMIDT (1982), *An investigation of the robustness of the Tobit estimators to non-normality*, "Econometrica", 50, pp. 1055-1069.

BANCA D'ITALIA (1988), *Modello mensile del mercato monetario*, Banca d'Italia, Servizio studi, Temi di discussione, n. 108, Roma.

E.R. BERNDT, B.H. HALL, R.E. HALL, J.A. HAUSMAN (1974), *Estimation and inference in non-linear structural models*, "Annals of Economic and Social Measurement", 3, pp. 653-665.

G. CALZOLARI, L. PANATTONI (1988a), *Alternative estimators of FIML covariance matrix: a Monte Carlo study*, "Econometrica", 56, pp. 701-714.

G. CALZOLARI, L. PANATTONI (1988b), *Finite sample performance of the robust Wald test in simultaneous equation systems*, "Advances in Econometrics", vol. 7, ed. by G.F. Rhodes Jr. and T.B. Fomby, JAI Press, Greenwich, pp. 163-191.

A. CHESHER, I. JEWITT (1987), *The bias of a heteroskedasticity consistent covariance matrix estimator*, "Econometrica", 55, pp. 1217-1222.

P.S. DHRYMES (1986), *Limited dependent variables*, "Handbook of Econometrics", ed. by Z. Griliches and M.D. Intrilligator, North-Holland Publishing Co., vol. 3, Amsterdam, pp. 1567-1631.

B. EFRON, D.V. HINKLEY (1978), *Assessing the accuracy of the maximum likelihood estimator: observed versus expected fisher information*, "Biometrika", 65, pp. 457-487.

R.C. FAIR (1977), *A note on computation of the Tobit estimates*, "Econometrica", 45, pp. 1723-1727.

R.P.H. FISHE, G.S. MADDALA, R.P. TROST (1979), *Estimation of a heteroscedastic Tobit model*, "Manuscript", University of Florida.

L. FLOOD, A. TASIRAN (1989), *A comparison of system Tobit estimations*, University of Gothenburg, Discussion Paper presented at the "European Meeting of the Econometric Society", Munich.

A.S. GOLDBERGER (1964), *Econometric Theory*, John Wiley & Sons, New York.

C. GOURIEROUX, A. MONFORT, A. TROGNON (1984), *Pseudo maximum likelihood methods: theory*, "Econometrica", 52, pp. 681-700.

W.H. GREEN (1985), *LIMDEP, user's manual*, Graduate School of Business Administration, New York University, New York.

W.E. GRIFFITHS, R.C. HILL, P.J. POPE (1987), *Small samples properties of probit model estimators*, "Journal of the American Statistical Association", 82, pp. 929-937.

B.H. HALL (1984), *Software for the computation of Tobit model estimates*, "Journal of Econometrics", 24, pp. 215-222.

J.J. HECKMAN (1976), *The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimators for such models*, "Annals of Economic and Social Measurement", 5, pp. 475-492.

J.G. MACKINNON, H. WHITE (1985), *Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties*, "Journal of Econometrics", 29, pp. 305-325.

G.S. MADDALA (1983), *Limited-dependent and qualitative variables in economics*, Cambridge University Press, Econometric Society Monographs n. 3.

T.A. MROZ (1987), *The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions*, "Econometrica", 55, pp. 765-799.

F.D. NELSON (1981), *A test for misspecification in the censored regression model*, "Econometrica", 49, pp. 1317-1329.

F.D. NELSON, L. OLSON (1978), *Specification and estimation of a simultaneous equation model with limited dependent variables*, "International Economic Review", 19, pp. 695-710.

F.D. NELSON, N.E. SAVIN (1988), *The nonmonotonicity of the power function of the wald test in nonlinear models*, The University of Iowa, Department of Economics, working paper.

W.K. NEWEY (1987), *Specification tests for distributional assumption in the Tobit model*, "Journal of Econometrics", 34, pp. 125-145.

R.J. OLSEN (1978), *Note on the uniqueness of the maximum likelihood estimator of the Tobit model*, "Econometrica", 46, pp. 1211-1215.

R.W. PARKS, N.E. SAVIN (1990), *The choice of coefficients covariance matrix estimator: outer product versus hessian*, The University of Iowa, Department of Economics, working paper.

J.L. POWELL (1983), *Asymptotic normality of the censored and truncated least absolute deviations estimators*, Stanford University, IMSSS, "Technical Report", n. 395.

I.R. PRUCHA (1984), *On the estimation of the variance covariance matrix of maximum likelihood estimators in nonlinear simultaneous equation systems: a Monte Carlo study*, University of Maryland, Department of Economics, Working paper nn.84-14.

P.M. ROBINSON (1982), *On the asymptotic properties of estimators of models containing limited dependent variables*, "Econometrica", 50, pp. 27-41.

P.A. RUUD (1986), *Consistent estimation of limited dependent variable models despite misspecification of distribution*, "Journal of Econometrics", 32, pp. 157-187.

R.C. SICKLES, P. SCHMIDT (1978), *Simultaneous equation models with truncated dependent variables: a simultaneous Tobit model*, "Journal of Economics and Business", 31, pp. 11-21.

R.J. SMITH (1987), *Testing the normality assumption in multivariate simultaneous limited dependent variable models*, "Journal of Econometrics", 34, pp. 105-123.

J. TOBIN (1958), *Estimation of relationships for limited dependent variables*, "Econometrica", 26, pp. 24-36.

H. WHITE (1980), *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*, "Econometrica", 48, pp. 817-838.

H. WHITE (1982), *Maximum likelihood estimation of misspecified models*, "Econometrica", 50, pp. 1-25.

H. WHITE (1983), *Corrigendum*, "Econometrica", 51, p. 513.

A.D. WITTE (1980), *Estimating the economic model of crime with individual data*, "Quarterly Journal of Economics", 94, pp. 57-84.

RIASSUNTO

Quando i coefficienti di un modello Tobit si stimano con la massima verosimiglianza, la loro matrice di covarianza è normalmente associata al metodo numerico impiegato per la massimizzazione della verosimiglianza. Gli stimatori usati nella pratica sono derivati da: (1) la matrice hessiana, (2) la matrice dei prodotti esterni delle derivate prime della funzione di verosimiglianza, (3) il valore atteso dell'hessiano, (4) una combinazione di 1 e 2 (la matrice di covarianza robusta di White). Differenze significative tra queste stime indicano di norma una possibile errata specificazione del modello. I risultati del nostro studio Monte Carlo sembrano contraddire quest'ultima affermazione, che rimane valida solo per campioni di notevoli dimensioni. Persino quando il modello è correttamente specificato si possono trovare delle differenze notevoli e sistematiche. La scelta di uno stimatore piuttosto che un altro non è neutrale e può quindi influire in maniera significativa sui risultati dei tests delle ipotesi.

SUMMARY

When the coefficients of a Tobit model are estimated by maximum likelihood their covariance matrix is typically, even if not necessarily, associated with the algorithm employed to maximize the likelihood. Covariance estimators used in practice are derived by: (1) the Hessian (observed information), (2) the matrix of outer products of the first derivatives of the log-likelihood (OPG version), (3) the expected Hessian (estimated information), (4) a mixture of 1 and 2 (White's QML covariance matrix). Significant differences among these estimates are usually interpreted as an indication of misspecification. From our Monte Carlo study this seems not to be true, unless the sample size is really very large. Even in absence of misspecification, large differences are encountered in small samples, and the sign of the differences is almost systematic. This suggests that the choice of the covariance estimator is not neutral and the results of hypotheses testing may be strongly affected by such a choice.