



UNIVERSITÀ POLITECNICA DELLE MARCHE
Dipartimento di Scienze Economiche e Sociali

A SIMPLE AND EFFECTIVE MISSPECIFICATION
TEST FOR THE DOUBLE-HURDLE MODEL

Riccardo Lucchetti, Claudia Pigni

QUADERNI DI RICERCA n. 397*

January 2014

(*) La numerazione progressiva continua dalla serie denominata “Quaderni di ricerca — Dipartimento di economia”

Abstract

The commonly-used version of the double-hurdle model rests on a rather restrictive set of statistical assumptions, which are very seldom tested by practitioners, mainly because of the lack of a standard procedure for doing so, although violation of such assumptions can lead to serious modelling flaws. We propose here a bootstrap-corrected conditional moment portmanteau test which is simple to implement and has good size and power properties.

JEL Class.: C12, C15, C02

Keywords: Information Matrix Test; Bootstrap; Double-Hurdle model

Indirizzo: Riccardo Lucchetti: Università Politecnica delle Marche, DISES, r.lucchetti@univpm.it
Claudia Pigni: Università degli studi di Perugia, Dipartimento di Economia, Finanza e Statistica, c.pigni@stat.unipg.it

A simple and effective misspecification test for the double-hurdle model

Riccardo Lucchetti, Claudia Pignini

1 Introduction and motivation

The double-hurdle model is a commonly used model for dealing with double censoring. This model is well suited for analysing situations where a sample selection effect occurs and a corner zero solution is possible in the optimisation process by the individual. As such, it has been used in countless applications, such as labour market studies, in which the dependent variable is the number of hours worked (the classic reference here is Blundell, Ham, and Meghir (1987)), migrant remittances (Bettin, Lucchetti, and Zazaro, 2012) or demand analysis for certain types of goods such as tobacco or alcohol, from Jones (1989) onward.

In the literature, the earliest reference is Cragg (1971); later, the model was extended to what is called the “dependent” model, which can be described by the equations

$$y_i^* = \beta' X_i + \varepsilon_i \quad (1)$$

$$s_i^* = \gamma' Z_i + u_i \quad (2)$$

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \text{ and } s_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N \left[0, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right] \quad (4)$$

Where y_i is the dependent variable of interest, which is subject to a censoring rule similar to Heckman’s sample selection model via the latent variable s_i^* , plus a Tobit-like constraint.

Estimation is typically carried out by Maximum Likelihood (ML), which is nowadays routine, especially since analytical expressions for the score vector and the Hessian matrix are readily derived. Clearly, the distributional assumption in equation (4) is crucial. Unfortunately, this assumption is rarely tested in practice. In fact, there is no standard way to carry out such

a test; this is quite surprising, as it can be easily shown that distributional misspecification leads to inconsistency of the ML estimator.¹

Put differently, a QMLE strategy (White, 1982) is inapplicable here. Even more importantly, alternative approaches robust to non-normality have not emerged in the literature, like they have for other closely related models, such as Heckman’s sample selection model (see for example Das, Newey, and Vella, 2003).

2 Our Proposed Test

The test we propose builds on a conditional-moment approach originally proposed by Smith (1987), supplemented with a bootstrap correction to improve its poor finite-sample properties, as suggested by Horowitz (1994). A similar strategy was recently proposed by Lucchetti and Pignini (2013), who focused on testing the bivariate normality assumption in the bivariate probit and sample selection models.

This test uses the fact that, under correct specification, the Information Matrix equality implies that the score variance plus the expected Hessian should be zero. This result provides a set of moment conditions that can be empirically tested. The Information Matrix test is therefore a test for $E(C_i) = 0$, where

$$C_i = \text{vech} \left[\frac{\partial^2 \ell_i}{\partial \theta \partial \theta'} + G_i G_i' \right], \quad (5)$$

ℓ_i is the log-likelihood for the i -th observation ($i = 1, \dots, n$), θ is the k -vector of parameters and $G_i \equiv \frac{\partial \ell_i}{\partial \theta}$; all quantities are evaluated at the “true” vector $\theta = \theta_0$. In what follows, we will adopt the notational convention to indicate individual elements of the vector C_i by superscripting the two elements of the vector θ with respect to which the derivatives are taken: for example, $C_i^{\beta, \sigma}$ indicates $\frac{\partial^2 \ell_i}{\partial \beta \partial \sigma} + \frac{\partial \ell_i}{\partial \beta} \cdot \frac{\partial \ell_i}{\partial \sigma}$.

The asymptotic version of the test is calculated via an Outer Product of the Gradient (OPG) regression (see Davidson and MacKinnon (2001)): the test statistic equals nR^2 of the regression of a vector of ones on a matrix M , with typical row $M_i' = [G_i', C_i']$. Under the null, the test statistic has an asymptotic χ^2 distribution with degrees of freedom given by $df = \text{rank}(M) - k$.

¹Arabmazar and Schmidt (1982) prove the inconsistency of the Tobit estimator when disturbances are non-normal. The authors argue that the same considerations apply to more general models with limited dependent variables, such as sample selection models.

Since the asymptotic version of the test is known² to suffer from very severe size bias in finite samples, a standard bootstrap correction is adopted: given the ML estimate $\hat{\theta}$, the test statistic T is calculated. Then, a number B of artificial data sets is generated, using $\hat{\theta}$ as parameters and the test on the simulated data (T_b , for $b = 1 \dots B$) is computed. The bootstrapped critical values are then recovered from the empirical quantiles of T_b .³

2.1 Computational aspects

The number of available moment conditions may be, in general, smaller than $k(k+1)/2$; this is because some may be collinear with elements of the score, so that the column rank of the matrix M is not full. For the double-hurdle model, this occurs under conditions that closely parallel those identified in Lucchetti and Pignini (2013) for Heckman’s sample selection model.

The upper bound of df is $k(k+1)/2 - 1$, since at least $C_i^{\rho,\rho}$ is always dropped in the OPG regression regardless of the model specification: this happens in the most favourable case of a model without intercepts where the sets of regressors X_i and Z_i are non overlapping and $C_i^{\rho,\rho}$ can be written as a linear combination of $C_i^{\gamma,\sigma}$, $C_i^{\gamma,\rho}$, $C_i^{\sigma,\rho}$, and G_i^ρ . In other (more common) cases, the number of df varies with different model setups.

Whatever the overall number of available moment conditions, however, the number of moment conditions to use in practice is open to choice. Such choice must be based on a mix of different considerations: small sample performance, ease of computation and scope of the alternative hypothesis.

We therefore focus on two variants: first, a test which includes all moment conditions (henceforth “All moments”, or AM test), which can be considered an asymptotically optimal portmanteau specification test versus general misspecification, including non-normality, non-linearity, heteroskedasticity and other forms of incorrect specification.

Next, we consider a reduced version of the test using only third and fourth moments of the joint distribution (henceforth “Third and fourth moments”, or TF test). The rationale is that, strictly speaking, this is interpretable as a normality test; however, other forms of misspecification, such as heteroskedasticity, omitted variables or non-linearity, would probably induce non-normality as a side effect, so it may be advantageous to consider a test based on a smaller set of moment conditions whose finite sample performance may be better. A similar choice appears to be rather effective in Lucchetti and Pignini (2013). Hence, for the TF test, we select from C_i only those

²See for example Cribari-Neto (1997) and references therein.

³Other approaches are possible; see Davidson and MacKinnon (2007).

columns containing third and fourth moment conditions to be included in the OPG regression. In the double-hurdle model, third moment conditions are in $C_i^{\beta,\sigma}$, $C_i^{\beta,\rho}$, $C_i^{\gamma,\sigma}$, $C_i^{\gamma,\rho}$ and fourth moment conditions are in $C_i^{\sigma,\sigma}$, $C_i^{\sigma,\rho}$ and $C_i^{\rho,\rho}$.

3 Monte Carlo study

Before going into the details of our Monte Carlo experiment, a word of warning is necessary. Numerical optimisation of the double-hurdle log-likelihood may be difficult in some cases for two reasons: first, as is well known among practitioners, the log-likelihood may be bimodal, especially in smaller samples;⁴ moreover, the maximal value of ρ may, in finite samples, lie arbitrarily close to unity, where the log-likelihood is non-differentiable and the whole scheme breaks down. Therefore, we discard all cases in which $|\hat{\rho}| > \tanh(5) \simeq 0.99991$. This happened especially often in the power analysis (unsurprisingly) when $n = 256$, but was a relatively rare occurrence with larger samples. The problem disappeared altogether for $n > 1024$. For this reason, we omit reporting the results we obtained for $n = 256$. We do not consider this a crucial omission: samples as small as 256 observations are nowadays rather uncommon. Besides, failure to obtain convergence of the numerical ML procedure should give by itself a qualitative indication of possible specification problems.

3.1 Experiment setup

We study the finite-sample properties of the proposed test in both the AM and the TF versions. We design each scenario in order to analyse how its performance relates to sample size, correlation coefficient, degree of censoring in the dependent variable and departures from normality. For each scenario, we run 1000 Monte Carlo and 400 bootstrap replications.

For the size analysis, we run Monte Carlo experiments considering values of ρ in the grid $(-0.5, 0, 0.5)$ and sample sizes of 512, 1024, 2048, and 4096 observations.⁵ In each experiment, the explanatory variables for equation (1) are a constant term and x while the explanatory variables for equation (2) are a constant and z ; x and z are independent standard normal random

⁴In order to circumvent this problem, Bettin, Lucchetti, and Zazzaro (2012) propose a grid search procedure, which we do not use here to maximise speed.

⁵We also ran a few experiments with more extreme values of ρ ; these, however, are not reported because the number of non-convergent estimates was rather problematic; nevertheless, the results are qualitatively similar to those reported.

variables. This makes the number of available moment conditions 15 for the AM test and 9 for the TF test.

The conditional mean parameters are $\beta = \gamma = [c, 1]$. The constant term c is computed so that, for a given value of ρ , a desired percentage of non-censored observations is returned after generating the data.⁶ The standard deviation σ in (4) is set to 1.

For the power analysis, we study the consequences of distributional misspecification in the main and in the selection equation error terms separately. In both cases, we hold one error term normally distributed and generate the other as either a standardised student- t_6 or a standardised χ_2^2 . This choice is made so to compare the normal distribution to a markedly asymmetric alternative and a symmetric (but heavily leptokurtic) one and have an idea of the relative impact of deviation from normality in both directions. The joint distribution is then derived through a Gaussian copula. We use the same grid for the correlation coefficient, covariates and standard deviation as for the size analysis and we run our experiment for sample sizes of 512, 1024 and 2048. The procedure for getting the desired value of non-censored observations becomes overly complex when distributions are not normal. Therefore, we set c as above (in all experiments, the actual observed proportion of censored cases p was very close to the desired one).

3.2 Size

Table 1 displays the results of the size analysis. Both the AM and the TF versions of the test exhibit good size properties, even though the TF version seems to be slightly less reliable in large samples.

The first experiment outcome we mention is important, but unsurprising, as it merely confirms a well-known result, uniformly agreed upon by the whole literature: the asymptotic version of the test exhibits a severe size bias as it over-rejects the null hypothesis far too often: the empirical size of the AM test is above 80% for a nominal size of 5% with 4096 observations in all the scenarios considered; in the same cases, the size bias of the TF test is more contained but far from its nominal value (around 20%).⁷ Hence, bootstrapping critical values is absolutely mandatory.

As can be seen in Table 1, the empirical size of the test approaches the nominal one as the sample size grows, which is also largely expected. A much

⁶In practice, we solve numerically for c the equation $p = \Phi_2\left(\frac{c}{\sqrt{2}}, \frac{c}{\sqrt{2}}, \frac{\rho}{1+\sigma^2}\right)$, where p is the desired percentage of non-censored observations and $\Phi_2(\cdot)$ is the bivariate standard normal CDF.

⁷Results are not presented here in table form for brevity but are available upon request.

more interesting result is one that closely parallels the findings in Lucchetti and Pignini (2013) for the sample selection model: the performance of the test worsens as the degree of censoring increases. This suggests that the size of the test is partly affected by the information loss due to censoring: for instance, there is a slight tendency to under-rejection with 75% of censored observations and $\rho = 0$, which is the most unfavourable case in terms of the quantity of information the test has at its disposal.

3.3 Power against non-normality

Tables 2 and 3 display the results of the power analysis that considers distributional misspecifications in the main and in the selection equation, respectively.

From Table 2, it clearly emerges that the TF version of the test has better power properties than the AM version: while the former is built only on the normality assumption, the latter version tests for all assumptions of correct specification that, except for normality, are not violated in the scenarios considered⁸. On the other hand, as discussed above, non-normality may be in turn a mere side effect of other forms of misspecification.

The results in Table 3 show instead that both tests have rather poor power properties when departures from normality are present only in the selection equation: the binary nature of the dependent variable affects mainly the proportion of observed data which, however, depends on the number of censored observations.

In general, the power properties of the test depend on the intensity of the departures from normality: the number of rejections is generally higher when the asymmetric alternative is considered. Moreover, there is a clear effect of the degree of censoring on power, which increases as the amount of information lost due to censoring decreases.

4 Conclusions

Testing for misspecification in the double hurdle model is an important task, that can be carried out very effectively by a conditional moment test.

Since the routine estimation technique is maximum likelihood, the relevant moment conditions can be chosen from those stemming from the Information Matrix equality. However, a bootstrap correction is absolutely

⁸This fact does not appear in Lucchetti and Pignini (2013), where power properties in the sample selection are good in general. Again, in the case of the sample selection model, much more information is available to the test than in the case of the double hurdle model.

Table 1: Empirical size

		AM, Censored observations: 25%											
		$n = 512$			$n = 1024$			$n = 2048$			$n = 4096$		
$\alpha \setminus \rho$		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
10%		0.051	0.067	0.090	0.096	0.068	0.105	0.111	0.104	0.101	0.103	0.091	0.108
5%		0.021	0.038	0.042	0.042	0.030	0.049	0.061	0.053	0.050	0.056	0.048	0.056
1%		0.001	0.005	0.004	0.007	0.004	0.008	0.012	0.011	0.007	0.009	0.014	0.007
		TF, Censored observations: 25%											
		$n = 512$			$n = 1024$			$n = 2048$			$n = 4096$		
$\alpha \setminus \rho$		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
10%		0.07	0.094	0.094	0.082	0.069	0.089	0.091	0.088	0.086	0.110	0.118	0.111
5%		0.031	0.046	0.055	0.040	0.038	0.031	0.057	0.046	0.045	0.049	0.050	0.054
1%		0.005	0.007	0.009	0.008	0.005	0.007	0.011	0.004	0.011	0.018	0.015	0.011
		AM, Censored observations: 75%											
		$n = 512$			$n = 1024$			$n = 2048$			$n = 4096$		
$\alpha \setminus \rho$		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
10%		0.065	0.053	0.088	0.097	0.082	0.084	0.109	0.106	0.107	0.092	0.094	0.096
5%		0.024	0.020	0.043	0.049	0.037	0.036	0.053	0.041	0.052	0.042	0.039	0.047
1%		0.002	0.003	0.007	0.008	0.006	0.005	0.012	0.008	0.007	0.009	0.004	0.011
		TF, Censored observations: 75%											
		$n = 512$			$n = 1024$			$n = 2048$			$n = 4096$		
$\alpha \setminus \rho$		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
10%		0.101	0.087	0.106	0.076	0.092	0.094	0.104	0.124	0.111	0.096	0.117	0.112
5%		0.049	0.036	0.050	0.038	0.038	0.045	0.047	0.073	0.049	0.045	0.065	0.058
1%		0.007	0.008	0.012	0.008	0.007	0.004	0.011	0.013	0.011	0.005	0.017	0.016

The bold font indicates that the value of the empirical size lies within the 95% confidence interval for the corresponding nominal size α .

Table 2: Power at 95%: Misspecification of the Main Equation

Cens. \ ρ		AM								
		$n = 512$			$n = 1024$			$n = 2048$		
		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
χ_2^2	23%	0.980	0.986	0.984	0.999	0.998	0.970	1.000	1.000	0.997
χ_2^2	77%	0.018	0.015	0.023	0.081	0.039	0.063	0.191	0.095	0.110
t_6	24%	0.023	0.024	0.041	0.066	0.033	0.069	0.077	0.050	0.049
t_6	75%	0.022	0.022	0.034	0.048	0.033	0.056	0.058	0.052	0.082

Cens. \ ρ		TF								
		$n = 512$			$n = 1024$			$n = 2048$		
		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
χ_2^2	23%	1.000	0.998	0.996	1.000	1.000	1.000	1.000	1.000	1.000
χ_2^2	77%	0.073	0.024	0.029	0.508	0.377	0.363	0.998	0.993	0.993
t_6	24%	0.035	0.041	0.057	0.090	0.150	0.187	0.322	0.550	0.584
t_6	75%	0.031	0.039	0.060	0.031	0.065	0.137	0.061	0.226	0.478

Table 3: Power at 95%: Misspecification of the Selection Equation

Cens. \ ρ		AM								
		$n = 512$			$n = 1024$			$n = 2048$		
		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
χ_2^2	23%	0.050	0.074	0.099	0.018	0.021	0.018	0.001	0.000	0.005
χ_2^2	77%	0.019	0.011	0.018	0.010	0.001	0.008	0.007	0.001	0.007
t_6	24%	0.020	0.022	0.040	0.032	0.030	0.047	0.055	0.047	0.056
t_6	75%	0.037	0.031	0.040	0.043	0.025	0.028	0.046	0.036	0.038

Cens. \ ρ		TF								
		$n = 512$			$n = 1024$			$n = 2048$		
		-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5
χ_2^2	23%	0.074	0.071	0.089	0.070	0.070	0.085	0.122	0.084	0.095
χ_2^2	77%	0.045	0.040	0.035	0.057	0.043	0.052	0.102	0.092	0.087
t_6	24%	0.040	0.032	0.039	0.037	0.035	0.031	0.060	0.045	0.035
t_6	75%	0.038	0.046	0.048	0.049	0.039	0.038	0.054	0.071	0.050

indispensable. Moreover, the choice of the actual moment conditions to use in practice may be an issue.

In this article, we propose two versions of a misspecification test based on the above principles and show, through comprehensive Monte Carlo simulations, that our proposed tests have good size properties even for relatively small samples.

Their power when the normality assumption is violated is also, on the whole, satisfactory in finite samples and increasing in the sample size; however, it must be noted that it heavily depends on the degree of censoring and of departure from normality. Moreover, power appears to be much lower when the selection equation is affected.

References

- ARABMAZAR, A., AND P. SCHMIDT (1982): “An Investigation of the Robustness of the Tobit Estimator to Non-Normality,” *Econometrica*, 50(4), 1055–1063.
- BETTIN, G., R. LUCCHETTI, AND A. ZAZZARO (2012): “Endogeneity and sample selection in a model for remittances,” *Journal of Development Economics*, 99(2), 370–384.
- BLUNDELL, R., J. HAM, AND C. MEGHIR (1987): “Unemployment and Female Labour Supply,” *Economic Journal*, 97(388a), 44–64.
- CRAGG, J. G. (1971): “Some statistical models for limited dependent variables with application to the demand for durable goods,” *Econometrica*, 39, 829–844.
- CRIBARI-NETO, F. (1997): “On the corrections to information matrix tests,” *Econometric Reviews*, 16(1), 39–53.
- DAS, M., W. K. NEWEY, AND F. VELLA (2003): “Nonparametric Estimation of Sample Selection Models,” *Review of Economic Studies*, 70(1), 33–58.
- DAVIDSON, R., AND J. G. MACKINNON (2001): “Artificial Regressions,” in *A Companion to Theoretical Econometrics*, ed. by B. Baltagi, pp. 16–37. Blackwell.
- (2007): “Improving the reliability of bootstrap tests with the fast double bootstrap,” *Computational Statistics & Data Analysis*, 51(7), 3259–3281.

- HOROWITZ, J. L. (1994): “Bootstrap-Based Critical Values for the Information Matrix Test,” *Journal of Econometrics*, 61, 365–411.
- JONES, A. M. (1989): “A Double-Hurdle Model of Cigarette Consumption,” *Journal of Applied Econometrics*, 4(1), 23–39.
- LUCCHETTI, R., AND C. PIGINI (2013): “A test for bivariate normality with applications in microeconomic models,” *Statistical Methods and Applications*.
- SMITH, R. J. (1987): “Testing the Normality Assumption in Multivariate Simultaneous Limited Dependent Variable Models,” *Journal of Econometrics*, 34(1-2), 105–123.
- WHITE, H. (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50(1), 1–25.