Economics Letters 21 (1986) 31–33
North-Holland

ALTERNATIVE SPECIFICATION TESTS FOR TOBIT AND RELATED MODELS

Peter KOOREMAN *

Tilburg University, 5000 LE Tilburg, The Netherlands

Received 13 January 1986

Recently Small and Hsiao (1985) proposed a specification test for the multinomial logit model as an alternative for the Hausman specification test. We show that Small and Hsiao's test procedure can also be applied to Tobit and related models.

1. Introduction

Under the assumption of no misspecification, the parameters of the Tobit model can be consistently estimated by either maximizing the Tobit likelihood function or by maximizing the truncated regression likelihood function. The latter estimator is inefficient relative to the former since it ignores the information contained in the zero-observations.

As has been set forth by Ruud (1984), the availability of this pair of estimators can be used to construct a Hausman test for misspecification. A practical problem with the Hausman test is that it requires the inversion of a matrix which is often nearly singular. As a consequence, the test result can be shaky due to computational inaccuracies.

In this note we show that the test procedure proposed by Small and Hsiao (1985) within the context of the multinomial logit model can serve as an alternative which avoids the computational problems of the Hausman test.

2. The Small–Hsiao specification test

Small and Hsiao (1985) are interested in testing the validity of the multinomial logit model as a specification in discrete choice analysis. Suppose we observe a discrete choice from a choice set C for each observation. Let D be a subset of the original choice set C. If the multinomial logit specification is correct (including its 'independence of irrelevant alternatives' property) the parameters can be consistently estimated either by maximizing the likelihood function over the complete sample, or by maximizing the conditional likelihood function based on probabilities conditional on choices within D.[1] The latter estimator is inefficient relative to the former since it ignores the information contained in the observations for which a choice from $\bar{D} = C - D$ is observed.

Small and Hsiao propose the following 'pseudo likelihood ratio test'. Divide the total sample randomly into two parts A and B of (asymptotically) equal sizes N^{A} and N^{B}. Let $\hat{\theta}_0^{\mathrm{A}}$ and $\hat{\theta}_0^{\mathrm{B}}$ be

* Helpful comments by Arie Kapteyn are gratefully acknowledged.

[1] Some parameters may not be identified from the restricted choice set.

estimates for the k vector of unknown parameters θ, obtained by maximizing the respective loglikelihood functions L^A and L^B, and let

$$\hat{\theta}_0^{AB} = \left(\tfrac{1}{2}\sqrt{2}\right)\hat{\theta}_0^{A} + \left(1 - \tfrac{1}{2}\sqrt{2}\right)\hat{\theta}_0^{B}. \tag{1}$$

Let $\hat{\theta}_1^B$ be the estimate obtained by maximizing the conditional loglikelihood function L_1^B over the N_1^B observations of subsample B whose choices belong to the restricted choice set D. Let

$$\Delta = -2\left[L_1^B(\hat{\theta}_0^{AB}) - L_1^B(\hat{\theta}_1^B)\right]. \tag{2}$$

The authors prove that Δ is asymptotically χ^2 distributed with k degrees of freedom.

A common feature of this test and the Hausman test is that they do not require specifying an explicit alternative model. Unlike the Hausman test statistic, the Small–Hsiao test statistic cannot assume negative values.

Obviously, there is a strong similarity between testing the multinomial logit model and testing the Tobit model. In both cases the consistent but inefficient estimator is based on a subsample where the endogenous variable only assumes values within a restricted range. In the proof of their proposition Small and Hsiao first observe that

$$\Delta \cong (\hat{\theta}_1^B - \hat{\theta}_0^{AB})' H_1^B(\hat{\theta}_1^B)(\hat{\theta}_1^B - \hat{\theta}_0^{AB}), \tag{3}$$

using a Taylor Series expansion of L_1^B around $\hat{\theta}_1^B$; H_1^B is the negative Hessian of L_1^B. Approximation (3) holds for any twice differentiable log likelihood function L_1^B. Next, it is sufficient to show that the asymptotic variance of $(N^B)^{1/2}.\ (\hat{\theta}_1^B - \hat{\theta}_0^{AB})$ equals the asymptotic limit of $(H_1^B/N^B)^{-1}$. This part of the proof only requires that the consistent but inefficient estimator is exclusively based on a part of the observations on which the efficient estimator is based. Hence, the Small–Hsiao test procedure cannot only be applied to the multinomial logit model but also to a much wider range of models, particularly Tobit and related models.

3. An empirical illustration

In Kooreman and Kapteyn (n.d.) the following endogenous switching household labor supply model was estimated:

$$l_f^* = g_f(w_m, w_f, p, \mu) + \epsilon_f, \tag{4}$$

$$l_f = l_f^*, \qquad l_m = g_m(w_m, w_f, p, \mu) + \epsilon_m \quad \text{if} \quad l_f^* < T, \tag{5),(6}$$

$$l_f = T, \qquad l_m = g_m(w_m, \bar{w}_f, p, \mu) + \epsilon_m^r \quad \text{if} \quad l_f^* \geq T. \tag{7),(8}$$

l_m and l_f denote male and female leisure, w_m and w_f are the male and female wage rate, respectively, p is the price of consumption and μ is non-labor family income; T is total time endowment. g_m and g_f are demand functions of the Almost Ideal Demand System and ϵ_f, ϵ_m and ϵ_m^r are error terms. The shadow wage $\bar{w}_f$ is implicitly defined by

$$T = g_f(w_m, \bar{w}_f, p, \mu). \tag{9}$$

Eq. (4) determines whether the female partner works in a paid job ($l_f^* < T$) or not ($l_f^* \geqslant T$) (all males are assumed to work). Eqs. (5) and (6) describe the number of hours worked by the partners in a two-earner household. Eq. (8) describes the male labor supply in a household with a non-participating female.

An interesting hypothesis to test is whether this model is capable of simultaneously describing these different aspects of the household's labor supply behavior. This seems to be an ideal application of the generalized Small–Hsiao specification test.

The model has been estimated twice. First only data on households where both partners worked in a paid job were employed (197 observations). Maximizing the appropriately conditioned likelihood function yielded a consistent (but inefficient) estimator. Next, an efficient estimator was obtained by maximizing the complete likelihood function over the 197 two-earner households plus 310 households where only the male partner worked in a paid job.

We first compared the two estimators using a Hausman test. The difference of the two relevant covariance matrices [cf. Ruud (1984, eq. (2.9))], however, did not satisfy the requirement of being positive semi-definite. The guaranteed positive semi-definite version was computed at 48.0 [Ruud (1984 eq. (2.10))]. The Small–Hsiao and the interchanged Small–Hsiao (obtained by reversing subsets A and B) test statistics are 46.5 and 53.8, respectively. So, in contrast to the empirical results reported by Small and Hsiao, the differences between the test statistics are quite small. All statistics reject equality of the two sets of estimates at any reasonable level of significance. [2,3]

References

Kooreman, P. and A. Kapteyn, n.d., Estimation of rationed and unrationed household labour supply functions using flexible functional forms, The Economic Journal, forthcoming.

Ruud, P., 1984, Tests of specification in econometrics, Econometric Reviews 3, 211–276.

Small, K.A. and C. Hsiao, 1985, Multinomial logit specification tests, International Economic Review 26, 619–627.

[2] Under the null hypothesis all statistics follow a $\chi^2_{(11)}$ distribution.

[3] In finite samples, the two test statistics could lead to conflicting inferences; see Small and Hsiao (1985) for a discussion.