Is it different for zeros? Discriminating between models for non-negative data with many zeros*

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

27 January 2011

Abstract

In many economic applications, the variate of interest is non-negative and its distribution is characterized by a mass-point at zero and a long right-tail. Many regression strategies have been proposed to deal with data of this type. 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, or to assess the validity of the preferred model, are not often used in practice. In this paper we propose a novel and simple regression-based specification test that can be used to test these models against each other.

JEL classification code: C12, C52

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

---

*We are grateful to Holger Breinlich, Francesco Caselli, Daniel Dias, Esmeralda Ramalho, and Joaquim Ramalho for helpful comments, and to John Mullahy for providing one of the datasets used in Section 5. Santos Silva also gratefully 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.

†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 empirical 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 datasets with these characteristics are also found in many other areas.[1] A stark example of data with a mixed distribution 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). What is common to all the cases we are considering is that the many zeros are not the result of some observability problem, but rather correspond to the existence of the so-called "corner-solutions" (Wooldridge, 2002).[2] In this type of application, researchers and policymakers are ultimately interested in the effects of the covariates on the distribution of the fully observable dependent variable y.

Linear models are generally inappropriate to describe corner-solutions data, and many alternative specifications have been suggested to try to accommodate the peculiar characteristics of this sort of data. The specifications that have been suggested can broadly be divided into one- and two-equation models. The one-equation models include different versions of the Tobit (Tobin, 1958, Eaton and Tamura, 1994), and models with an exponential conditional expectation function (Mullahy, 1998, and Santos Silva and Tenreyro, 2006). The two-equation models allow the covariates to

[1] See Jones (2000) for a survey of applications in health economics, and Helpman, Melitz, and Rubinstein (2008), and Anderson and Yotov (2010), for recent examples of applications in international economics. La Porta, López-de-Silanes, and Zamarripa (2003) is an example of the use of this type of data in finance.

[2] Therefore, this situation is different from the sample selection found in labour economics where the zeros in wage data are a just a convenient way to indicate that the individual does not participate in the labour market.

affect the conditional distribution of y in two different ways, modelling separately the probability of observing a zero and the positive observations, or, more frequently, their logarithm. The class of two-equation models includes the so-called two-part model of Duan, Manning, Morris, and Newhouse (1983), and models based on Heckman's (1979) sample selection estimator.

Although there has been a long debate in the literature on the relative merits of different models for corner-solutions data (see, e.g., Duan et al., 1983 and 1984, Hay and Olsen, 1984, Manning, Duan, and Rogers, 1987, and Mullahy, 1998), in practice, the choice between the competing specifications is often based on convenience, or on the beliefs of the researcher about how the zeros where generated. Strikingly, formal statistical tests are rarely used to help choosing the most suitable specification, or to assess the validity of the preferred model.

However, having an appropriate test to choose between competing models is important for several reasons. First, because none of the proposed specifications 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 one-equation specification with minimal distributional assumptions at the estimation stage; instead, the model of Helpman, Melitz and Rubinstein (2008) leads to a two-equation specification. The gain in flexibility provided by the two-part model in this case comes at the cost of stronger distributional assumptions at the estimation stage; the test we propose implicitly weighs in these costs and benefits to judge the appropriateness of the different specifications for a given dataset. Even if the researcher favours one specification on theoretical grounds, it is important to check its adequacy by testing it against competing specifications; this can help confirm (or reject) the researcher's views on the models. Finally, the choice of the correct model 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.

One of the reasons that may explain why statistical tests are not routinely used to choose between competing models for corner-solutions data is that it is not obvious how such tests could be performed. Indeed, because the various specifications that have been proposed are based on very different modelling strategies, it is not immediately clear what test can be used to choose between the competing alternatives.

In this paper we argue that existing econometric tests are inappropriate to discriminate between alternative models for corner-solutions data and we propose a novel regression-based test that can achieve this goal. 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, implicitly or explicitly they all define the conditional expectation of y. 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.

The remainder of the paper is organised as follows. The next section summarizes the more popular competing specifications that have been used to model corner-solutions data. Section 3 describes the testing strategy we adopt and the proposed specification test. Section 4 presents the results of a small simulation study illustrating the finite sample performance of the proposed test, and Section 5 employs two well-known datasets to illustrate the practical use of the approach we suggest. Finally, Section 6 contains brief concluding remarks and an Appendix gives technical details on the proposed testing procedure.

2. A CATALOGUE OF MODELS

Table 1 lists some prominent nonlinear models that have been used in an attempt to deal with the challenges posed by corner-solutions data.[3] Besides the basic description of the model, Table 1 also includes the form of $\mathrm{E}[y|x]$, the conditional expectation of y given a set of covariates x, implied by the different specifications.

Model 1, the Exponential Conditional Expectation (ECE) model, is the simplest specification and it can be estimated using a pseudo-maximum likelihood estimator of the family considered by Gourieroux, Monfort and Trognon (1984). This specification has been used, for example, by Santos Silva and Tenreyro (2006) and Anderson and Yotov (2010) in the context of trade data. Mullahy (1998) suggested this specification to model the demand for health care, and Manning and Mullahy (2001) studied its performance for the case in which the dependent variable is strictly positive. Although it is not pursued here, we note that this model can be made more flexible by the introduction of a shape parameter, as in Wooldridge (1992) and Basu, Arondekar and Rathouz (2006).

TABLE 1 ABOUT HERE

Model 2 is the two-part model (2PM) proposed by Duan et al. (1983). This specification has been extensively used to model demand for health care and it is described in Wooldridge's (2002) textbook. The use of this model has led to some controversy

[3] This list is by no means exhaustive. Other models that have been used in this context include the four-part model of Duan et al. (1983), the threshold Tobit of Eaton and Tamura (1994), the generalized gamma model of Manning, Basu, and Mullahy (2005), and the two-equation model of Helpman, Melitz and Rubinstein (2008), among many others.

in the literature, which is elegantly summarized in Jones (2000). Model 3 is the modified two-part model (M-2PM) of Mullahy (1998), in which the first part is also a binary model for the probability that $y_i > 0$.[4] The second part is an ECE model estimated on the positive observations only. Although this model has seen little use in practice, it is an attractive alternative to the standard 2PM because it leads to a specification of $\mathrm{E}[y|x]$ which does not depend on incidental assumptions about the distribution of the errors in the second stage.

Model 4 is the Tobit (Tobin, 1958), which has often been used to model non-censored data with many zeros. See Wooldridge (2002) for a discussion of the use of the Tobit in this context and for a derivation of $\mathrm{E}[y|x]$. Leading examples of the use of the Tobit to model non-censored data with mixed distributions are La Porta, López-de-Silanes and Zamarripa (2003), and Rose (2004).[5]

Finally, Models 5 and 6 are based on Heckman's (1979) sample selection estimator and have been extensively used to model data with mixed distributions.[6] Wooldridge (2002) discusses the use of the sample selection models in this context and gives the expression of $\mathrm{E}[y|x]$ for the model in levels (Model 5). The expression of $\mathrm{E}[y|x]$ for the model in logs (Model 6) can be traced back to van de Ven and van Praag (1981). For examples of the use of these models in health economics, see, among many others, the discussion in Mullahy (1998) and the survey of Jones (2000). Hallack (2006) and Helpman, Melitz and Rubinstein (2008) are leading examples of the use of this approach in modelling international trade.

[4] Mullahy (1998) models the first stage as a logit, but he notes that other binary choice models can be used. Here we use a probit in the first part because that tends to be the choice of most practitioners.

[5] Eaton and Tamura (1994) proposed a modified Tobit estimator that has been used mainly in the context of the estimation of gravity models.

[6] When Models 5 and 6 are used to describe genuine sample selection data, $\Pr(y_i > 0|x_i)$ should be interpreted as the conditional probability that y_i is observed.

It is important to note that the models in Table 1 not only specify $\mathrm{E}[y|x]$, but they also prescribe a method to estimate the parameters of interest. Therefore, although some models (1, 2, 3, and 6) essentially specify conditional means of the form $\exp\left(x_i'\beta_j^* + \ln\Phi\left(x_i'\gamma_j^*\right)\right)$,[7] in practice, the estimates of $\mathrm{E}[y|x]$ they lead to can be very different because they are evaluated at different estimates of the parameters, whose probability limits will in general be distinct. Notice also that for Models 2 and 6, the parameters of interest cannot be estimated using just information from the conditional mean as it does not identify the intercepts of at least one of the two parts of the model.

3. TESTING STRATEGIES

Common to all the models regularly used to describe corner-solutions data is that, implicitly or explicitly, they all specify the conditional mean of y given x. Moreover, $\mathrm{E}[y|x]$ is actually the object of interest in the kind of applications we have in mind 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, like Mullahy (1998), we compare the different models on the basis of the adequacy of the implied conditional expectations.

Naturally, as done by Mullahy (1998), one can use standard tests to separately check the specification of $\mathrm{E}[y|x]$ in each model, for example using some version of the simple RESET test (Ramsey, 1969, and Ramsey and Schmidt, 1976). However, this approach ignores the information provided by the alternative models and therefore we suggest exploring this information by 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 hypothesis.

[7] The notation β_j^* and γ_j^* is used to denote β_j and γ_j with the intercepts appropriately shifted. The ECE specification is obtained, for example, by letting the intercept in γ^* pass to infinity.

The motivation for using tests for 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 Heckit does not nest, and is not nested by, its logarithmic specification. But, perhaps less obviously, we argue that the use of the non-nested hypotheses framework is justified (and indeed 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, as noted above, the models imply not only a functional form for $\mathrm{E}[y|x]$, but also an estimation method for its parameters. 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 points, 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.

Most tests for non-nested hypotheses require the specification of the entire conditional density of y given x (Cox, 1961, Atkinson, 1970, Quandt, 1974, Gourieroux and Monfort, 1994, and Santos Silva, 2001), and therefore are not appropriate in this context. However, Davidson and MacKinnon (1981) introduced several tests for non-nested hypotheses that only require the specification of the conditional mean, and therefore are appropriate for our purpose. In this section we build on the results of Davidson and MacKinnon (1981) to develop testing procedures to discriminate between competing models for corner-solutions data.

3.1. The *P* and C tests

As in Davidson and MacKinnon (1981), suppose that we want to test Model A, characterized by

$$M_A : \mathrm{E}[y_i|x_i] = f_A(x_i, \lambda),$$

against Model B, which implies

$$M_B : \mathrm{E}\left[y_i|x_i\right] = f_B\left(x_i, \mu\right).$$

The approach suggested by Davidson and MacKinnon (1981) to test M_A against alternatives in the direction of M_B is based on the nesting of M_A and M_B within an artificial compound model of the form

$$M_C : \mathrm{E}\left[y_i|x_i\right] \; = \left(1-\alpha\right) f_A\left(x_i, \lambda\right) + \alpha f_B\left(x_i, \mu\right),$$

where the correct specification of M_A can be checked by testing $H_0 : \alpha = 0$ against $H_1 : \alpha = 1$.[8] Essentially, this test checks a moment condition of the form

$$E\left[\left(y_i - f_A\left(x, \lambda\right)\right)\left(f_B\left(x_i, \mu\right) - f_A\left(x_i, \lambda\right)\right)\right] = 0.$$

That is, the test checks the correct specification of $f_A\left(x, \lambda\right)$ by testing whether the errors $y - f_A\left(x, \lambda\right)$ have zero expectation, giving more weight to the observations for which the difference between the conditional means implied by M_A and M_B are larger.

In general, unrestricted estimation of M_C does not identify α. Therefore, Davidson and MacKinnon (1981) suggest performing the test conditioning on the estimates of μ obtained under the alternative. Moreover, Davidson and MacKinnon (1981) note that estimation of this non-linear model can be avoided by using a linearization of M_C around $\lambda = \hat{\lambda}$, where $\hat{\lambda}$ denotes the estimate of λ under the null. That is, the test can be performed as a t-test for the significance of α in the auxiliary linear regression

$$y_i - f_A\left(x_i, \hat{\lambda}\right) = \nabla_\lambda f_A\left(x_i, \hat{\lambda}\right)\delta + \alpha\left(f_B\left(x_i, \hat{\mu}\right) - f_A\left(x_i, \hat{\lambda}\right)\right) + \eta_i,$$

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

where $\nabla_\lambda f_A\left(x_i,\hat{\lambda}\right)$ denotes a vector containing the derivatives of $f_A(x,\lambda)$ with respect to λ, evaluated at $\hat{\lambda}$, and $\hat{\mu}$ represents any consistent estimate of μ. This procedure is named P test by Davidson and MacKinnon (1981).[9]

The P test can be used to discriminate between many of the models described in the previous section. In particular, it can be used whenever $f_B(x_i,\mu)$ is not a special case of $f_A(x_i,\lambda)$. For example, the P test can be used to test the Heckit against its logarithmic specification, and vice-versa, and it is also appropriate when the ECE is the null and the Heckit or the two-part models are the alternative.

However, because the P test does not condition on the estimates obtained under the null, the test will have power equal to size when $f_A(x_i,\lambda)$ and $f_B(x_i,\mu)$ have the same functional form, or when $f_B(x_i,\mu)$ is a special case of $f_A(x_i,\lambda)$.[10] 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. Therefore, the P test cannot be used to compare some of the more popular models, like the Heckit in logs and the 2PM, because both these models specify conditional expectations of the form $\exp\left(x_i'\beta_j^*+\ln\Phi\left(x_i'\gamma_j^*\right)\right)$, where the notation β_j^* and γ_j^* is used, as before, to denote β_j and γ_j with the intercepts appropriately shifted.

A possible avenue to follow at this point would be to resort to the seldom-used C test (Davidson and MacKinnon, 1981), and estimate M_C conditioning both on $\hat{\lambda}$ and on $\hat{\mu}$. That is, the C test would check the significance of α in the auxiliary linear regression

$$y_i-f_A\left(x_i,\hat{\lambda}\right)=\alpha\left(f_B\left(x_i,\hat{\mu}\right)-f_A\left(x_i,\hat{\lambda}\right)\right)+\zeta_i.$$

The disadvantage of this procedure, however, is that the standard t-statistic for the significance of α ignores that the dependent variable in the regression is evaluated at

[9]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 test because its implementation is cumbersome when the null is a nonlinear model.

[10]This problem is documented, in a related context, by Ramalho, Ramalho and Murteira, (2010).

$\hat{\lambda}$ rather than at λ, and consequently has an asymptotic variance smaller than one. Therefore, a test based on this statistic is likely to be severely undersized and will lack power. Davidson and MacKinnon (1981) provide a valid estimator of the variance of the estimate of α with which it is possible to construct an asymptotically valid test, but that procedure is likely to be too cumbersome for most practitioners to use it routinely.

The C test is the only approach that is feasible when $f_A(x_i, \lambda)$ and $f_B(x_i, \mu)$ have the same functional form. However, when this is not the case, a more attractive procedure is available. In the remainder of this section we use the results of Davidson and MacKinnon (1981) to develop a new test that is specifically designed to deal with the situation where $f_B(x_i, \mu)$ is a special case of, but not identical to, $f_A(x_i, \lambda)$ like, for example, when the Heckit in logs is tested against the ECE. The test can also be used when $f_A(x_i, \lambda)$ is a special case of, but not identical to, $f_B(x_i, \mu)$, being closely related to the P test in this case. The new test thus builds on the C and P tests, and it explicitly takes into consideration that this type of data is typically characterized by strong heteroskedasticity.

3.2. The proposed test

In what follows, we will focus on models that specify conditional expectations of the form $\exp\left(x_i'\beta_j^* + z_i^j\right)$, where $z_i^j = \ln\Phi\left(x_i'\gamma_j^*\right)$.[11] Suppose that we want to test Model j, $j \in \{1, 2, 3, 6\}$, implying

$$M_j : \mathrm{E}[y_i|x_i] = \exp\left(x_i'\beta_j^* + z_i^j\right),$$

against Model k, $k \in \{1, 2, 3, 6\}$ and $k \neq j$, which leads to

$$M_k : \mathrm{E}[y_i|x_i] = \exp\left(x_i'\beta_k^* + z_i^k\right).$$

[11] It is trivial to develop an analogous procedure for more general specifications of the conditional mean, but that is not pursued here.

Although the conditional expectations implied by both models have the same functional form, the different models imply that these functions are evaluated at different sets of parameters and in that sense they are non-nested.

As in Davidson and MacKinnon (1981), we start by nesting the competing specifications in a compound model of the form

$$M_l : \mathrm{E}\left[y_i|x_i\right] = (1-\alpha)\exp\left(x_i'\beta_j^* + z_i^j\right) + \alpha\exp\left(x_i'\beta_k^* + z_i^k\right),$$

and want to check the correct specification of M_j by testing $H_0 : \alpha = 0$ against $H_1 : \alpha = 1$. As before, α is not generally identified and therefore the test has to be performed by conditioning on parameter estimates. In particular, as in the C test, we propose testing $H_0 : \alpha = 0$ versus $H_1 : \alpha = 1$ conditioning on estimates obtained both under the alternative and under the null. However, unlike in the C test, we will not condition on all estimates under the null, but only on the estimates of γ_j^*, the parameters determining the probability of observing $y_i = 0$. That is, β_j^* is allowed to be freely estimated, and the hypothesis of interest is tested in the artificial regression

$$y_i = (1-\alpha)\exp\left(x_i'\beta^* + \hat{z}_i^j\right) + \alpha\left(\exp\left(x_i'\hat{\beta}_k^* + \hat{z}_i^k\right)\right) + \xi_i, \tag{1}$$

where $\hat{z}_i^a = \ln\Phi\left(x_i'\hat{\gamma}_a^*\right)$ and $\hat{\beta}_a^*$ and $\hat{\gamma}_a$ denote estimates obtained under model $a \in \{j, k\}$.

Like in the P test, estimation of (1) can be avoided by linearizing the model around $\beta^* = \hat{\beta}_j^*$. Moreover, in the specific context we have in mind, the variance of ξ_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 $\mathrm{Var}\left[y_i|x_i\right] \propto \exp\left(x_i'\beta_j^* + z_i^j\right)$. This modification is not only likely to improve the performance of the test in finite samples, but it also has a second interesting consequence. Indeed, the test based on the weighted regression checks a moment condition of the form

$$\mathrm{E}\left[\left(y_i - \exp\left(x_i'\beta_j^* + z_i^j\right)\right)\frac{\exp\left(x_i'\beta_k^* + z_i^k\right) - \exp\left(x_i'\beta_j^* + z_i^j\right)}{\exp\left(x_i'\beta_j^* + z_i^j\right)}\right] = 0.$$

That is, like the tests proposed by Davidson and MacKinnon (1981), the test proposed here 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, a percentage difference between the two conditional means is used as a weight. The use of a percentage difference is appropriate and attractive 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, which are critical in distinguishing between one- and two-equation models, would be essentially ignored if the weights were just the difference between the two sets of fitted values, as in the P test, and not a percentage difference, as in the proposed test.

Implementing the test in practice is very simple. Defining $\hat{y}_i^a = \exp\left(x_i'\hat{\beta}_a^* + \hat{z}_i^a\right)$ and $\delta = \hat{\beta}_k^* - \beta$, the proposed test is just a t-test for $H_0 : \alpha = 0$ against $H_1 : \alpha = 1$ in the OLS estimation of an artificial model of the form

$$\frac{y_i - \hat{y}_i^j}{\sqrt{\hat{y}_i^j}} = \sqrt{\hat{y}_i^j} x_i'\delta + \alpha \frac{\hat{y}_i^k - \hat{y}_i^j}{\sqrt{\hat{y}_i^j}} + \nu_i, \qquad (2)$$

which can be conveniently performed, for example, in Stata (StataCorp., 2009), as a least squares regression of $\left(y_i - \hat{y}_i^j\right) / \hat{y}_i^j$ on x and $\left(\hat{y}_i^k - \hat{y}_i^j\right) / \hat{y}_i^j$, using $\hat{y}_i^j$ as weights.

Like in the C test (Davidson and MacKinnon, 1981), the asymptotic variance of the usual t-statistic for the significance of α in (2) is not equal to 1 because the test does not take into consideration that $\hat{z}_i^j$ is evaluated at estimates under the null. Indeed, as it is shown in the Appendix (see also Davidson and MacKinnon, 1981, and Pierce, 1982), when $\hat{z}_i^j$ is is evaluated at the maximum likelihood estimates of γ_j^*, the standard t-test for $H_0 : \alpha = 0$ in (2) will be asymptotically undersized. Still, we expect this problem to be much less severe than in the C test because the proposed procedure does not condition on $\hat{\beta}_j^*$. Naturally, an asymptotically correct estimator of the variance of the estimate of α can be obtained using the misspecification-robust version of the methods presented in Davidson and MacKinnon (1981), as detailed in the Appendix. However, in the next section we present some Monte Carlo evidence which suggests that the test suffers only from small size distortions even when it is based on the uncorrected estimate of the variance. Therefore, we conjecture that, for most empirical applications, the additional computational burden of correcting the covariance estimator may not be justified.

It is noteworthy that, when the null is the ECE model ($j = 1$), the variance of the estimate of α does not need to be corrected because $\hat{z}_i^j = 0$ and therefore, under the null, $\hat{\gamma}_j^*$ vanishes from the model. In this case, the test based on (2) is just an heteroskedasticity adjusted P test, and consequently the standard t-statistic will asymptotically have the correct size.

Finally, it is important to mention that, as is standard with tests for non-nested hypotheses, the roles of the null and alternative can be reversed. This leads to three possible outcomes of the proposed 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 proposed test has the ability to reject both specifications when neither is appropriate. On the other hand, if two specifications are very close and the sample is not rich enough, the test may be unable to discriminate between the two competitors.

4. SIMULATION RESULTS

In this section we report the results of a small scale Monte Carlo study evaluating the performance of the proposed test. More specifically, two sets of simulations were performed. In the first set, we focus on the tests comparing the Heckit in logs, estimated by maximum likelihood, with the ECE model, estimated by Poisson pseudo-maximum likelihood (PPML). In the second set of simulations the models being compared are the ECE, again estimated by PPML, and the M-2PM for which the first part is a probit and the second part is an ECE estimated by PPML. This choice of models to include in the simulations 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 Heckit in logs or the M-2PM is the corect model, data are generated as

$$\Pr\left(y_i > 0|x_i\right) = \Pr\left(0.2x_{i1} + 0.8x_{i2} + e_i > 0|x_i\right),$$
$$\text{for } y_i > 0 : \ln\left(y_i\right) = 1 + 0.8x_{i1} + 0.0x_{i2} + u_i,$$
$$\begin{bmatrix} e_i \\ u_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}\right).$$

In the first set of experiments, when the null is the Heckit, we set $\rho = 0.5$ and $\sigma = 0.5$. For the second set of experiments, when the null is the M-2PM, $\rho = 0$ and $\sigma = 0.5$.

When the ECE model is used to generate the data, we follow Santos Silva and Tenreyro (2009) and obtain y_i as a χ^2 random variable with η_i degrees of freedom, where the η_i are draws from a negative-binomial distribution with

$$\mathrm{E}\left[\eta_i|x_i\right] = \exp\left(0.55 + 0.88x_{i1} + 0.2x_{i2}\right),$$
$$\mathrm{Var}\left[\eta_i|x_i\right] = 5\mathrm{E}\left[\eta_i|x_i\right].$$

In all data generation processes, x_1 is obtained as a random draw from the standard normal distribution, and x_2 is a dummy variable with $\Pr\left(x_2 = 1\right) = 0.4$. New sets of regressors are drawn for each Monte Carlo replication. In the Heckit and M-2MP designs, y_i is equal to zero for about 40 percent of the observations, whereas in the ECE design this percentage is close to 45 percent.

All tests are one-tailed and are based on the t-statistic for $H_0 : \alpha = 0$ versus $H_1 : \alpha = 1$ in (2). When the Heckit in logs or the M-2PM is the null, the test is performed both with and without the correction to the covariance matrix. Obviously, when the ECE is the null, the correction is not needed.

The simulation results were obtained for samples of sizes 1000, 2000, and 5000 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 Heckit in logs and the ECE are the competing models. Table 3 presents the corresponding results for the cases in which the M-2PM and the ECE are considered.

TABLES 2 & 3 ABOUT HERE

The results in Tables 2 and 3 suggest that, with these designs, the asymptotically valid tests are generally mildly oversized, whereas the uncorrected test is slightly undersized. Overall, under the null, the tests display reasonable behaviour, even for samples which are much smaller than those currently used in most empirical applications. When the null is false, the tests have reasonable power and, naturally, the power increases with the sample size.

The behaviour of the tests under the alternative deserves a couple of additional comments. First, we notice that, as expected, the uncorrected test is less powerful than the version based on the corrected covariance matrix. However, this loss of power is relatively small, and it vanishes reasonably quickly when the sample size increases. Therefore, at least with reasonably sized samples, the additional cost incurred in computing the corrected test-statistic may not be justified. The second interesting point to note is that the tests are substantially less powerful when the null is the ECE than when the null is either the Heckit in logs or the M-2PM. This issue deserves further exploration, but we conjecture that this difference in power results from the fact that the ECE 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 Table 1, the ECE is the only one that directly estimates $\mathrm{E}\left[y_i|x_i\right]$, 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 is flexible enough to approximate the $\mathrm{E}\left[y_i|x_i\right]$ implied by the Heckit in logs or by the M-2PM. A more complete study of the ability of the ECE model to approximate the functional form of the conditional mean of other models is, however, beyond the scope of the present paper.

5. EMPIRICAL ILLUSTRATIONS

In this section we illustrate the performance of the proposed test in applications using two well-known data sets, one in international trade and the other in the demand for health care. In both cases, the ECE model, estimated by PPML, is tested against a two-equation model and vice-versa. In view of the simulation results presented in the previous section, when the null is the two-equation model, only the results of the test based on the uncorrected covariance matrix are presented because this is the version of the test that is more likely to be used in practice.

5.1. 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 using a variety of methods. 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).[12] 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).

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

Table 4 presents the main estimation results obtained with the Heckit in logs (estimated by maximum likelihood) and with the ECE model (estimated by PPML), both with and without the multilateral resistance terms suggested by Anderson and van Wincoop (2003).[13] 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 test of the sample selection estimator against the ECE model and vice-versa.

Comparing the R^2s for the competing models, it is possible to see that for both specifications the ECE model fits the data substantially better than the Heckit. However, goodness-of-fit statistics give no indication about the adequacy of the models being contrasted and therefore it is interesting to use the proposed procedure to test the two models against each other. The results in the last row of Table 4 show that, either with or without the multilateral resistance terms, the proposed test clearly rejects the Heckit specification, while providing no evidence of departures of the ECE model, estimated by PPML, in the direction of its competitor. Naturally, this result is specific to this particular example and it should therefore not be viewed as indicating that the ECE model is generally preferable to the Heckit in applications describing bilateral export flows.

TABLE 4 ABOUT HERE

5.2. Much ado about two redux

In a landmark paper, Mullahy (1998) studied the choice between one- and two-equation 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

[13] The multilateral resistance terms are importer and exporter fixed effects.

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.[14] In this particular application, the R^2s of the models are all but the same, which may suggest that there is little to choose between the two models. However, the results of the proposed test provide no evidence against the M-2PM, while clearly rejecting the ECE model estimated by PPML. These results are in line with those of Mullahy (1998) who, using a number specification tests and goodness-of-fit criteria, also finds that the M-2PM specification is preferable to the ECE model in this particular data set. Again, we emphasize that this result 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 ABOUT HERE

6. 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 for non-nested hypotheses and propose an easily implementable regression-based test which is particularly suited to discriminate between one- and two-equation models. Moreover, the proposed test explicitly takes into account the heteroskedasticity that is likely to be present in data of this type,

[14] Both the second part of the M-2PM and the ECE are estimated by PPML. Notice that the estimates reported here for these two Poisson regressions do not match exactly those reported by Mullahy (1998). This is possibly due to more sophisticated algorithm now available for the estimation of this type of models. Notice also that the first equation is estimated using a probit, not a logit as in Mullahy (1998). The results hardly change if a logit is used in the first stage.

and has an intuitive interpretation in terms of orthogonality conditions. We present the results of a small-scale 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. Two illustrative applications show that the test can be quite useful in practice.

The test proposed here can also easily be adapted to other contexts where the researcher wants to choose between one- and two-equation models. Examples include the choice between count data models and their zero-inflated counterparts (Mullahy, 1986), or models for fractional data with mass-points at either bound (Ramalho, Ramalho and Murteira, 2010).

APPENDIX: ADJUSTED COVARIANCE MATRIX

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

$$\frac{y_i - \hat{y}_i^j}{\sqrt{\hat{y}_i^j}} = \sqrt{\hat{y}_i^j} x_i' \delta + \alpha \frac{\hat{y}_i^k - \hat{y}_i^j}{\sqrt{\hat{y}_i^j}} + u_i.$$

The easiest way of obtaining an asymptotically valid covariance matrix for the OLS estimates of $\theta = (\delta, \alpha)$, say $\hat{\theta} = \left(\hat{\delta}, \hat{\alpha}\right)$, is to consider the joint estimation of θ and $\phi_j = \left(\beta_j^*, \gamma_j^*\right)$ by system GMM (see, Newey, 1984).[15] Let S_1 and S_2 denote the vector of moment conditions for the model under the null and for the test equation, respectively. Moreover, let H denote the expectation of the matrix of derivatives of S_1 with respect to ϕ_j and H_1 and H_2 denote the expectation of the derivatives of S_2 with respect to ϕ_j and θ, respectively. Then, the covariance matrix of the vector $\left(\hat{\beta}_j^*, \hat{\gamma}_j^*, \hat{\delta}, \hat{\alpha}\right)$ in the just-identified system-GMM estimator is given by $M^{-1} \Sigma M^{-1\prime}$,

[15] The results in this appendix are presented for the case in which β_j^* and γ_j^* are jointly estimated by maximum likelihood, like in the Heckit. For models such as the 2-PM in which γ_j^* can be estimated independently of β_j^*, the same results are valid if one considers only the moment conditions for the joint estimation of γ_j^* and θ.

with

$$M = \begin{bmatrix} H & 0 \\ H_1 & H_2 \end{bmatrix}, \qquad \Sigma = \mathrm{E}\begin{bmatrix} S_1 S_1' & S_1 S_2' \\ S_2 S_1' & S_2 S_2' \end{bmatrix}.$$

Noting that

$$M^{-1} = \begin{bmatrix} H^{-1} & 0 \\ -H_2^{-1} H_1 H^{-1} & H_2^{-1} \end{bmatrix},$$

the variance of $\hat{\theta}$ can then be written as

$$\begin{aligned} \mathrm{V}\left(\hat{\theta}\right) &= H_2^{-1}\mathrm{E}\left(S_2 S_2'\right) H_2^{-1\prime} - H_2^{-1}\mathrm{E}\left(S_2 S_1'\right) H^{-1\prime} H_1' H_2^{-1\prime} \\ & \quad -H_2^{-1} H_1 H^{-1}\mathrm{E}\left(S_1 S_2'\right) H_2^{-1\prime} + H_2^{-1} H_1 H^{-1}\mathrm{E}\left(S_1 S_1'\right) H^{-1\prime} H_1' H_2^{-1\prime}, \end{aligned}$$

or

$$\mathrm{V}\left(\hat{\theta}\right) = \mathrm{V}_{\hat{\theta}} + H_2^{-1}\left\{H_1 \mathrm{V}\left(\hat{\phi}_j\right) H_1' - \mathrm{E}\left(S_2 S_1'\right) H^{-1\prime} H_1' - H_1 H^{-1}\mathrm{E}\left(S_1 S_2'\right)\right\} H_2^{-1\prime},$$

where $\mathrm{V}\left(\hat{\phi}_j\right)$ is the estimated variance of $\hat{\phi}_j = \left(\hat{\beta}_j^*, \hat{\gamma}_j^*\right)$ 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.[16] In the context of the proposed test, it is of special interest to consider the case where the two-equation model is estimated by maximum likelihood. In this case, $\mathrm{V}\left(\hat{\phi}_j\right) = H^{-1}$ and $\mathrm{E}\left(S_2 S_1'\right) = H_1$, and therefore

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

implying that $\mathrm{V}\left(\hat{\theta}\right)$ is smaller than $\mathrm{V}_{\hat{\theta}}$ (see Pierce, 1982). Therefore, when the two-equation model 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.

[16] 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.

Table 1: Models for corner solutions data

	Model	Specification	$E[y\|x]$
1	ECE	$E[y_i\|x_i] = \exp(x_i'\beta_1)$	$\exp(x_i'\beta_1)$
2	2PM	$\Pr(y_i > 0\|x_i) = \Phi(x_i'\gamma_2)$ for $y_i > 0 : \ln(y_i) = x_i'\beta_2 + e_i$ $e_i \sim \mathcal{N}(0, \sigma_2^2)$	$\exp\left(x_i'\beta_2 + \frac{\sigma_2^2}{2} + \ln\Phi(x_i'\gamma_2)\right)$
3	M-2PM	$\Pr(y_i > 0\|x_i) = \Phi(x_i'\gamma_3)$ $E[y_i\|x_i, y_i > 0] = \exp(x_i'\beta_3)$	$\exp(x_i'\beta_3 + \ln\Phi(x_i'\gamma_3))$
4	Tobit	$y_i = \max\{0, x_i'\beta_4 + e_i\}$ $e_i \sim \mathcal{N}(0, \sigma_4^2)$	$x_i'\beta_4\Phi(x_i'\beta_4/\sigma_4) + \sigma_4\phi(x_i'\beta_4/\sigma_4)$
5	Heckit	$\Pr(y_i > 0\|x_i) = \Pr(x_i'\gamma_5 + e_{1i} > 0\|x_i)$ for $y_i > 0 : y_i = x_i'\beta_5 + e_{2i}$ $\begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_5\sigma_5 \\ \rho_5\sigma_5 & \sigma_5^2 \end{bmatrix}\right)$	$x_i'\beta_5\Phi(x_i'\gamma_5) + \rho_5\sigma_5\phi(x_i'\gamma_5)$
6	Heckit (in logs)	$\Pr(y_i > 0\|x_i) = \Pr(x_i'\gamma_6 + e_{1i} > 0\|x_i)$ for $y_i > 0 : \ln(y_i) = x_i'\beta_6 + e_{2i}$ $\begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_6\sigma_6 \\ \rho_6\sigma_6 & \sigma_6^2 \end{bmatrix}\right)$	$\exp\left(x_i'\beta_6 + \frac{\sigma_6^2}{2} + \ln\Phi(x_i'\gamma_6 + \rho_6\sigma_6)\right)$

Table 2: Rejection frequencies at the 5% nominal level

	Null is true			Null is false		
$n =$	1000	2000	5000	1000	2000	5000
The null is the Heckit	0.0661	0.0614	0.0594	0.9514	0.9987	1.0000
(Uncorrected test)	0.0510	0.0484	0.0509	0.9313	0.9970	1.0000
The null is the ECE	0.0855	0.0717	0.0702	0.2723	0.3771	0.6421

Table 3: Rejection frequencies at the 5% nominal level

	Null is true			Null is false		
$n =$	1000	2000	5000	1000	2000	5000
The null is the M-2PM	0.0634	0.0578	0.0539	0.9588	0.9973	0.9992
(Uncorrected test)	0.0444	0.0414	0.0435	0.9324	0.9952	0.9990
The null is the ECE	0.0789	0.0685	0.0564	0.3046	0.4183	0.7031

Table 4: Gravity equations for trade

Estimator:	Heckit 1st part	Heckit 2nd part	ECE	Heckit 1st part	Heckit 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		18360	18360		18360
R^2	0.580		0.862	0.391		0.928
Nonnested test p-values	0.000		0.999	0.029		1.000

Table 5: Demand for health care

Estimator:	M-2PM 1st part	M-2PM 2nd part	ECE
Age	0.004	−0.008	−0.006
	(0.001)	(0.000)	(0.000)
Male	−0.535	−0.106	−0.299
	(0.015)	(0.005)	(0.005)
White	0.090	0.151	0.185
	(0.020)	(0.006)	(0.006)
Schooling	0.061	0.031	0.051
	(0.003)	(0.001)	(0.001)
Married	0.070	−0.136	−0.111
	(0.016)	(0.005)	(0.005)
Excellent	−0.792	−1.575	−1.817
	(0.030)	(0.008)	(0.007)
Very Good	−0.592	−1.311	−1.476
	(0.030)	(0.007)	(0.007)
Good	−0.500	−0.847	−0.983
	(0.030)	(0.006)	(0.006)
Observations	36111	27598	36111
R^2	0.078		0.077
Nonnested test p-values	0.752		0.001

REFERENCES

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

Anderson J, 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.

Basu, A., Arondekar, B.V., and Rathouz, P.J. (2006). "Scale of interest versus scale of estimation: comparing alternative estimators for the incremental costs of a comorbidity," *Health Economics*, 15, 1091-1107.

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.

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.

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

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.

Hay, J.W. and Olsen, R.J. (1984). "Let them eat cake: A note on comparing alternative models of the demand for health care," *Journal of Business and Economic Statistics* 2, 279–282.

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. & 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.

Manning, W.G., Basu A., and Mullahy J. (2005). "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data," *Journal of Health Economics* 24, 465-488.

Manning,W.G., Duan, N., and Rogers,W.H. (1987). "Monte Carlo evidence on the choice between sample selection and two-part models," *Journal of Econometrics* 35, 59–82.

Manning, W.G. and Mullahy, J. (2001). "Estimating Log Models: To Transform or Not to Transform?," *Journal of Health Economics* 20, 461-494.

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.

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

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. (2010). "Alternative estimating and testing empirical strategies for fractional regression models," *Journal of Economic Surveys*, forthcoming.

Ramsey, J.B. (1969). "Tests for specification errors in classical linear least squares regression analysis," *Journal of the Royal Statistical Society B*, 31, 350-371.

Ramsey, J.B. and Schmidt, P. (1976). "Some further results on the use of OLS and BLUS residuals in specification error tests," *Journal of the American Statistical Association*, 71, 389–390.

Rose, A.K. (2004). "Do we really know that the WTO increases trade?," *American Economic Review*, 94, 98-114.

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. (2009). *Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator*, Department of Economics, University of Essex, Discussion Paper No. 666.

StataCorp. (2009). *Stata Release 11. Statistical Software.* College Station (TX): StataCorp LP.

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.

Wooldridge, J.M. (1992). "Some Alternatives to the Box-Cox Regression Model," *International Economic Review*, 33, 935-955.

Wooldridge, J.M. (2002). *Econometric analysis of cross section and panel data*, Cambridge, MA: MIT Press.