# Testing non-nested models for non-negative data with many zeros*

J.M.C. Santos Silva†, Silvana Tenreyro‡, Frank Windmeijer§

16 February 2014

**Abstract**

In economic applications it is often the case that the variate of interest is non-negative and its distribution has a mass-point at zero. Many regression strategies have been proposed to deal with data of this type but, although there has been a long debate in the literature on the appropriateness of different models, formal statistical tests to choose between the competing specifications are not often used in practice. We use the non-nested hypothesis testing framework of Davidson and MacKinnon (1981, "Several tests for model specification in the presence of alternative hypotheses," *Econometrica*, 49, 781-793) to develop a novel and simple regression-based specification test that can be used to discriminate between these models.

*JEL codes*: C12, C52

*Key words*: Health economics, international trade, non-nested hypotheses, $C$ test, $P$ test.

*We are grateful to the editor Jason Abrevaya and to an anonymous referee for many helpful and constructive suggestions. We also thank Holger Breinlich, Francesco Caselli, Daniel Dias, Esmeralda Ramalho, Joaquim Ramalho, and Rainer Winkelmann for helpful comments, and John Mullahy for providing one of the datasets used in Section 4. Santos Silva acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010). Tenreyro acknowledges financial support from ERC and Bank of Spain, the latter through the Bank of Spain Associate Professorship. Windmeijer acknowledges financial support from ERC grant 269874 - DEVHEALTH.

†University of Essex and CEMAPRE, jmcss@essex.ac.uk.

‡London School of Economics, CREI, CEP and CEPR. S.Tenreyro@lse.ac.uk.

§University of Bristol, f.windmeijer@bristol.ac.uk.

## 1. INTRODUCTION

In many applications the variate of interest, say $y$, is non-negative and has a mixed distribution characterized by the coexistence of a long right-tail and a mass-point at zero. Applications using this sort of data are typical in health and international economics, but data with these characteristics are also found in many other areas. An example in international economics is the case of bilateral trade flows where the zeros may result from the existence of fixed costs or access costs that preclude firms or countries to sell into some destinations (see, for example, Melitz, 2003, Helpman, Melitz, and Yeaple, 2004, Chaney, 2008, and Arkolakis, 2008). An example in health economics is medical expenditures, which are zero for those individuals that do not utilise health care, and positive for those who do.[1]

What is common to the cases we are considering here is that the zeros are not the result of some observability problem but rather correspond to the existence of the so-called "corner-solutions". In this case researchers and policymakers are ultimately interested in the effects of the covariates on the distribution of the fully observable dependent variable $y$ (see Wooldridge, 2002, and Dow and Norton, 2003).

One question that is central in specifying models for this type of data is whether the zero and positive observations are generated by the same mechanism or whether the zeros are somehow different. When it is assumed that a single mechanism is at work the data are typically described by a single-index model such as the Tobit (Tobin, 1958, Eaton and Tamura, 1994) or by models with an exponential conditional expectation function (Mullahy, 1998, and Santos Silva and Tenreyro, 2006). If the zeros are believed to be generated by a different process, the covariates are allowed to affect the conditional distribution of $y$ in two different ways, leading to double-index models such as the two-part models of Duan, Manning, Morris, and Newhouse (1983) and Mullahy (1998), models based on Heckman's (1979) sample selection estimator, or zero inflated models such as the p-Tobit of Deaton and Irish (1984).

[1] See Jones (2000) for a survey of applications in health economics. La Porta, López-de-Silanes, and Zamarripa (2003) is an example of the use of this type of data in finance.

In this paper we propose a statistical test to discriminate between competing models for corner-solutions data. The test is specifically designed to discriminate between single- and double-index models and therefore it can be used to test whether or not the zero and positive observations are generated by different mechanisms. Developing tests for this purpose is not trivial because the competing models can imply very different estimation methods and it is not immediately clear what test can be used to choose between them. Our approach is based on the observation that, although the models being considered are based on very different modelling approaches and differ widely in the nature of the assumptions they make, they all define the conditional expectation of $y$ given a set of covariates $x$. Therefore, the suitability of each of the competing models can be gauged by testing the corresponding conditional expectation against that of any of the alternatives being considered. Heuristically, our test will check whether the estimate of the conditional expectation of $y$ obtained under the alternative can be used to improve the prediction of $y$ obtained under the null. If that is the case, we have evidence against the null because this implies that the model under the null is not explaining some features of the data that are captured by the alternative.

Of course, there are general specification tests that can be used to check the adequacy of the models we consider here (e.g., Bierens, 1982 and 1990, and Wooldridge, 1992). However, these tests do not use information about the precise alternative that is being considered and, more importantly, despite being available for decades these tests were never really adopted by practitioners. Therefore, our purpose is to introduce a test that has good performance and is simple enough to be appealing to practitioners; after all, a test needs to be performed to have non-zero power.

Having an appropriate test to choose between competing models is important for several reasons. First, because none of the proposed specifications nests or generally dominates its competitors, deciding which of the models is more appropriate is an empirical question that has to be answered for each specific dataset the researcher is considering. Second, and related, the test may help to empirically discriminate among competing theories and thus shed more light on the mechanisms affecting the variable of interest. Thus, for example, the structural gravity model for trade of Anderson and Yotov (2010), which in turn builds on Anderson and van Wincoop (2003), leads to a single-index specification with minimal distributional assumptions at the estimation stage; instead, the model of Helpman, Melitz and Rubinstein (2008) leads to a double-index specification and relies on strong distributional assumptions. Even if the researcher favours one specification on theoretical grounds, it is important to check its adequacy by testing it against competing specifications because this can help to confirm (or reject) the researcher's views on the models. Finally, the model choice plays a critical role in the estimation of marginal effects and elasticities that are often used to assess the impact of different public policies and, as said, corner-solutions data are of high prevalence in key areas of public policy such as health and international economics.

The remainder of the paper is organised as follows. Section 2 develops the proposed specification test and compares and contrasts it with alternative testing procedures. Section 3 presents the results of a simulation study illustrating the finite sample performance of the proposed test, and Section 4 employs two well-known datasets to illustrate the practical use of the approach we suggest. Finally, Section 5 contains brief concluding remarks and an Appendix gives technical details on the proposed testing procedure and presents a complementary result on the relation between two-part and sample selection models.

## 2. THE PROPOSED TEST

### 2.1. The testing strategy

A feature that is common to all the models regularly used to describe corner-solutions data is that, implicitly or explicitly, they specify $\mathrm{E}[y|x]$, the conditional expectation of $y$ given $x$. Moreover, $\mathrm{E}[y|x]$ is often the object of interest in empirical applications because it is the function needed to compute key quantities of interest, such as marginal effects and elasticities, which in turn can shed light on welfare effects (e.g., Arkolakis, Costinot and Rodríguez-Clare, 2009). Therefore, we compare the different models on the basis of the adequacy of the implied conditional expectations. In particular, we suggest testing the specification of $\mathrm{E}[y|x]$ implied by one model against alternatives in the direction of competing specifications. This can be done by framing the problem as a test of non-nested hypotheses.

The motivation for using tests of non-nested hypotheses is obvious when the purpose is to compare models whose implied conditional expectations are non-nested, in the sense that they cannot be obtained by imposing restrictions on the parameters of the competing specifications. For instance, the Tobit does not nest, and is not nested by, the exponential conditional expectation model. But, perhaps less obviously, we argue that the use of the non-nested hypotheses framework is needed even when the functional form of the conditional expectation of one model is identical to, or nested within, that of the competing alternative. This is because the models imply not only a functional form for $\mathrm{E}[y|x]$ but they also prescribe a method to estimate the parameters of interest. Therefore, even if two models specify the same functional form for $\mathrm{E}[y|x]$, the implied conditional expectations will generally be different because they are evaluated at different parameter values, even asymptotically. In this case, none of the models leads to an estimated conditional expectation that nests the others in the sense that it will always fit the data at least as well as that of its competitors. For example, Heckman's selection model nests the two-part model of Duan et al. (1983), but when these models are used to describe corner-solutions data there is no guaranty that the conditional expectation implied by the sample selection model will fit the data better than the conditional expectation implied by the two-part model.[2]

[2] Therefore, using the usual t-statistic on the inverse Mills ratio coefficient to check if the two-part model is a valid simplification of the sample selection model gives no information about the ability of these models to describe corner-solutions data (cf. Dow and Norton, 2003 and Norton, Dow and Do, 2008). In an often cited paper, Duan, Manning, Morris and Newhouse (1984) argue that there can be correlation between the error terms in the two parts of the two-part model and that therefore

**Table 1: Some models for corner solutions data**

| Model | Specification | $\mathrm{E}[y \mid x]$ |
|---|---|---|
| ECE | $\mathrm{E}[y \mid x] = \exp(x'\beta)$ | $\exp(x'\beta)$ |
| 2PM | $\Pr(y > 0 \mid x) = \Phi(x'\gamma)$<br>for $y > 0 : \ln(y) = x'\beta + e$<br>$e \mid x \sim \mathcal{N}(0, \sigma^2)$ | $\exp\left(x'\beta + \frac{\sigma^2}{2}\right)\Phi(x'\gamma)$ |
| M-2PM | $\Pr(y > 0 \mid x) = \Phi(x'\gamma)$<br>$\mathrm{E}[y \mid x, y > 0] = \exp(x'\beta)$ | $\exp(x'\beta)\Phi(x'\gamma)$ |
| Sample Selection (in logs) | $\Pr(y > 0 \mid x) = \Pr(x'\gamma + e_1 > 0 \mid x)$<br>for $y > 0 : \ln(y) = x'\beta + e_2$<br>$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \mid x \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}\right)$ | $\exp\left(x'\beta + \frac{\sigma^2}{2}\right)\Phi(x'\gamma + \rho\sigma)$ |

Notes: ECE is Exponential Conditional Expectation; 2PM is Two-Part Model; M-2PM is Modified Two-Part Model.

Most tests of non-nested hypotheses require the specification of the entire conditional density of $y$ given $x$ (Cox, 1961, Atkinson, 1970, Quandt, 1974, Pesaran and Deaton, 1978, Vuong, 1989, Gourieroux and Monfort, 1994, and Santos Silva, 2001), and therefore are not appropriate in this context. An exception is Davidson and MacKinnon (1981), who introduced the $P$ and $C$ tests of non-nested hypotheses that only require the specification of the conditional mean. In the next subsection we build on these results to develop a testing procedure specifically designed to discriminate between competing models for corner-solutions data.

In the development of the test below, we specifically consider single- and double-index models of the type illustrated in Table 1. The models in Table 1 are the exponential conditional expectation (ECE); the two-part model (2PM) of Duan et al. (1983);[3] Mullahy's (1998) modified two-part model (M-2PM), and the sample

the two models are not nested. We show in the Appendix that their example is misleading and does not lead to the conclusion that the errors can be correlated.

[3]Our results also apply to the case where normality of $e$ is not assumed and the conditional expectation of the 2PM is estimated by smearing, as suggested by Duan (1983). In this case, $\mathrm{E}[y \mid x] = \exp(x'\beta + s)\Phi(x'\gamma)$, where $s$ is the log of the scaling factor estimated by smearing.

selection model in logs.[4] The conditional means of these models can all be represented as $\mathrm{E}[y|x] = g(x'\beta) F(x'\gamma)$, where $F(x'\gamma) = \Phi(x'\gamma)$ for the double-index models (with implicit intercept shifts where applicable) and $F(x'\gamma) \equiv 1$ for the single-index ECE model, and $g(x'\beta) = \exp(x'\beta)$ in both cases. Our results below apply not only to these models but also to any models with $\mathrm{E}[y|x] = g(x'\beta) F(x'\gamma)$, with $g(\cdot) > 0$ and $0 < F(\cdot) \leq 1$.

### 2.2. The test statistic

We build on Davidson and MacKinnon's (1981) seminal work to develop a testing procedure specifically designed to discriminate between competing models for corner-solutions data. Let $\{y_i, x_i'\}_{i=1}^n$ be an i.i.d. sample of size $n$ and assume that standard regularity conditions are satisfied (see Appendix A1). Furthermore, as in Davidson and MacKinnon (1981), suppose that we want to test Model $A$, implying

$$M_A : \mathrm{E}[y_i|x_i] = g_A(x_i'\beta_A) F_A(x_i'\gamma_A),$$

against Model $B$,

$$M_B : \mathrm{E}[y_i|x_i] = g_B(x_i'\beta_B) F_B(x_i'\gamma_B),$$

where, for $j \in \{A, B\}$, $g_j(\cdot) > 0$, $0 < F_j(\cdot) \leq 1$, $\beta_j$ and $\gamma_j$ are vectors of parameters, and the models have a single index when $F_j(\cdot) \equiv 1$.

As noted before, Models $A$ and $B$ will be non-nested even if $g_A(\cdot) F_A(\cdot)$ and $g_B(\cdot) F_B(\cdot)$ have the same form because in general these functions are evaluated at sets of parameter estimates with different probability limits. Therefore, as in Davidson and MacKinnon (1981), we start by nesting the competing specifications in a compound model of the form

$$M_C : \mathrm{E}[y_i|x_i] = (1-\alpha) g_A(x_i'\beta_A) F_A(x_i'\gamma_A) + \alpha g_B(x_i'\beta_B) F_B(x_i'\gamma_B),$$

and want to check the correct specification of $M_A$ by testing $H_0 : \alpha = 0$ against $H_1 : \alpha = 1$.[5] As noted by Davidson and MacKinnon (1981), in general $\alpha$, $\beta_A$, $\gamma_A$, $\beta_B$,

[4] The conditional expectation for this model was obtained by van de Ven and van Praag (1981).

[5] A two-tailed test could also be used. However, here we follow Fisher and McAleer (1979), who argue that, when the purpose is to discriminate between two competing models, one-sided (in

and $\gamma_B$ are not separately identified and therefore the test has to be performed by conditioning on parameter estimates. In particular, Davidson and MacKinnon (1981) consider two cases: the $P$ test, which conditions on the estimates obtained under the alternative, and the $C$ test that conditions on estimates obtained under both the null and alternative.[6] For reasons that will be clear in Subsection 2.3 below, we propose conditioning on estimates obtained under both the null and alternative, as in the $C$ test; however, unlike in the $C$ test, we will not condition on all parameter estimates under the null, but only on the estimates of $\gamma_A$. That is, $\beta_A$ is allowed to be freely estimated, and the hypothesis of interest is tested in the artificial regression

$$y_i = (1-\alpha)\, g_A\left(x_i'\beta_A\right) F_A\left(x_i'\hat{\gamma}_A\right) + \alpha g_B\left(x_i'\hat{\beta}_B\right) F_B\left(x_i'\hat{\gamma}_B\right) + \xi_i, \tag{1}$$

where $\hat{\beta}_j$ and $\hat{\gamma}_j$ denote estimates obtained under model $j \in \{A, B\}$.

Like in the $P$ test, estimation of (1) can be avoided by linearizing the model around $\beta_A = \hat{\beta}_A$. Moreover, in the specific context we have in mind, the variance of $\xi_i$ is likely to increase with $\mathrm{E}\left[y_i|x_i\right]$ and therefore we suggest estimating the linearization of (1) by weighted least squares, under the assumption that $\mathrm{Var}\left[y_i|x_i\right] \propto g_A\left(x_i'\beta_A\right) F_A\left(x_i'\gamma_A\right)$. As discussed in the next subsection, this modification is not only likely to improve the performance of the test in finite samples but, more importantly, it also leads to an interesting interpretation of the test.

Implementing the test in practice is very simple. Defining $\hat{y}_i^j = g_j\left(x_i'\hat{\beta}_j\right) F_j\left(x_i'\hat{\gamma}_j\right)$ and $\delta = (1-\alpha)\left(\hat{\beta}_A - \beta_A\right)$, the proposed test is just a (robust) t-test for $H_0 : \alpha = 0$ against $H_1 : \alpha = 1$ in the OLS estimation of an artificial regression of the form

$$\frac{y_i - \hat{y}_i^A}{\sqrt{\hat{y}_i^A}} = \frac{\nabla g_A\left(x_i'\hat{\beta}_A\right) F_A\left(x_i'\hat{\gamma}_A\right) x_i'}{\sqrt{\hat{y}_i^A}}\delta + \alpha\frac{\hat{y}_i^B - \hat{y}_i^A}{\sqrt{\hat{y}_i^A}} + \nu_i, \tag{2}$$

where $\nabla g_A\left(x_i'\hat{\beta}_A\right)$ denotes the derivative of $g_A(\cdot)$ evaluated at $x_i'\hat{\beta}_A$. This regression can be conveniently performed, for example in Stata (StataCorp.,

the direction of the alternative) tests should be used. This is in line with the seminal procedure developed by Cox (1961).

[6] The $J$ test described in Davidson and MacKinnon (1981) could also be used in this context. However, this procedure is not as attractive as the $P$ and $C$ tests because its implementation is cumbersome when the null is a nonlinear model.

2013), as a least squares regression of $\left(y_i - \hat{y}_i^A\right) / \nabla g_A\left(x_i'\hat{\beta}_A\right) F_A\left(x_i'\hat{\gamma}_A\right)$ on $x$ and $\left(\hat{y}_i^B - \hat{y}_i^A\right) / \nabla g_A\left(x_i'\hat{\beta}_A\right) F_A\left(x_i'\hat{\gamma}_A\right)$, using $\left(\nabla g_A\left(x_i'\hat{\beta}_A\right) F_A\left(x_i'\hat{\gamma}_A\right)\right)^2 / \hat{y}_i^A$ as weights.[7] The test based on (2) will be referred to as the $HPC$ test because it has features of the $P$ and $C$ tests and accounts for the presence of heteroskedasticity.

As in the $C$ test, the asymptotic variance of the t-statistic for the significance of $\alpha$ in (2) is not equal to 1 because the test does not take into consideration that $\gamma_A$ is evaluated at estimates under the null. Indeed, as sketched in Appendix A.1 (see also Davidson and MacKinnon, 1981, Pierce, 1982, and Lee, 2010), when $\gamma_A$ is evaluated at its maximum likelihood estimates, the t-statistic for $H_0 : \alpha = 0$ in (2) will be asymptotically normal with variance smaller than 1 and therefore the test based on it will be undersized. Still, we expect this problem to be much less severe than in the $C$ test because the $HPC$ test does not condition on $\hat{\beta}_A$. Naturally, as detailed in the Appendix, an asymptotically correct estimator of the variance of the estimate of $\alpha$ can be obtained using the misspecification-robust version of the methods presented in Davidson and MacKinnon (1981) and the t-statistic for $H_0 : \alpha = 0$ based on this estimator will be asymptotically distributed as a standard normal. Alternatively, an asymptotically valid covariance matrix estimator can be obtained using a simple pairs-bootstrap approach (Freedman, 1981).[8] However, in the next section we present Monte Carlo evidence suggesting that the $HPC$ test suffers only from small size distortions even when it is based on the uncorrected estimate

[7] Note that for the case where $g\left(x'\beta\right) = \exp\left(x'\beta\right)$ the artificial regression simplifies to

$$\frac{y_i - \hat{y}_i^A}{\sqrt{\hat{y}_i^A}} = x_i'\sqrt{\hat{y}_i^A}\delta + \alpha\frac{\hat{y}_i^B - \hat{y}_i^A}{\sqrt{\hat{y}_i^A}} + \nu_i,$$

which can be performed as a least squares regression of $\left(y_i - \hat{y}_i^A\right) / \hat{y}_i^A$ on $x$ and $\left(\hat{y}_i^B - \hat{y}_i^A\right) / \hat{y}_i^A$, using $\hat{y}_i^A$ as weights. In this case the test can also be performed with the `hpc` command we have written for Stata (StataCorp., 2013).

[8] Alternatively, it is possible to compute bootstrap p-values using a pairs-bootstrap procedure and a modified test for which the null is valid in the bootstrap samples (see, e.g., Davidson and MacKinnon, 2006, p. 822). However, bootstrapping the t-statistic is likely to require more bootstrap draws than the computation of the covariance matrix and, because the uncorrected test statistic is not asymptotically pivotal, it does not lead to the usual asymptotic refinements.

of the variance. Therefore, we conjecture that, for most empirical applications, the computational effort of correcting the covariance estimator may not be necessary.

The important case where the null is a single-index model is noteworthy. In this case $F_A(x_i'\gamma_A) \equiv 1$ and therefore $\hat{\gamma}_A$ vanishes from (1). Hence, when the null is a single-index model, the variance of the estimate of $\alpha$ does not need to be corrected. In this case the test based on (2) is just a weighted version of the $P$ test and consequently the (robust) t-test will asymptotically have the correct size.[9] This means that the $HPC$ test is particularly easy to perform when the null is a single-index model, which is often of interest in applied work.

Finally, it is important to mention that, as is standard with tests of non-nested hypotheses, the roles of the null and alternative can be reversed. This leads to three possible outcomes of the $HPC$ test: one model may be rejected and the other accepted, both models may be accepted, or both rejected. Therefore, unlike model selection criteria that always choose one of the models being compared, the $HPC$ test has the ability to reject both specifications when neither is appropriate. Conversely, if the two specifications are very close and the sample is not rich enough, the test may be unable to discriminate between the two competitors.

### 2.3. Comparison with other tests

The $HPC$ test is obviously closely related to the $P$ and $C$ tests of Davidson and MacKinnon (1981). In our context, these tests essentially check a moment condition of the form

$$E\left\{\left[y_i - g_A(x_i'\beta_A)\, F_A(x_i'\gamma_A)\right]\left[g_B(x_i'\beta_B)\, F_B(x_i'\gamma_B) - g_A(x_i'\beta_A)\, F_A(x_i'\gamma_A)\right]\right\} = 0.$$

That is, the tests check the correct specification of $M_A$ by testing whether the errors $y - g_A(x_i'\beta_A)\, F_A(x_i'\gamma_A)$ have zero expectation, giving more weight to the observations

[9] Likewise, when $g_A(\cdot) \neq g_B(\cdot)$ and $F_A(\cdot) \neq F_B(\cdot)$, the test can be performed without conditioning on the estimates of $\gamma_A$. In this case the test will again be just an heteroskedasticity adjusted version of the $P$ test, and therefore the (robust) t-test will asymptotically have the correct size.

for which the difference between the conditional means implied by $M_A$ and $M_B$ is larger.

In contrast, the test based on (2) checks a moment condition of the form

$$\mathrm{E}\left\{\left[y_i - g_A\left(x_i'\beta_A\right)F_A\left(x_i'\gamma_A\right)\right]\frac{g_B\left(x_i'\beta_B\right)F_B\left(x_i'\gamma_B\right) - g_A\left(x_i'\beta_A\right)F_A\left(x_i'\gamma_A\right)}{g_A\left(x_i'\beta_A\right)F_A\left(x_i'\gamma_A\right)}\right\} = 0.$$

That is, like the $P$ and $C$ tests, the $HPC$ test checks whether the errors of the model under the null have zero expectation when the weight given to each observation depends on the difference between the conditional expectations of the two models. The difference here is that, because of the weights accounting for the presence of heteroskedasticity in the $HPC$ auxiliary regression, a percentage difference between the two conditional means is used as a weight in the moment condition. The use of a percentage difference is appropriate and appealing in this particular context because all models being considered imply specifications of the conditional mean of $y_i$ which are strictly positive, but can be close to zero for a large proportion of the observations in the sample. These observations with fitted values close to zero, which are critical in distinguishing between single- and double-index models, are all but ignored when the weights are just the difference between the two sets of fitted values, as in the $P$ and $C$ tests, and not a percentage difference, as in the proposed test.

Therefore, we expect the $HPC$ test to outperform the $P$ and $C$ tests because it partially accounts for the presence of heteroskedasticity and consequently it is better suited to the particular context we have in mind.[10] Moreover, as will be illustrated, the $P$ test has very low power when the null is a single index model that can be obtained as a limiting case of the alternative when $F_B\left(x_i'\gamma_B\right)$ approaches 1.[11] This

[10] Notice that the $P$ and $C$ tests can be made robust to heteroskedasticity by using an appropriate covariance matrix, as we do in the simulations in Section 3. What makes the $HPC$ test different, however, is that the moment condition it checks is weighted to account for the likely presence of heteroskedasticity.

[11] This happens because the $P$ test checks whether there is some set of parameters such that the conditional mean under the null is correctly specified, but it does not check whether the estimation method implied by the null identifies these parameters. This problem is illustrated, in a related context, by the results of Ramalho, Ramalho and Murteira, (2010).

is an important drawback because the null and alternative hypotheses satisfy this relation in many interesting cases, e.g., the ECE model and the M-2PM of Mullahy (1998). The $HPC$ test bypasses this problem by conditioning on the estimates of $\gamma_A$. However, because it does not condition on estimates of $\beta_A$, its performance is likely to be reasonable even if the corrected covariance estimator is not used, which makes it more appealing than the $C$ test. The simulation results in the next section suggest that conditioning on $\gamma_A$ but not on $\beta_A$ is indeed a good compromise.

In his seminal paper, Mullahy (1998) considers several tests to check the specification of models for corner-solutions data which are related to the test proposed here. The first is a split-sample test having as the null the two-part model (Duan et al., 1983), and as alternative the M-2PM. The test is based on a random partition of the original sample into two sub-samples, and compares the slope parameter estimates $\hat{\beta}_A$ obtained from one sub-sample with the slope parameter estimates $\hat{\beta}_B$ obtained in the other. The second test has the ECE model as the null and again the M-2PM as the alternative, and it is a test for whether the slope parameters in $\gamma_B$ are equal to 0.[12] Therefore, like the $HPC$ test, these procedures check the adequacy of a model in the direction of an alternative of interest. However, these tests are very specific and are difficult to extended to check the adequacy of other models, such as the M-2PM or models based on Heckman's (1979) sample selection estimator.

Mullahy (1998) also considers a set of conditional moment tests (Newey, 1985, Tauchen, 1985) which check the orthogonality between the residual $\left(y_i - \hat{y}_i^A\right)$ and regressors and their squares and cross products. Unlike the two other tests described above, these tests can be used to check the validity of any model. However, this approach has a number of drawbacks. First, it is not clear what is the most relevant set of orthogonality conditions to check. Second, checking separately the validity of different sets of moments conditions, as done by Mullahy (1998), makes it difficult to control the global significance level of the family of tests. Finally, although the tests are likely to have power against a wide range of alternatives, they are unlikely to be

[12] Mullahy (1998) shows that the modified two-part model can be estimated in one or two steps. However, this test is valid only when the one-step approach is used.

particularly powerful in directions of special interest. The $HPC$ test is closely related to these conditional moment tests in the sense that it can also be seen as a check of a moment condition. However, it avoids the drawbacks of the procedures proposed by Mullahy (1998) by identifying a single moment condition that is likely to lead to a test that is particularly powerful against alternatives of interest.

Finally, although we focus on corner-solutions data, our results can be extended to the case where $y$ is bounded between 0 and 1, or is a count. Therefore, for example, the $HPC$ test can be used to check for zero-inflation in the models of Gaundry and Dagenais (1979) and Mullahy (1986), and for the presence of under-reporting or under-recording in the models of Feinstein (1989), Winkelmann and Zimmermann (1993), and Mukhopadhyay and Trivedi (1995). In particular, the test can be used to check for zero-inflation in count data models, even when the null hypothesis only specifies the conditional expectation of interest.[13] Stata (StataCorp., 2013) offers the possibility of testing Poisson and negative-binomial count data models against their zero inflated counterparts using Vuong's (1989) test. However, Vuong's (1989) test is likelihood based and therefore it is invalid if the purpose is to compare the ECE model estimated by pseudo-maximum likelihood with a zero inflated counterpart.[14]

## 3. SIMULATION EVIDENCE

In this section we report the results of a simulation study evaluating the finite sample performance of the $HPC$ test and contrasting it with that of some of the

[13]Indeed, if the zero inflation is defined by a probit, the conditional expectation of zero-inflated Poisson or negative binomial models are identical to that of the M-2PM in Table 1. Therefore, the test comparing these zero-inflated models and their standard counterparts can be performed exactly as the test comparing the M-2PM and the ECE. Similar tests can be obtained when the zero inflation is defined by a logit or any other binary choice model. In any case, the test comparing the zero-inflated model and the standard Poisson or negative binomial regression can be performed using the `hpc` command we have written for Stata (StataCorp., 2013).

[14]Additionally, Stata (StataCorp., 2013) implements the version of Vuong's (1989) test for non-nested models when the models being compared are actually overlapping. Therefore the procedure is invalid even if the null hypothesis is that the conditional distribution of $y$ is Poisson.

related procedures discussed before. More specifically, two sets of simulations were performed. In the first set we study the performance of tests comparing the logarithmic specification of Heckman's selection model, estimated by maximum likelihood, with an ECE model, estimated by Poisson pseudo-maximum likelihood (PPML). The second set of simulations studies tests comparing an ECE model, again estimated by PPML, and the M-2PM of Mullahy (1998), in which the first part is a probit and the second part is an ECE model estimated by non-linear least squares. This choice of models to compare is motivated by the empirical illustrations in the next section in which these specifications are tested against each other.

In the experiments in which the double-index models are used to generate the data, the following data generating process is used

$$\Pr\left(y_i > 0|x_i\right) = \Pr\left(0.2x_{i1} + 0.8x_{i2} + e_i > 0|x_i\right),$$
$$\text{for } y_i > 0 : \ln\left(y_i\right) = 1 + 0.8x_{i1} + u_i,$$
$$\begin{bmatrix} e_i \\ u_i \end{bmatrix} |x_i \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}\right),$$

where $x_i = [x_{i1}, x_{i2}]'$. In the first set of experiments, when the data are generated by the M-2PM, we set $\rho = 0$ and $\sigma = 0.5$. For the second set of experiments, when the data are generated using the selection model, $\rho = 0.75$ and $\sigma = 0.5$.

When the ECE model is used to generate the data, we follow Santos Silva and Tenreyro (2011) and obtain $y_i$ as a $\chi^2$ random variable with $\eta_i$ degrees of freedom, where $\eta_i$ are draws from a negative-binomial distribution with

$$\mathrm{E}\left[\eta_i|x_i\right] = \exp\left(0.5 + 0.9x_{i1} + 0.4x_{i2}\right),$$
$$\mathrm{Var}\left[\eta_i|x_i\right] = 3\mathrm{E}\left[\eta_i|x_i\right].$$

In all data generation processes, $x_1$ is obtained as a random draw from the standard normal distribution, $x_2$ is a dummy variable with $\Pr\left(x_2 = 1\right) = 0.4$, and new sets of regressors are drawn for each Monte Carlo replication.[15] In all designs, $y_i$ is equal to zero for about 40 percent of the observations.

The $HPC$ test is performed as a heteroskedasticity-robust t-statistic for $H_0 : \alpha = 0$ versus $H_1 : \alpha = 1$ in (2). When a double-index model is the null, the test is performed in three different ways: with the uncorrected covariance matrix, with the valid covariance matrix obtained as decribed in Appendix A.1, and with the covariance matrix estimated using 50 bootstrap samples. When the single-index model is the null the corrections are not needed.

In addition to the proposed test, these simulations also consider some related test procedures. In particular, we consider heteroskedasticity-robust versions of the $P$ and $C$ tests of Davidson and MacKinnon (1981), and conditional moment (CM) tests as in Mullahy (1998).[16] We note that when the null is a single-index model the heteroskedasticity-robust $P$ test is closely related to the $HPC$ test (the difference being that the former does not use weights to partially account for the heteroskedasticity of the data), and also that the $C$ test is performed with the uncorrected covariance matrix, with the valid covariance matrix obtained along the lines described in Appendix A.1, and with the covariance matrix estimated using 50 bootstrap samples. Like Mullahy (1998), we will also consider CM tests based on the orthogonality between the residuals of the model and covariates and their squares and cross-products. However, the choice of the specific conditional moments to use requires some care. In

[15] Experiments in which $x_1$ is generated as draws from a uniform or $\chi^2_{(8)}$ distribution were also performed but are not reported in detail because the results are not substantially different from those reported here.

[16] The other two tests proposed by Mullahy (1998) are not considered here. The first of these tests compares the traditional two-part model with its modified version proposed by Mullahy. The nature of this problem is very different from the one we consider here and it has been comprehensively discussed by Manning and Mullahy (2001) and Santos Silva and Tenreyro (2006). The second test proposed by Mullahy (1998) compares the exponential conditional function model with the modified two-part model. As noted before, this test is valid only when the double-index model is estimated in a single step. However, this tends to be a difficult task and convergence is often not possible, making the estimator difficult to use in simulations.

our simple models, one obvious choice is to use all relevant moments.[17] Because this approach is unlikely to be feasible in realistic settings, we also consider a one degree of freedom test based only on the square of $x_1$, the only continuous regressor. In what follows, these tests will be labelled CM-Full and CM-$x_1^2$, respectively.

The simulation results were obtained for samples of size 500, 2000, and 8000 and are based on $10,000$ Monte Carlo replications. Table 2 below presents the rejection frequencies at the conventional 5% level for the first set of simulations in which the competing models are the M-2PM and the ECE model. Table 3 presents the corresponding results for the cases in which the sample selection and the ECE models are considered.

The results in Tables 2 and 3 suggest that the $HPC$ test has a reasonable behaviour under the null, at least for sample sizes currently used in empirical applications. This is true whether or not the test is performed using the corrected estimators of the covariance matrix, confirming our conjecture that this correction is relatively unimportant. The closely related $P$ test is also reasonably well behaved under the null, although it is more prone to over-reject. Therefore, the heteroskedasticity correction used by the proposed test seems to improve its behaviour under the null. In contrast, the performance of the $C$ test, either with or without the correction of the covariance, is erratic. Indeed, the $C$ test is sometimes under-sized, but can also be severely oversized, especially when the null is a double-index model. As for the CM tests, CM-$x_1^2$ is always reasonably well behaved under the null, but CM-Full is severely oversized when the null is the M-2PM.

[17] In our setting, this test has 5 degrees of freedom when the null is a double-index model, and 3 degrees of freedom when the null is the single-index model. This is because, by construction, the PPML residuals are orthogonal to the regressors.

**Table 2: Rejection frequencies at the 5% nominal level (M-2PM and ECE models)**

| | Null is true | | | Null is false | | |
|---|---|---|---|---|---|---|
| $n =$ | 500 | 2000 | 8000 | 500 | 2000 | 8000 |
| The null is the modified two-part model | | | | | | |
| $HPC$ Test | 0.0471 | 0.0598 | 0.0502 | 0.7365 | 0.9931 | 0.9984 |
| (Bootstrapped) | 0.0589 | 0.0575 | 0.0489 | 0.7314 | 0.9918 | 0.9980 |
| (Uncorrected) | 0.0700 | 0.0561 | 0.0484 | 0.7257 | 0.9905 | 0.9976 |
| $C$ Test | 0.1634 | 0.1076 | 0.0528 | 0.1622 | 0.6613 | 0.9813 |
| (Bootstraped) | 0.2429 | 0.1763 | 0.1056 | 0.4121 | 0.8017 | 0.9862 |
| (Uncorrected) | 0.0706 | 0.0682 | 0.0462 | 0.1744 | 0.6101 | 0.9524 |
| $P$ Test | 0.0763 | 0.0885 | 0.0850 | 0.0835 | 0.1344 | 0.2529 |
| CM-Full | 0.6098 | 0.6607 | 0.7048 | 0.7757 | 0.9999 | 1.0000 |
| CM-$x_1^2$ | 0.0885 | 0.0751 | 0.0601 | 0.5115 | 0.8135 | 0.9683 |
| The null is the exponential conditional expectation model | | | | | | |
| $HPC$ Test | 0.0750 | 0.0638 | 0.0560 | 0.2213 | 0.4227 | 0.8314 |
| $C$ Test | 0.3539 | 0.1955 | 0.1181 | 0.2382 | 0.3438 | 0.5240 |
| (Bootstrapped) | 0.4280 | 0.2426 | 0.1393 | 0.2643 | 0.3438 | 0.5208 |
| (Uncorrected) | 0.1558 | 0.1122 | 0.0881 | 0.1231 | 0.1958 | 0.3661 |
| $P$ Test | 0.0910 | 0.0857 | 0.0827 | 0.1739 | 0.2055 | 0.3051 |
| CM-Full | 0.0829 | 0.0570 | 0.0533 | 0.3121 | 0.4558 | 0.8108 |
| CM-$x_1^2$ | 0.0702 | 0.0556 | 0.0526 | 0.2289 | 0.2429 | 0.3011 |

The behaviour of the tests under the alternative reveals some interesting features. Given its erratic performance under the null, the results for the $C$ test are difficult to interpret and will not be discussed further. As for the $P$ test, its rejection frequencies are always quite low, even for the largest sample size considered, confirming that this test is not attractive for the purpose we consider here. The performances of the CM tests are somewhat mixed, with CM-Full comprehensively outperforming CM-$x_1^2$. Lastly, the performance of the $HPC$ test is encouraging and, as expected, the loss of power resulting from using the uncorrected covariance matrix is very small and it vanishes reasonably quickly when the sample size increases. Therefore, at least with reasonably sized samples, it may not be necessary to incur the additional costs in computing the corrected test-statistic.

**Table 3: Rejection frequencies at the 5% nominal level (Sample Selection and ECE models)**

| | Null is true | | | Null is false | | |
|---|---|---|---|---|---|---|
| $n =$ | 500 | 2000 | 8000 | 500 | 2000 | 8000 |
| The null is the sample selection model | | | | | | |
| $HPC$ Test | 0.0316 | 0.0413 | 0.0444 | 0.7647 | 0.9995 | 1.0000 |
| (Bootstraped) | 0.0544 | 0.0525 | 0.0459 | 0.7546 | 0.9998 | 1.0000 |
| (Uncorrected) | 0.0527 | 0.0436 | 0.0426 | 0.7590 | 0.9991 | 1.0000 |
| $C$ Test | 0.4470 | 0.2988 | 0.1618 | 0.9424 | 0.9998 | 1.0000 |
| (Bootstraped) | 0.5457 | 0.3494 | 0.1863 | 0.9755 | 0.9999 | 1.0000 |
| (Uncorrected) | 0.2079 | 0.1482 | 0.0973 | 0.9534 | 0.9999 | 1.0000 |
| $P$ Test | 0.1177 | 0.1243 | 0.1006 | 0.1108 | 0.1510 | 0.2329 |
| CM-Full | 0.0668 | 0.0513 | 0.0489 | 0.8569 | 0.9958 | 1.0000 |
| CM-$x_1^2$ | 0.0509 | 0.0460 | 0.0482 | 0.7700 | 0.9636 | 1.0000 |
| The null is the exponential conditional expectation model | | | | | | |
| $HPC$ Test | 0.0843 | 0.0681 | 0.0585 | 0.1966 | 0.3687 | 0.7659 |
| $C$ Test | 0.0234 | 0.0460 | 0.0637 | 0.0069 | 0.0286 | 0.1497 |
| (Bootstraped) | 0.0654 | 0.0802 | 0.0823 | 0.1032 | 0.1540 | 0.2491 |
| (Uncorrected) | 0.0343 | 0.0316 | 0.0236 | 0.0307 | 0.0838 | 0.2170 |
| $P$ Test | 0.1027 | 0.0780 | 0.0623 | 0.1642 | 0.1927 | 0.2929 |
| CM-Full | 0.0839 | 0.0552 | 0.0493 | 0.2799 | 0.3939 | 0.7121 |
| CM-$x_1^2$ | 0.0769 | 0.0551 | 0.0510 | 0.2060 | 0.2211 | 0.2684 |

Overall, the proposed test compares favourably with its competitors. Indeed, the only test that regularly outperforms the $HPC$ test is CM-Full, but this test is less reliable under the null and it is unlikely to be practical in realistic applications with many regressors. It is also interesting to recall that, when the null is the single-index model, the only difference between the $HPC$ and the heteroskedasticity-robust $P$ test is the way heteroskedasticity is taken into account. Therefore, the heteroskedasticity correction used in the $HPC$ test not only improves its behaviour under the null, but it also greatly improves its power. In summary, not only is the $HPC$ test easy to implement, it is also reasonably well behaved both under the null and under the alternative.

It is also noteworthy that all tests are substantially less powerful when the null is the single-index model. This issue deserves further exploration, but we conjecture that this difference in power results from the fact that the ECE model is able to reasonably approximate the true conditional expectation, even when the data is generated by the competing model, while the reverse is not true. Indeed, of all the models considered in these experiments, the ECE model is the only one that directly estimates $\mathrm{E}[y_i|x_i]$ and, therefore, delivers an estimate of the conditional expectation of $y_i$ which is optimal in some sense (depending on the estimation method), even when the model is misspecified. This suggests that, at least for the designs considered here, the ECE model often is flexible enough to approximate the $\mathrm{E}[y_i|x_i]$ implied by the double-index models. A more complete study of the ability of the ECE model to approximate the functional form of the conditional mean of other models is beyond the scope of the present paper.

## 4. EMPIRICAL ILLUSTRATIONS

In this section we illustrate the use of the $HPC$ test with two examples based on well-known data sets, one in international trade and the other in the demand for health care. In both cases, the ECE model is tested against a double-index model and vice-versa. In view of the simulation results presented in the previous section, when the null is the double-index model, only the results of the test based on the bootstrapped and uncorrected covariance matrices are presented, because these are the versions of the test that are more likely to be used in practice.

### 4.1. A gravity model for trade

Santos Silva and Tenreyro (2006) use cross-sectional bilateral export flows data from 137 countries to estimate different specifications of the gravity equation for trade, which is an ECE model. Besides the dependent variable, the dataset includes traditional gravity regressors, such as the GDP of importer and exporter, bilateral distance, and dummies indicating contiguity, common language, colonial ties, access to water, and the existence of preferential-trade agreements. Further details on the data, including sources and descriptive statistics, are provided in Santos Silva and Tenreyro (2006).[18] In this section we use the same data to illustrate the application of the proposed test by testing a gravity equation estimated by the PPML, as in Santos Silva and Tenreyro (2006), against a logarithmic specification of Heckman's (1979) sample selection estimator, used in this context by Hallack (2006); a related estimator is also used by Helpman, Melitz and Rubinstein (2008).

Table 4 presents the main estimation results obtained with the sample selection estimator (estimated by maximum likelihood) and with the ECE model, both with and without the multilateral resistance terms suggested by Anderson and van Wincoop (2003).[19] The last few lines of the Table also include the $R^2$ for each model (computed as the square of the correlation between the dependent variable and the estimated conditional mean), and the $p$-value of the $HPC$ test of the sample selection estimator against the ECE model and vice-versa. We computed the $HPC$ test using both the uncorrected covariance matrix (which is valid when the null is the ECE model) and a covariance matrix estimated using 1000 pairs-bootstrap draws.

Comparing the $R^2$s for the competing models it is possible to see that for both specifications the ECE model fits the data substantially better than the sample selection estimator. However, goodness-of-fit statistics give no indication about the adequacy of the models being contrasted and therefore it is important to use the proposed procedure to test the two models against each other. The results in the last two rows of Table 4 show that, either with or without the multilateral resistance terms, the $HPC$ test clearly rejects the sample selection model, while providing no evidence of departures of the ECE model in the direction of its competitor. The results also show that there is little difference between the $p$-values of the $HPC$ test computed with and without the bootstrap.

[18] These data are available at `http://privatewww.essex.ac.uk/~jmcss/LGW.html`.

[19] The multilateral resistance terms are importer and exporter dummies.

**Table 4: Gravity Equations for Trade**

| Specification: | Selection model 1st part | Selection model 2nd part | ECE | Selection model 1st part | Selection model 2nd part | ECE |
|---|---|---|---|---|---|---|
| Log distance | −0.452 | −1.200 | −0.784 | −0.730 | −1.349 | −0.750 |
| | (0.025) | (0.034) | (0.055) | (0.029) | (0.031) | (0.041) |
| Log exp.'s GDP | 0.461 | 0.979 | 0.733 | — | — | — |
| | (0.009) | (0.012) | (0.027) | — | — | — |
| Log imp.'s GDP | 0.329 | 0.826 | 0.741 | — | — | — |
| | (0.008) | (0.012) | (0.027) | — | — | — |
| Log exp.'s GDP per capita | 0.102 | 0.215 | 0.157 | — | — | — |
| | (0.010) | (0.017) | (0.053) | — | — | — |
| Log imp.'s GDP per capita | 0.110 | 0.115 | 0.135 | — | — | — |
| | (0.010) | (0.017) | (0.045) | — | — | — |
| Common border | −0.491 | 0.256 | 0.193 | −0.657 | 0.170 | 0.370 |
| | (0.112) | (0.129) | (0.104) | (0.118) | (0.128) | (0.091) |
| Common language | 0.334 | 0.709 | 0.746 | 0.320 | 0.408 | 0.383 |
| | (0.039) | (0.067) | (0.135) | (0.050) | (0.067) | (0.093) |
| Colonial ties | 0.158 | 0.412 | 0.024 | 0.301 | 0.668 | 0.079 |
| | (0.040) | (0.070) | (0.150) | (0.053) | (0.069) | (0.134) |
| Landlocked exp. | 0.054 | −0.061 | −0.864 | — | — | — |
| | (0.033) | (0.062) | (0.157) | — | — | — |
| Landlocked imp. | −0.065 | −0.672 | −0.697 | — | — | — |
| | (0.034) | (0.061) | (0.141) | — | — | — |
| Exp.'s remoteness | 0.132 | 0.485 | 0.660 | — | — | — |
| | (0.051) | (0.079) | (0.134) | — | — | — |
| Imp.'s remoteness | −0.043 | −0.204 | 0.561 | — | — | — |
| | (0.052) | (0.085) | (0.118) | — | — | — |
| Free-trade agreement | 1.156 | 0.480 | 0.181 | 1.097 | 0.3058 | 0.376 |
| | (0.163) | (0.100) | (0.088) | (0.181) | (0.098) | (0.077) |
| Openness dummy | 0.295 | −0.130 | −0.107 | — | — | — |
| | (0.027) | (0.053) | (0.131) | — | — | — |
| Multilateral resistance terms | No | No | No | Yes | Yes | Yes |
| Observations | 18360 | 9613 | 18360 | 18360 | 9613 | 18360 |
| $R^2$ | 0.580 | | 0.862 | 0.391 | | 0.928 |
| $HPC$ test p-values | | | | | | |
| Uncorrected | 0.000 | | 0.999 | 0.029 | | 1.000 |
| Bootstraped | 0.000 | | 0.998 | 0.025 | | 1.000 |

Naturally, this result is specific to this particular example and therefore it should not be viewed as indicating that the ECE model is generally preferable to the sample selection specification in applications describing bilateral export flows.

### 4.2. Much ado about two redux

Mullahy (1998) studied the choice between single- and double-index models for the demand for health care. To illustrate the methods considered in the paper, Mullahy (1998) estimates different models for the number of doctor visits during the previous year. The data used are a sample of $36,111$ observations from the 1992 National Health Interview Survey. Besides the dependent variable, the data contains information on a number of covariates: age of the respondent, gender, ethnic background, schooling, marital status, and dummies for health status. Mullahy (1998) provides descriptive statistics and more information about the data.

Table 5 presents the estimation results for the M-2PM proposed by Mullahy (1998) and for the ECE model, also considered by Mullahy (1998), as well as the $R^2$ values and $HPC$ test $p$-values computed as described above.[20] In this particular application, the $R^2$s of the models are virtually identical, which may suggest that there is little to choose between the two models. However, the results of the $HPC$ test provide no evidence against the M-2PM, while clearly rejecting the ECE model. Again, the results of the $HPC$ test with and without the bootstrap are almost identical.

Up to a point, these results are in line with those of Mullahy (1998) who, using a number CM tests and goodness-of-fit criteria, also finds that the M-2PM specification is preferable to the ECE model in this particular data set. However, our results contrast with those of Mullahy (1998) in that he finds that both models fail the CM tests checking the orthogonality between the residuals and regressors and their

[20] The first part of the modified two-part model is a logit, as in Mullahy (1998). Both the second part of the modified two-part model and the exponential conditional expectation model are estimated by non-linear least squares, as in Mullahy (1998). In this particular example, the qualitative results hardly change if a these models are estimated by Poisson pseudo-maximum likelihood, as recommended by Santos Silva and Tenreyro (2006).

squares and cross products. In view of the simulation results presented in the previous section, the rejection of the M-2PM may be just a consequence of the tendency of the CM tests to severely over-reject this particular null hypothesis.

Finally, we emphasize that the finding that the M-2PM outperforms the ECE model is specific to the particular example being considered and should not be taken as evidence that the M-2PM should in general be preferred to the ECE model in health care utilization applications.

**Table 5: Demand for Health Care**

| Specification: | M-2PM $1^{\text{st}}$ part | M-2PM $2^{\text{nd}}$ part | ECE |
|---|---|---|---|
| Age | 0.007 | −0.009 | −0.007 |
| | (0.001) | (0.002) | (0.002) |
| Male | −0.913 | −0.047 | −0.199 |
| | (0.026) | (0.045) | (0.047) |
| White | 0.151 | 0.132 | 0.175 |
| | (0.034) | (0.052) | (0.054) |
| Schooling | 0.103 | 0.037 | 0.055 |
| | (0.004) | (0.007) | (0.007) |
| Married | 0.120 | −0.148 | −0.130 |
| | (0.028) | (0.041) | (0.043) |
| Health: Excellent | −1.394 | −1.612 | −1.828 |
| | (0.057) | (0.048) | (0.049) |
| Health: Very Good | −1.056 | −1.340 | −1.480 |
| | (0.057) | (0.045) | (0.046) |
| Health: Good | −0.898 | −0.856 | −0.973 |
| | (0.057) | (0.044) | (0.045) |
| Observations | 36,111 | 27,598 | 36,111 |
| $R^2$ | 0.078 | | 0.078 |
| $HPC$ test p-values | | | |
| Uncorrected | 0.682 | | 0.000 |
| Bootstapped | 0.673 | | 0.000 |

## 5. CONCLUDING REMARKS

The choice of the most appropriate model for corner-solutions data has been the subject of numerous studies and even some controversy. In this paper we argue that this problem should be addressed as a test of non-nested hypotheses and propose an easily implementable regression-based test which is particularly suited to discriminate between single- and double-index models. Moreover, the proposed test explicitly takes into account the heteroskedasticity that is likely to be present in data of this type, and has an intuitive interpretation in terms of orthogonality conditions. We present the results of a simulation study which suggest that the proposed test is reasonably well behaved both under the null and under the alternative, at least for the sample sizes that are commonly used in empirical studies. Moreover, the test compares very favourably with alternative procedures that could be used for this purpose. Two illustrative applications show that the test can be quite useful in practice.

## APPENDIX

### A1. Asymptotic distribution and adjusted covariance matrix

The proposed test is based on the OLS estimation of an artificial model of the form

$$\frac{y_i - \hat{y}_i^A}{\sqrt{\hat{y}_i^A}} = \frac{\nabla g_A\left(x_i'\hat{\beta}_A\right) F_A\left(x_i'\hat{\gamma}_A\right) x_i'}{\sqrt{\hat{y}_i^A}}\delta + \alpha\frac{\hat{y}_i^B - \hat{y}_i^A}{\sqrt{\hat{y}_i^A}} + \nu_i.$$

The easiest way of obtaining the asymptotic distribution of the OLS estimates of $\theta = (\delta, \alpha)$, say $\hat{\theta} = \left(\hat{\delta}, \hat{\alpha}\right)$, is to consider the joint estimation of $\theta$ and $\phi_A = (\beta_A, \gamma_A)$ by system GMM as in Newey (1984). The results in this appendix are presented for the case in which $\beta_A$ and $\gamma_A$ are jointly estimated by maximum likelihood, as in Heckman's (1979) selection model. For cases such as the two-part model in which $\gamma_A$ can be estimated independently of $\beta_A$, the same results are valid if one considers only the moment conditions for the joint estimation of $\gamma_A$ and $\theta$.

Let $S_1$ and $S_2$ denote the vector of moment conditions for the model under the null and for the test equation, respectively, and define $S(\lambda) = (S_1', S_2')'$, with

$\lambda = (\beta'_A, \gamma'_A, \delta, \alpha)'$. The just-identified system-GMM estimator of $\lambda$ is defined as the solution of

$$n^{-1}\sum_{i=1}^{n} S\left(\hat{\lambda}\right) = 0,$$

where $\hat{\lambda} = \left(\hat{\beta}_A, \hat{\gamma}_A, \hat{\delta}, \hat{\alpha}\right)$, and we assume the following standard regularity conditions (see, e.g., theorems 2.6 and 3.4 in Newey and McFadden, 1994).

**A1** $\mathrm{E}\left(S\left(\lambda\right)\right) = 0$ only if $\lambda = \lambda_0$, where $\lambda_0$ denotes the true value of $\lambda$.

**A2** $\lambda_0 \in$ interior of $\Lambda$, which is compact.

**A3** $S\left(\lambda\right)$ is continuous at each $\lambda \in \Lambda$ with probability one.

**A4** With probability approaching one $S\left(\lambda\right)$ is continuously differentiable in a neighborhood $\varsigma$ of $\lambda_0$.

**A5** $\mathrm{E}\left(\sup_{\lambda\in\Lambda}\left\|S\left(\lambda\right)\right\|\right) < \infty$, $\mathrm{E}\left(\left\|S\left(\lambda_0\right)\right\|^2\right) < \infty$, and $\mathrm{E}\left(\sup_{\lambda\in\varsigma}\left\|\frac{\partial}{\partial\lambda'}S\left(\lambda\right)\right\|\right) < \infty$.

**A6** The matrix $M'M$ is non-singular, where $M = \mathrm{E}\left(\frac{\partial}{\partial\lambda'}S\left(\lambda_0\right)\right)$.

Then, the results in Newey and McFadden (1994) imply that

$$\sqrt{n}\left(\hat{\lambda} - \lambda\right) \xrightarrow{d} \mathcal{N}\left(0, M^{-1}\Sigma M^{-1\prime}\right),$$

where

$$\Sigma = \mathrm{E}\begin{bmatrix} S_1 S_1' & S_1 S_2' \\ S_2 S_1' & S_2 S_2' \end{bmatrix}.$$

Noting that

$$M^{-1} = \begin{bmatrix} H^{-1} & 0 \\ -H_2^{-1}H_1H^{-1} & H_2^{-1} \end{bmatrix},$$

where $H$ denotes the expectation of the matrix of derivatives of $S_1$ with respect to $\phi_A$ and $H_1$ and $H_2$ denote the expectation of the derivatives of $S_2$ with respect to $\phi_A$ and $\theta$, respectively, the variance of $\hat{\theta}$ can then be written as

$$\begin{aligned} \mathrm{V}\left(\hat{\theta}\right) &= H_2^{-1}\mathrm{E}\left(S_2S_2'\right)H_2^{-1\prime} - H_2^{-1}\mathrm{E}\left(S_2S_1'\right)H^{-1\prime}H_1'H_2^{-1\prime} \\ &\quad - H_2^{-1}H_1H^{-1}\mathrm{E}\left(S_1S_2'\right)H_2^{-1\prime} + H_2^{-1}H_1H^{-1}\mathrm{E}\left(S_1S_1'\right)H^{-1\prime}H_1'H_2^{-1\prime}, \end{aligned}$$

or

$$\mathrm{V}\left(\hat{\theta}\right) = \mathrm{V}_{\hat{\theta}} + H_2^{-1}\left\{H_1\mathrm{V}\left(\hat{\phi}_A\right)H_1' - \mathrm{E}\left(S_2S_1'\right)H^{-1\prime}H_1' - H_1H^{-1}\mathrm{E}\left(S_1S_2'\right)\right\}H_2^{-1\prime},$$

where $\mathrm{V}\left(\hat{\phi}_A\right)$ is the estimated variance of $\hat{\phi}_A = \left(\hat{\beta}_A, \hat{\gamma}_A\right)$ and $\mathrm{V}_{\hat{\theta}}$ is the uncorrected estimated variance of $\hat{\theta}$.

Whether $\mathrm{V}\left(\hat{\theta}\right)$ is smaller, larger, or equal to $\mathrm{V}_{\hat{\theta}}$, in the positive semidefinite sense, depends on the particular case being considered. For example, if $H_1 = 0$, the two matrices are equal and when $\mathrm{E}\left(S_2 S_1'\right) = 0$, $\mathrm{V}\left(\hat{\theta}\right)$ is larger than $\mathrm{V}_{\hat{\theta}}$ in the positive semidefinite sense.

In the context of the $HPC$ test, it is of special interest to consider the case where $\gamma_A$ is estimated by maximum likelihood. In this case, $\mathrm{V}\left(\hat{\phi}_A\right) = -H^{-1}$ and $\mathrm{E}\left(S_2 S_1'\right) = -H_1$, and therefore

$$\mathrm{V}\left(\hat{\theta}\right) = \mathrm{V}_{\hat{\theta}} - H_2^{-1} H_1 \left(-H^{-1}\right) H_1' H_2^{-1\prime},$$

implying that $\mathrm{V}\left(\hat{\theta}\right)$ is smaller than $\mathrm{V}_{\hat{\theta}}$ (see Pierce, 1982, and Lee, 2010, pp. 104-5). Therefore, when $\gamma_A$ is estimated by maximum-likelihood, the test-statistic constructed using the uncorrected covariance will have variance smaller than 1 and, therefore, the test will be asymptotically undersized.

Finally, we reiterate that the correction of the covariance matrix is needed only when $M_A$ is a double-index model.

## A2. Correlation in the Two-Part Model

In Duan et al. (1984) an example is given that argues that there can be correlation between the two error terms in the two-part model and that therefore this model is not nested by the sample selection model, in the sense that the two-part model cannot be obtained by imposing a restriction on the selection model. Since then, numerous papers have quoted this result, e.g. Leung and Yu (1996) and Norton et al. (2008). Here we argue that the example is misleading.

In the notation of Duan et al. (1984), the two-part model is given by

$$\begin{aligned}
I_i &= x_i'\delta_1 + \eta_{1i}, \;\; \eta_{1i}|x_i \sim N(0,1) \\
\ln(y_i) &= x_i'\delta_2 + \eta_{2i} \\
(\eta_{2i}|I_i > 0, x_i) &\sim f\left(0, \sigma^2\right) \\
(\eta_{2i}|I_i \leq 0, x_i) &\equiv -\infty \qquad (y_i \equiv 0)
\end{aligned}$$

where $f$ is a continuous distribution with mean zero and variance $\sigma^2$. Hence, $(\ln(y_i)\,|I_i > 0, x_i) \sim f(x_i'\delta_2, \sigma^2)$.

To show that correlation between $\eta_1$ and $\eta_2$ is possible Duan et al. (1984) constructed the following example (pages 285-286): Let $Z_{1i}$ and $Z_{2i}$ follow a standard bivariate normal distribution with correlation coefficient $\rho$. Let $G_i$ be the left- and $H_i$ be the right-truncated standard normal cdf, with $-x_i'\delta_1$ as truncation point:

$$\begin{aligned}
G_i(u) &= \int_{-x_i'\delta_1}^{u} \phi(z)\,dz/\Phi(x_i'\delta_1), \qquad -x_i'\delta_1 \leq u, \\
H_i(v) &= \int_{-\infty}^{v} \phi(z)\,dz/\Phi(x_i'\delta_1), \qquad v \leq -x_i'\delta_1,
\end{aligned}$$

where $\phi$ denotes the standard normal pdf.

Construct $(\eta_{1i}, \eta_{2i})$ as follows: With probability $\Phi(x_i'\delta_1)$, define

$$\eta_{1i} = G_i^{-1}(\Phi(Z_{1i}))\,; \quad \eta_{2i} = f^{-1}(\Phi(Z_{2i}))\,.$$

With probability $(1 - \Phi(x_i'\delta_1))$ define

$$\eta_{1i} = H_i^{-1}(\Phi(Z_{1i}))\,; \quad \eta_{2i} = -\infty.$$

Then the two-part model assumptions are satisfied and there is correlation between $\eta_{1i}$ and $\eta_{2i}$. Duan et al. (1984) show that when $f$ is assumed to be normal then the conditional expectation is given by

$$\begin{aligned}
E(\eta_{2i}|\eta_{1i}) &= \rho\sigma\Phi(G_i(\eta_{1i}))\,, \qquad \eta_{1i} > -x_i'\delta_1, \\
E(\eta_{2i}|\eta_{1i}) &\equiv -\infty, \qquad \eta_{1i} \leq -x_i'\delta_1
\end{aligned}$$

and consequently $\eta_{1i}$ and $\eta_{2i}$ are stochastically dependent and positively associated.

The problem with this argument lies in the fact that with probability $\Phi(x_i'\delta_1)$ we draw an $\eta_{1i}$ such that $\eta_{1i}$ is larger than $-x_i'\delta_1$. This essentially introduces a new uniformly distributed random variable, say $\zeta_i$, and changes the model to

$$\begin{aligned}
\zeta_i|x_i &\sim U(0,1) \\
I_i &= x_i'\delta_1 + \eta_{1i} \\
I_i &> 0 \quad \text{if} \quad \zeta_i < \Phi(x_i'\delta_1),
\end{aligned}$$

so $\zeta_i$ determines the outcome $I_i > 0$ and is independent of $\eta_{2i}$. Therefore, there is no selection problem, as

$$\begin{aligned}
E(\ln(y_i)|I_i > 0) &= x_i'\delta_2 + E(\eta_{2i}|I_i > 0) \\
&= x_i'\delta_2 + E(\eta_{2i}|\zeta_i < \Phi(x_i'\delta_1)) \\
&= x_i'\delta_2.
\end{aligned}$$

Clearly, the model of the example can be specified as

$$\begin{aligned}
\zeta_i|x_i &\sim U(0,1) \\
I_i^* &= 1(\zeta_i < \Phi(x_i'\delta_1)) \\
\ln(y_i) &= x_i'\delta_2 + \eta_{2i} \\
(\eta_{2i}|I_i^* = 1, x_i) &\sim f\left(0, \sigma^2\right) \\
(\eta_{2i}|I_i^* = 0, x_i) &\equiv \infty \qquad (y_i \equiv 0),
\end{aligned}$$

with $\zeta_i$ independently distributed of $\eta_{2i}$ and the value of $\eta_{1i}$ is immaterial. Therefore, this example does not show that the errors $\eta_{1i}$ and $\eta_{2i}$ in the original model can be correlated.

In summary, under the maintained assumptions, there is no evidence to support the view that the two-part model cannot be obtained by imposing a restriction on the sample selection model. However, the assumptions of the sample selection model are unlikely to hold when it is used to describe corner solutions data, and in that case there is no guaranty that the conditional expectation implied by the sample selection model will fit the data better than the conditional expectation implied by the two-part model. For example, if $\eta_{2i}$ is homoskedastic but non-normal, the two-part model can be used to consistently estimate the conditional expectation of $y_i$, while that is not possible with the sample selection model. In that sense, the two models are indeed not nested.

## REFERENCES


Anderson, J. and Yotov, Y. (2010). "The changing incidence of geography," *American Economic Review* 100, 2157-2186.

Anderson J. and van Wincoop, E. (2003). "Gravity with gravitas: a solution to the border puzzle," *American Economic Review* 93, 170-192.

Arkolakis, C. (2008). *Market penetration costs and the new consumers margin in international trade*, NBER working paper No. 14214.

Arkolakis, C., Costinot, A. and Rodríguez-Clare, A. (2009). *New trade models, same old gains?*, NBER Working Paper No. 15628.

Atkinson, A.C. (1970). "A method for discriminating between models," *Journal of the Royal Statistical Society*, Series B, 32, 323-353.

Bierens, H.J. (1982). "Consistent model specification tests," *Journal of Econometrics*, 20, 105-134.

Bierens, H.J. (1990). "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.

Chaney, T. (2008). "Distorted gravity: The intensive and extensive margins of international trade," *American Economic Review* 98, 1707-1721.

Cox, D.R. (1961). "Tests of separate families of hypotheses," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, University of California Press, Berkeley, 105-123.

Davidson, R. and MacKinnon, J.G. (1981). "Several tests for model specification in the presence of alternative hypotheses," *Econometrica* 49, 781-793.

Davidson, R. and MacKinnon, J.G. (2006). "Bootstrap methods in econometrics," in Mills, T.C. and Patterson, K. (eds.) *Palgrave Handbook of Econometrics*, Vol. 1. Ch. 23, 821-838, London: Palgrave Macmillan.

Deaton, A. and Irish, M. (1984), "Statistical models for zero expenditures in household budgets," *Journal of Public Economics* 23, 59-80.

Dow, W.H. and Norton, E.C. (2003). "Choosing between and interpreting the Heckit and two-part models for corner solutions," *Health Services & Outcomes Research Methodology*, 4, 5-18.

Duan, N. (1983). "Smearing estimate: a nonparametric retransformation method," *Journal of theAmerican Statistical Association*, 78, 605-610.

Duan, N., Manning, W.G., Morris, C.N., and Newhouse, J.P. (1983). "A comparison of alternative models for the demand for medical care," *Journal of Business and Economic Statistics* 1, 115-126.

Duan, N., Manning, W.G., Morris, C.N., and Newhouse, J.P. (1984). "Choosing between the sample-selection model and the multi-part model," *Journal of Business and Economic Statistics* 2, 283-289.

Eaton, J. and A. Tamura (1994). "Bilateralism and regionalism in Japanese and US trade and direct foreign investment patterns," *Journal of the Japanese and International Economics* 8, 478-510.

Feinstein, J.S. (1989). "The safety regulation of U.S. nuclear power plants: Violations inspections, and abnormal occurrences," *Journal of Political Economy* 97, 115-154.

Fisher, G., and McAleer. M. (1979). "On the interpretation of the Cox test in econometrics," *Economics Letters* 4, 145-150.

Freedman, D.A. (1981). "Bootstrapping regression models." *Annals of Statistics* 9, 1218-1228.

Gaundry, M.J.I. and Dagenais, M.G. (1979). "The dogit model," *Transportation Research Part B: Methodological*, 13, 105-111.

Gourieroux, C. and Monfort, A. (1994). "Testing non-nested hypotheses", in Engle, R.F. and McFadden, D. (eds.) *Handbook of econometrics*, Vol. IV, Ch. 44, 2583-2637, Amsterdam: Elsevier.

Hallak, J.C. (2006). "Product quality and the direction of trade," *Journal of International Economics* 68, 238-265.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica* 47, 153-161.

Helpman, E., Melitz, M.J. and Rubinstein, Y. (2008), "Estimating trade flows: Trading partners and trading volumes," *Quarterly Journal of Economics* 123, 441-487.

Helpman, E., Melitz, M.J. and Yeaple, S.R. (2004). "Export versus FDI with heterogeneous firms," *American Economic Review* 94, 300-316.

Jones, A.M. (2000). "Health Econometrics," in Newhouse, J.P. and Culyer, A.J. (eds.) *Handbook of Health Economics*, Vol. 1A, Ch. 6, 265-344, Amsterdam: Elsevier.

La Porta, R., López-de-Silanes, F. and Zamarripa, G. (2003). "Related lending," *The Quarterly Journal of Economics* 118, 231-268.

Lee, M.-J. (2010). *Micro-econometrics: Methods of moments and limited dependent variables*, 2nd ed., New York, NY: Springer.

Leung, S.F. and Yu., S. (1996). "On the choice between sample-selection and two-part models," *Journal of Econometrics* 72, 197-229.

Mukhopadhyay, K. and Trivedi, P.K. (1995). *Regression models for under-recorded count data*, Paper presented at the Econometric Society 7th World Congress, Tokyo.

Melitz, M.J. (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica* 71, 1695-1725.

Mullahy, J. (1986). "Specification and testing in some modified count data models," *Journal of Econometrics* 33, 341-365.

Mullahy, J. (1998). "Much ado about two: Reconsidering retransformation and the two-part model in health econometrics," *Journal of Health Economics* 17, 247-282.

Newey, W.K. (1984). "A method of moments interpretation of sequential estimators," *Economics Letters* 14, 201-206.

Newey, W.K. (1985). "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047-1070.

Newey, W.K. and McFadden, D. (1994). "Large sample estimation and hypothesis testing," in Engle, R.F. and McFadden, D. (eds.), *Handbook of Econometrics*, Vol. 4, Ch. 36, 2111-2245, Amsterdam: Elsevier.

Norton, E.C., Dow, W.H. and Do, Y.K. (2008). "Specification tests for the sample selection and two-part models," *Health Services and Outcomes Research Methodology* 8, 201-208.

Pierce, D.A. (1982). "The asymptotic effect of substituting estimators for parameters in certain types of statistics," *Annals of Statististics* 10, 475-478.

Pesaran, M.H. and Deaton, A.S. (1978), "Testing non-nested nonlinear regression models," *Econometrica* 46, 677-694.

Quandt, R.E. (1974). "A comparison of methods for testing non-nested hypotheses," *Review of Economics and Statistics* 56, 251-255.

Ramalho, E.A., Ramalho, J.J.S. and Murteira, J.M.R. (2011). "Alternative estimating and testing empirical strategies for fractional regression models," *Journal of Economic Surveys* 25, 19-68.

Santos Silva, J.M.C. (2001). "A score test for non-nested hypotheses with applications to discrete data models," *Journal of Applied Econometrics* 16, 577-597.

Santos Silva, J.M.C. and Tenreyro, S. (2006), "The log of gravity," *The Review of Economics and Statistics* 88, 641-658.

Santos Silva, J.M.C. and Tenreyro, S. (2011). "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator," *Economics Letters* 112, 220-222.

StataCorp. (2013). *Stata Release 13. Statistical Software*, College Station (TX): StataCorp LP.

Tauchen, G. (1985). "Diagnostic testing and evaluation of maximum likelihood models," *Journal of Econometrics*, 30, 415-443.

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

van de Ven, W.P. and van Praag, B.M. (1981). "Risk aversion of deductibles in private health insurance: Application of an adjusted Tobit model to family health care expenditures," in van der Gaag, J. and Perlman, M. (eds.) *Health, Economics and Health Economics*, 125-148, Amsterdam: North Holland.

Vuong, Q.H. (1989). "Likelihood Ratio tests for model selection and non-nested hypotheses," *Econometrica* 57, 307-333.

Winkelmann, R. and Zimmermann, K.F. (1993). *Poisson logistic regression*, Department of Economics, University of Munich, Working Paper No 93-18.

Wooldridge, J.M. (1992). "A Test for Functional Form Against Nonparametric Alternatives," *Econometric Theory* 8, 452-475.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.