

de Luna, Xavier; Johansson, Per

Working Paper

Testing exogeneity under distributional misspecification

Working Paper, No. 2001:9

Provided in Cooperation with:

IFAU - Institute for Evaluation of Labour Market and Education Policy, Uppsala

Suggested Citation: de Luna, Xavier; Johansson, Per (2001) : Testing exogeneity under distributional misspecification, Working Paper, No. 2001:9, Institute for Labour Market Policy Evaluation (IFAU), Uppsala

This Version is available at:

<https://hdl.handle.net/10419/82155>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



IFAU - OFFICE OF LABOUR
MARKET POLICY EVALUATION

Testing exogeneity under distributional misspecification

Xavier de Luna, Per Johansson

WORKING PAPER 2001:9

Testing exogeneity under distributional misspecification*

Xavier de Luna[†] and Per Johansson[‡]

18th July 2001

Abstract

We propose a general test for exogeneity that is robust against distributional misspecification. The test can also be used to identify other types of misspecifications, such as the presence of a random coefficient. The idea is to sort the data with respect to a variable (a sorting score) and then split the sample into two parts. Using a Chow test, it can then be tested whether estimated parameters in the two sub-samples are different. We give conditions under which it is possible to test for exogeneity by using the (supposedly) endogenous variable as a sorting score. The resulting test does not need instrumental variables. Evidence from a Monte Carlo study and an empirical application suggests that the test can be useful for practitioners.

Keywords: Absenteeism at Work; Endogeneity; Linear exponential family; Random effect; Random coefficient; Selectivity.

JEL: C12, C14, C15, C21, C31, C52.

*We are grateful to Joshua Angrist, Kurt Brännäs, Sune Karlsson, and participants at seminars given at the Dept. of Economic Statistics, Stockholm School of Economics, the Dept. of Statistics, Uppsala University, and the Dept. of Economics, Umeå University, for helpful comments on previous drafts of this paper.

[†]Department of Economics, Umeå University, 901 87 Umeå, Sweden. E-mail: xavier@econ.umu.se.

[‡]IFAU - Office of Labour Market Policy Evaluation, 751 20 Uppsala, Sweden. Email: Per.Johansson@ifau.uu.se.

1 Introduction

1.1 Background and purpose

A challenging issue in observational studies is to assess whether causal conclusions may be drawn from a given analysis. The concept of exogeneity is closely connected to this issue, see, e.g., Koopmans (1950) and Imbens (1997).¹

Engle, Hendry and Richard (1983), building on Koopmans' work, were probably the first in providing a rigorous definition of exogeneity within an inferential framework: In regression models for cross-sectional data where a dependent variable y is explained, an explanatory variable z was defined as being exogenous² for a parameter vector of interest if this parameter could be efficiently estimated by basing the likelihood on the conditional density of y given z , thereby ignoring the marginal density of z . Observe that this definition rest on a well-specified model and is, hence, not meaningful under distributional misspecification.

In particular situations, methods have been developed avoiding the specification of econometric regression models. For instance, if the issue of interest is to assess the effect of a treatment/program – $z = 1$ (treatment group) $z = 0$ (control group) – on a given socio-economic variable y of interest (e.g. the effect of a labor program on wages), then instead of considering a regression model, an experimental study may be mimic to a certain point, by using matching methods (see, e.g., Rosenbaum and Rubin, 1983; Rosenbaum, 1995; Heckman, Ichimura and Todd, 1998). Here exogeneity is replaced with a concept of ignorability of the treatment assignment.

However, for most economic empirical applications, regression models remain the most useful tool in performing inference. In this paper we propose a general test for exogeneity of a regressor under distributional misspecification by using the framework of quasi maximum likelihood (QML) inference developed in White (1982).³ To achieve this, a definition

¹See also Holland (1986, with discussion) for an excellent overview on different definitions of causality, and their implications on causal inference. An early reference on causal inference is Stroz and Wold (1960).

²Note that in the cross-sectional context of this paper the concepts of weak and strong exogeneity are equivalent.

³For a related use of the QML framework see Wooldridge (1991), who suggests extensions to conditional moment tests under distributional misspecification.

of exogeneity is proposed which is operational under distributional misspecification.

1.2 Testing exogeneity

In order to propose a test for model misspecification of a regression model for cross-sectional data we define a Kullback-Leibner (KL) (Kullback and Leibner, 1951) invariant parameter for an explanatory variable, z , in a parametric regression model. Loosely, a parameter vector of interest is said to be KL -invariant for z , if it is invariant when conditioning the inference on any value for this variable. To test the hypothesis that a parameter is KL -invariant for z , the available sample is sorted with respect to the variable z and divided into two parts. Using a quasi maximum likelihood estimator (QMLE) the parameter is then estimated on the two sub-samples. Given the conditions in White (1982), we can test whether these two estimated parameters are equal. A Chow test statistic is used for that purpose.⁴ We show that this Chow test is a test for the KL -invariance of the parameter of interest. This test is consistent and has power against non-invariance. Furthermore, the test is simple to carry out with any software allowing for maximum likelihood estimation.

Given the definition of KL -invariance, we define exogeneity under distributional misspecification. A variable is said to be exogenous for a parameter when this parameter is KL -invariant for any existing variable (observed or unobserved, known or unknown) which happens to be dependent on the former variable. This definition of exogeneity is conceptually different from the definition of Engle et al. (1983), although in usual econometric models of endogeneity (non-exogeneity, see, e.g., Vella, 1992) the two definitions agree. The concept of exogeneity proposed herein has, however, the advantage of allowing for distributional misspecification and is, we believe, closer to a notion of causality.

The Chow test described above becomes a test for exogeneity in situations where we know that the non-invariance of a parameter for a variable can only be due to the endogeneity of this variable. In this respect, we show that KL -invariance is guaranteed under exogeneity when the assumed parametric model belongs to the linear exponential family (see

⁴For the particular case of linear models, such sorting was used in de Luna and Johansson (2001) to give graphical diagnostics of endogeneity based on recursive residuals. Here, the Chow test is a more general tool.

Gouriéroux, Monfort and Trognon, 1984), and this model is possibly misspecified except for the form of the conditional expectation.

In contrast to the Hausman test (e.g., White, 1982), the introduced Chow test does not need a fully well-specified model under the alternative hypothesis. The latter also usually needs instruments while an instrument-free Chow test is available if we can directly sort with respect to the suspected endogenous variable. However, such sorting is not always implementable and, for instance, if the suspected endogenous variable is a dummy variable, then instruments are needed to sort the data and hence to perform the test.

Finally, other types of misspecification leading to non-invariance may also be tested, such as the presence of unobserved heterogeneity in the form of a random coefficient. As usual with misspecification tests, which misspecification leads to the test having power is a delicate question and the answer should be based on available theoretical grounds. Nevertheless, a general strategy for using the Chow test in practical situations is suggested in Section 6.

1.3 Organization of the paper

In Section 2, the framework is described; in particular we define *KL*-invariance and exogeneity. In Section 3, the Chow test is described and its consistency shown. Its power is also discussed, for instance in relation to well-known regression specifications where endogeneity (or selectivity) is modelled. Based on these models, Section 4 studies the small sample performance through a Monte Carlo study. Section 5 gives an empirical illustration using data on the number of days blue collar workers in Sweden are absent from work, see Johansson and Palme (1996). In their paper, the primary interest was to study the economic incentives on absenteeism. The cost of being absent is expected to be endogenous, however. We sort the data on the cost variable and form two data sets of equal size; one with low costs and one with high costs of being absent. Binomial regression models are estimated separately (using a QMLE) on the two samples and, on basis of the Chow test, the exogeneity of the cost variable is rejected.

2 Kullback-Leibner invariance and exogeneity

Let y_i and z_i be two variables observed on individual i .⁵ It is assumed that (y_i, z_i) , $i = 1, 2, \dots$, are independently and identically distributed random variables. We denote the joint, conditional and marginal densities $g(y_i, z_i)$, $g(y_i|z_i)$, $g(z_i|y_i)$, $g(y_i)$, $g(z_i)$, etc. Note that we use small letters, e.g. y_i , both to denote a random variable and its observed realization.

For scientific purposes (e.g., explaining an outcome y_i of a treatment z_i), a parametrized regression model with density $f(y_i|z_i; \theta)$, for y_i and z_i is specified up to the parameter vector θ taking value in a given space. Note that we do not assume that $f(y_i|z_i; \theta) = g(y_i|z_i)$ for one value of θ , i.e. the model is possibly misspecified. This parameter is estimated using a quasi maximum likelihood estimator (QMLE), i.e.,

$$\frac{1}{n} \sum_{i=1}^n \log f(y_i|z_i; \theta) \quad (1)$$

is maximized with respect to θ , for the random sample, $(y_1, z_1), \dots, (y_n, z_n)$, of size n . We denote this QMLE $\hat{\theta}$. The theory associated with quasi likelihood inference was presented in White (1982), whose results are now briefly summarized.

Under assumptions A1 – A3 in White (1982), such an estimator exists and, as $n \rightarrow \infty$,

$$\hat{\theta} \xrightarrow{a.s.} \theta^*,$$

where *a.s.* stands for almost sure convergence. Here, θ^* is the parameter vector minimizing the Kullback and Leibner (1951) Information Criterion (KLIC), i.e.,

$$\theta^* = \arg \min_{\theta} I(g : f, \theta), \quad (2)$$

where⁶

$$I(g : f, \theta) = E_{y,z} \left(\log \left[\frac{g(y_i|z_i)}{f(y_i|z_i; \theta)} \right] \right). \quad (3)$$

⁵In order to enhance the presentation, we neglect observed exogenous variables \mathbf{x}_i .

⁶It is equivalent to compare $f(y|z; \theta)$ to $g(y, z)$ or $g(y|z)$, because the former can be factorized as $g(y, z) = g(y|z)g(z)$, and $g(z)$ does not depend on the parameter on which the optimization takes place.

The expectation operator $E_{y,z}(\cdot)$ is defined with respect to the true distribution of the subscript variables, here $g(y_i, z_i)$. Further, under assumptions A1 – A6 in White (1982), as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N(0, \mathbf{V}(\boldsymbol{\theta}^*)),$$

and $\mathbf{V}(\hat{\boldsymbol{\theta}}) \xrightarrow{a.s.} \mathbf{V}(\boldsymbol{\theta}^*)$, where d stands for convergence in distribution. Here, $\mathbf{V}(\boldsymbol{\theta}^*) = \mathbf{J}^{-1} \mathbf{I} \mathbf{J}^{-1}$ and $\mathbf{V}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{J}}^{-1} \hat{\mathbf{I}} \hat{\mathbf{J}}^{-1}$, where

$$\mathbf{J} = E_{y,z} \left(-\frac{\partial^2 \log f(y_i|z_i; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right),$$

$$\mathbf{I} = E_{y,z} \left(\frac{\partial \log f(y_i|z_i; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \left(\frac{\partial \log f(y_i|z_i; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \right)' \right)$$

and

$$\hat{\mathbf{J}} = \frac{1}{n} \sum_{i=1}^n -\frac{\partial^2 \log f(y_i|z_i; \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'},$$

$$\hat{\mathbf{I}} = \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \log f(y_i|z_i; \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \left(\frac{\partial \log f(y_i|z_i; \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \right)' \right].$$

Hence, even if the model for y_i is misspecified (i.e. $f(y_i|z_i; \boldsymbol{\theta})$ is different from $g(y_i|z_i)$ for any $\boldsymbol{\theta}$), the QML estimator is strongly consistent (with respect to $\boldsymbol{\theta}^*$), asymptotic normal and has an estimable covariance matrix.

However, misspecification must be restricted for the fitted model $f(y_i|z_i; \hat{\boldsymbol{\theta}})$ to have scientific content (other than predictive). Thus, this leads us to consider situations where the estimated parameter is interpretable.

Definition 2.1 *Let*

$$t_z(z_i) = \arg \min_{\boldsymbol{\theta}} I(g : f, \boldsymbol{\theta} | z_i),$$

where

$$I(g : f, \boldsymbol{\theta} | z_i) = E_{y|z} \left(\log \frac{g(y_i|z_i)}{f(y_i|z_i; \boldsymbol{\theta})} \right) \Big|_{z_i} \quad (4)$$

is the KLIC conditional on z_i . The parameter $\boldsymbol{\psi} = \phi(\boldsymbol{\theta})$, a function of $\boldsymbol{\theta}$ only, is said to be KL-invariant for z_i if $\phi(t_z(z_i)) = \boldsymbol{\psi}^*$, a constant, for any value of z_i .

Remark 2.1 A parameter which is not a function of θ only is not *KL*-invariant for z_i by definition. When $\phi(\theta) = \theta$, then $\psi^* = \theta^*$. For simplicity, θ will be used to denote both θ and ψ in the sequel, while making clear from the context which is the parameter of interest.

We have chosen to call a parameter with such a property "invariant" because of its invariance when conditioning on the variable of interest, z_i .

A trivial example where the parameter is invariant is when there is no misspecification, i.e. $f(y_i|z_i; \theta) = g(y_i|z_i)$ for a given value of θ . The following definition and result provide another class of situations where *KL*-invariance in z_i is guaranteed as soon as the conditional expectation $E_{y|z}(y_i|z_i)$ is well-specified.

Definition 2.2 The density $f(y_i|z_i; \theta)$ belongs to the linear exponential family (*LEF*) if it can be written in the form

$$\log f(y_i|z_i; \theta) = A(m(z_i, \theta)) + B(y_i) + C(m(z_i, \theta))y_i,$$

where $m(z_i, \theta)$ is the conditional mean associated with $f(y_i|z_i; \theta)$, and $A(m), B(y_i)$ and $C(m)$ are scalar functions.

Proposition 2.1 Assume that $f(y_i|z_i; \theta)$ belongs to the *LEF*. Under regularity assumptions (see Gouriéroux et al., 1984, Appendix 1) and if $m(z_i, \theta)$ is well specified, i.e. $E_{y|z}(y_i|z_i) = m(z_i, \theta_0)$ for θ_0 a particular value of θ ,⁷ we have that θ is *KL*-invariant for z_i .

Proof. The conditional *KLIC* (4) is minimized with respect to θ by maximizing

$$E_{y|z}(\log f(y_i|z_i; \theta)|z_i) = A(m(z_i, \theta)) + E_{y|z}[B(y_i)|z_i] + C(m(z_i, \theta))m_0,$$

where $m_0 = m(z_i, \theta_0)$. However, this is equivalent to maximizing

$$A(m(z_i, \theta)) + C(m(z_i, \theta))m_0$$

since $E_{y|z}[B(y_i)|z_i]$ does not depend on θ . By Property 4 in Gouriéroux et al. (1984), this expression is maximized at $\theta = \theta_0$, for any z_i . Thus, θ is *KL*-invariant for z_i . ■

⁷By Theorem 1 in Gouriéroux et al. (1984) we know that $\theta^* = \theta_0$ in (2).

Example 2.1 *Let us consider the linear model*

$$y_i = \gamma z_i + \nu_i,$$

where ν_i is such that $E_{\nu|z}(\nu_i|z_i) = \rho z_i$, $\rho \neq 0$. Here $\theta = \gamma + \rho$ is the parameter entering $E(y_i|z_i) = m(z_i, \theta)$. Proposition 2.1 says that when using a QMLE for θ based on the assumption that $y_i|z_i$ has a distribution from the LEF (e.g. ordinary least squares, OLS, estimator), then θ is KL-invariant for z_i . On the other hand, γ is not KL-invariant for z_i if $\rho \neq 0$, because then $\gamma = \theta - \rho$ is not a function of θ only, see Remark 2.1.

Remark 2.2 *Note that the result in Proposition 2.1 does not depend directly on the density $g(y_i, z_i)$, except in that $m(z_i, \theta)$ must be well specified. On the other hand, the density $f(y_i|z_i; \theta)$ chosen to obtain a QMLE must be from the LEF. In such cases, if θ is not KL-invariant for z_i then it implies that $E_{y|z}(y_i|z_i)$ is misspecified.*

Exogeneity of a variable has been defined for well-specified models, see Engle et al. (1983) or Gouriéroux and Monfort (1995, Def. 1.8). In our context, if $g(y_i, z_i) = f(y_i|z_i; \theta)f(z_i|\alpha)$ and θ and α vary over a product space, then z_i is said to be exogenous for θ . Mistakenly assuming exogeneity most often leads to invariant misspecification, i.e. θ is not KL-invariant for z_i if z_i is endogenous, see Example 3.1 below. This prompts us to propose a definition of exogeneity valid in situations where the model is possibly misspecified, although invariant.

Definition 2.3 *For a random variable u_i , let*

$$t_u(u_i) = \arg \min_{\theta} I(g : f, \theta | u_i),$$

where

$$I(g : f, \theta | u_i) = E_{y,z|u} \left(\log \frac{g(y_i|z_i)}{f(y_i|z_i; \theta)} \right) \Big| u_i.$$

Then, the variable z_i is said to be exogenous for $\psi = \phi(\theta)$, a function of θ only, if this parameter is KL-invariant for z_i , and for any other random variable u_i (observed or not observed, known or not known) dependent on z_i , we have $\phi(t_u(u_i)) = \phi(t_z(z_i)) = \psi^*$. That is ψ is also KL-invariant for any u_i dependent on z_i .

These two definitions of exogeneity are conceptually different as the two following examples illustrate.

Example 2.2 Consider the model for y_i and z_i :

$$\begin{aligned} y_i &= \delta + \gamma z_i + \varepsilon_i \\ z_i &= \gamma + \eta_i, \end{aligned}$$

where $E_\varepsilon(\varepsilon_i) = E_\eta(\eta_i) = 0$, and ε_i and η_i are independent and jointly normal random variables. Here z_i is not exogenous for γ by the Engle et al. (1983) definition. Indeed, one cannot neglect the second equation without losing information on γ . By Definition 2.3, z_i is exogenous for γ in the first equation, since η_i is the only random variable dependent on z_i , and clearly the maximum likelihood estimator of γ based on $f(y_i|z_i; \delta, \gamma)$, here a normal density, is consistent independently on the values taken by η_i .

The model in this example, although peculiar, shows that the Engle et al. (1983) definition of exogeneity for well specified models is not a special case of Definition 2.3. The former is also concerned with efficiency of estimation, an important but not essential issue.

Example 2.3 Consider again the model in Example 2.1, completed as follows:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \gamma z_i + u_i + \varepsilon_i \tag{5}$$

$$z_i = \mathbf{x}_i^* \boldsymbol{\delta} + \eta_i \tag{5}$$

$$u_i = \lambda \eta_i + \omega_i, \tag{6}$$

where $E(\omega_i|\eta_i) = E(\eta_i|\mathbf{x}_i^*) = E(\varepsilon_i|\mathbf{x}_i, z_i, u_i) = 0$ and \mathbf{x}_i and \mathbf{x}_i^* are vectors of exogenous variables. This is a standard model of endogeneity for z_i . Assuming further that all variables in the model are jointly normal, we can write

$$E_{y|\mathbf{x}, z}(y_i|\mathbf{x}_i, z_i) = \mathbf{x}_i' \boldsymbol{\beta} + (\gamma + \rho) z_i,$$

where ρ is a function of λ and $\boldsymbol{\delta}$, and the conditional density $f(y_i|\mathbf{x}_i, z_i; \boldsymbol{\theta})$ is a normal density with location parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}', \theta_2)'$, where $\theta_2 = \gamma + \rho$. The interest in such a model lies typically in γ . However, the latter is a

function of $\boldsymbol{\delta}$, since $\gamma = \theta_2 - \rho$, and therefore is not a function of $\boldsymbol{\theta}$ only. The consequence is that z_i is not exogenous for γ neither in the Engle et al. (1983) sense nor in the sense of Definition 2.3. Let us now assume that the interest lies in parameter θ_2 . Then, z_i is exogenous for θ_2 in the Engle et al. (1983) sense. However, it is not exogenous in the sense of Definition 2.3, although as we saw in Example 2.1, θ_2 is *KL*-invariant for z_i . To see this, consider the sub-population defined by $\eta_i < c$, where c is a given constant within the sample space of η_i . Then

$$E_{y|\mathbf{x},z}(y_i|\mathbf{x}_i, z_i, \eta_i < c) = \mathbf{x}_i' \boldsymbol{\beta} + (\gamma + \rho_c) z_i \quad (7)$$

where $|\rho_c| < |\rho|$. For the sub-population of interest the maximum likelihood estimator is consistent for $\boldsymbol{\theta}_c = (\boldsymbol{\beta}', \gamma + \rho_c)'$. This dependence of ρ_c on η_i , a random variable dependent on z_i , implies that z_i is not exogenous for θ_2 and thereby $\boldsymbol{\theta}$.

This example shows that exogeneity in the Engle et al. (1983) sense does not imply exogeneity as defined in Definition 2.3. The meaning of the latter definition is that although $\boldsymbol{\theta}$ is estimated consistently and efficiently (in particular $\boldsymbol{\theta}$ is *KL*-invariant for z_i), z_i is not exogenous for $\boldsymbol{\theta}$. Of course, $\boldsymbol{\theta}$ is typically not the parameter of interest, and when γ is the focus both definitions imply non-exogeneity of z_i . These two examples are given here to emphasize the conceptual differences between the two exogeneity definitions. These are thus not equivalent in general, but will be in agreement (assuming well specified distributions) in typical econometric models of endogeneity such as those treated in the following sections. An advantage of Definition 2.3 is that it is operational under distributional misspecification. In the sequel we refer to this definition whenever mentioning exogeneity.

Note that when a variable is exogenous for a parameter (in particular, this parameter is *KL*-invariant for the variable), then we may say that the QML estimator converges to a value with a causal interpretation.

3 A Wald-Chow test

Assume that we have a random sample of n individuals for which a scalar variable s_i has been observed. The data set can then be sorted with respect to s (in ascending or descending order for s_i , called a *sorting score*). Based

on this sorting, the sample is separated into two parts, one with $s_i < c$ and one with $s_i > c$, where c is a fixed truncation point. Then, the parameter $\boldsymbol{\theta}$ in $f(y_i|z_i; \boldsymbol{\theta})$ is estimated separately for each sub-sample using a QMLE, yielding $\hat{\boldsymbol{\theta}}_j$ and $\mathbf{V}(\hat{\boldsymbol{\theta}}_j)$, $j = 1$ ($s_i < c$) and $j = 2$ ($s_i > c$). Then, we can use the Wald statistic

$$W = (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)' \left(\mathbf{V}(\hat{\boldsymbol{\theta}}_1) + \mathbf{V}(\hat{\boldsymbol{\theta}}_2) \right)^{-1} (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2) \quad (8)$$

to test H_0 : " $\boldsymbol{\theta}$ is KL -invariant for s_i ." This is justified within the next two sections.

3.1 Size of the test

The proposition below shows that, under H_0 , the above Wald test statistic is asymptotically Chi-squared distributed with p (dimension of $\boldsymbol{\theta}$) degrees of freedom. We call the resulting test a Chow test, in reference to Chow (1960).

Proposition 3.1 *Let A1–A6 in White (1982) hold (also when conditioning with respect to s_i). Then, under H_0 : " $\boldsymbol{\theta}$ is KL -invariant for s_i ", the Wald statistic W is asymptotically $\chi^2(p)$ distributed. Moreover, as $n \rightarrow \infty$,*

$$\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2 \xrightarrow{a.s.} 0 \quad (9)$$

and

$$\mathbf{V}(\hat{\boldsymbol{\theta}}_1) + \mathbf{V}(\hat{\boldsymbol{\theta}}_2) \xrightarrow{a.s.} \mathbf{V}(\boldsymbol{\theta}_1^*) + \mathbf{V}(\boldsymbol{\theta}_2^*). \quad (10)$$

Proof. Let $\boldsymbol{\theta}_1^* = \arg \max_{\boldsymbol{\theta}} E_{y,z}[\log f(y_i|z_i, \boldsymbol{\theta}) | s_i < c]$.⁸ By the invariance of $\boldsymbol{\theta}$ for s_i , we have that

$$\begin{aligned} \boldsymbol{\theta}_1^* &= \arg \max_{\boldsymbol{\theta}} E_{y,z}[\log f(y_i|z_i, \boldsymbol{\theta}) | s_i < c] \\ &= \arg \max_{\boldsymbol{\theta}} E_{y,z}[\log f(y_i|z_i, \boldsymbol{\theta})] = \boldsymbol{\theta}^*. \end{aligned}$$

Then, $\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^*) \xrightarrow{d} N(0, \mathbf{V}(\boldsymbol{\theta}_1^*))$, $\hat{\boldsymbol{\theta}}_1 \xrightarrow{a.s.} \boldsymbol{\theta}^*$ and $\mathbf{V}(\hat{\boldsymbol{\theta}}_1) \xrightarrow{a.s.} \mathbf{V}(\boldsymbol{\theta}_1^*)$ (White's (1982) results summarized above). Note that, although $\boldsymbol{\theta}_1^* = \boldsymbol{\theta}^*$, $\mathbf{V}(\boldsymbol{\theta}_1^*)$ and $\mathbf{V}(\boldsymbol{\theta}^*)$ may differ.

⁸Note that this is equivalent to minimizing (3), where the expectation is conditional on $s_i < c$.

Let $\boldsymbol{\theta}_2^* = \arg \max_{\boldsymbol{\theta}} E_{y,z}[\log f(y_i|z_i, \boldsymbol{\theta})|s_i > c]$. Then, similarly $\sqrt{n}(\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^*) \xrightarrow{d} N(0, \mathbf{V}(\boldsymbol{\theta}_2^*))$, $\widehat{\boldsymbol{\theta}}_2 \xrightarrow{a.s.} \boldsymbol{\theta}_2^*$ and $\mathbf{V}(\widehat{\boldsymbol{\theta}}_2) \xrightarrow{a.s.} \mathbf{V}(\boldsymbol{\theta}_2^*)$. Therefore, under H_0 , (9) and (10) hold, and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_2) \xrightarrow{d} N(0, \mathbf{V}(\boldsymbol{\theta}_1^*) + \mathbf{V}(\boldsymbol{\theta}_2^*))$, implying the asymptotic chi-square distribution of the quadratic form W . ■

Thus, the Chow test associated with an adequate sorting score is consistent, i.e. has the correct size asymptotically. Note that the choice of c is irrelevant for consistency. We now give a result useful in the sequel.

Proposition 3.2 *If $\boldsymbol{\theta}$ is KL-invariant for z_i , and y_i and u_i are two random variables which are independent conditional on z_i , then $\boldsymbol{\theta}$ is also KL-invariant for u_i .*

Proof. We have

$$\begin{aligned} E_{y,z|u}[\log f(y_i|z_i; \boldsymbol{\theta})|u_i] &= E_{z|u}\{E_{y|z,u}[\log f(y_i|z_i; \boldsymbol{\theta})|z_i, u_i]|u_i\} \\ &= E_{z|u}\{E_{y|z}[\log f(y_i|z_i; \boldsymbol{\theta})|z_i]|u_i\} \\ &= E_{z|u}\{h(z_i, \boldsymbol{\theta})|u_i\}. \end{aligned}$$

The second equality is a consequence of y_i and u_i being independent, conditional on z_i under H_0 .

By the KL-invariance of $\boldsymbol{\theta}$ for z_i ,

$$I(g : f, \boldsymbol{\theta}|z_i) = E_{y|z} \left(\log \frac{g(y_i|z_i)}{f(y_i|z_i; \boldsymbol{\theta})} \Big| z_i \right)$$

is minimized for $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ for any z_i . Hence, $\arg \max_{\boldsymbol{\theta}} h(z_i, \boldsymbol{\theta}) = \boldsymbol{\theta}^*$ for any z_i . Therefore, we have, that $h(z_i, \boldsymbol{\theta}) \square h(z_i, \boldsymbol{\theta}^*)$ for all z_i , which implies

$$\int h(z_i, \boldsymbol{\theta})g(z_i|u_i)dz_i \square \int h(z_i, \boldsymbol{\theta}^*)g(z_i|u_i)dz_i \text{ for all values for } u_i.$$

Hence,

$$\arg \max_{\boldsymbol{\theta}} E_{z|u}\{h(z_i, \boldsymbol{\theta})|u_i\} = \boldsymbol{\theta}^*.$$

Thus, $\boldsymbol{\theta}$ is also KL-invariant for u_i . ■

Remark 3.1 *By Proposition 3.2, z_i is exogenous for $\boldsymbol{\theta}$ as soon as there is no random variable u_i which is both dependent on z_i , and dependent on y_i conditionally on z_i .*

3.2 Power of the test

When H_0 in Proposition 3.1 does not hold, then in general $\theta_1^* \neq \theta_2^*$, see e.g. Example 2.3. The above Chow test will then have power depending on the actual difference $\theta_1^* - \theta_2^*$, which depends on the cut point c . The choice of c also affects the power through its influence on $\mathbf{V}(\hat{\theta}_1^*)$ and $\mathbf{V}(\hat{\theta}_2^*)$, e.g., the value of c determines the number of observations available in the two sub-samples. In both cases, the influence of c on the power is case dependent and as a general rule, c may be chosen as the median of the observed values for the sorting score, in order to have 50% of the observations in each sub-sample.

Two main situations can be distinguished: (i) θ is KL -invariant for z_i even when the latter variable is endogenous and (ii) θ is KL -invariant for z_i only if z_i is exogenous.

Under the first situation (see e.g. Example 2.3) Proposition 3.2 suggests sorting scores guaranteeing the size of the test of exogeneity, i.e. choose s_i such that under H_0 , " z_i is exogenous for θ " then y_i and s_i are independent conditional on z_i . When H_0 does not hold, then there exists, by definition, an unobserved variable u_i dependent on z_i , such that θ is not KL -invariant for u_i . To obtain power for the test an obvious choice is $s_i = u_i$. The latter is, however, typically unobserved. Instead assume that we have a sorting score s_i which is dependent on u_i (and therefore on z_i) and such that y_i and s_i are independent, conditional on u_i and z_i .^{9,10} Then, we can write

$$\begin{aligned} E_{y,z}[\log f(y_i|z_i; \theta)|s_i < c] &= E_{u,z}\{E_{y|u,z}[\log f(y_i|z_i; \theta)|u_i, z_i, s_i < c]|s_i < c\} \\ &= E_{u,z}\{E_{y|u,z}[\log f(y_i|z_i; \theta)|u_i, z_i]|s_i < c\} \\ &= E_{u,z}\{h(u_i, z_i, \theta)|s_i < c\}. \end{aligned}$$

Because θ is not KL -invariant for u_i , $\arg \max_{\theta} h(u_i, z_i, \theta) = \theta(u_i)$. Thus, in general,

$$\arg \max_{\theta \in \Theta} E_{u,z}\{h(u_i, z_i, \theta)|s_i < c\} \neq \arg \max_{\theta} E_{u,z}\{h(u_i, z_i, \theta)\},$$

thereby implying that $\theta_1^* \neq \theta_2^*$.

⁹This latter condition is not restrictive, and means that s_i would not need to be introduced in the model to explain y_i , were z_i and u_i both available.

¹⁰In practice, such a sorting score is obtained with the help of instrumental variables, see Example 3.1.

For the second situation when the model is specified such that $\boldsymbol{\theta}$ is *KL*-invariant for z_i only if z_i is exogenous for $\boldsymbol{\theta}$ it is possible to test H_0 : " z_i is exogenous" by using $s_i = z_i$. This is a test for exogeneity without instruments and it will have power since, in general, $\arg \max_{\boldsymbol{\theta}} E_y[\log f(y_i | z_i; \boldsymbol{\theta}) | z_i < c] \neq \arg \max_{\boldsymbol{\theta}} E_y[\log f(y_i | z_i; \boldsymbol{\theta}) | z_i > c]$.

Example 3.1 *We conclude this section by revisiting Example 2.3, without making any distributional assumption. We have*

$$E_{y|\mathbf{x},z,u}(y_i|\mathbf{x}_i, z_i, u_i) = \alpha + \mathbf{x}_i' \boldsymbol{\beta} + \gamma z_i + u_i. \quad (11)$$

If $E_{u|z}(u_i|z_i) = 0$, i.e. $\lambda = 0$ in (6), then z_i is exogenous for γ . On the other hand, if $E_{u|z}(u_i|z_i) = f(z_i)$ then z_i is endogenous for γ , see Example 2.3. Two cases are distinguished: (i) $E_{u|z}(u_i|z_i)$ is non-linear and (ii) $E_{u|z}(u_i|z_i)$ is linear in z_i . Under (i) $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}', (\gamma + \rho))'$ is not *KL*-invariant for z_i and, hence, the Chow test associated with the sorting score $s_i = z_i$ has power to identify the endogeneity. On the other hand, under (ii) the same parameter $\boldsymbol{\theta}$ is *KL*-invariant for z_i , and another sorting score than z_i needs to be used to detect the endogeneity. Thus, we need to find another variable for which $\boldsymbol{\theta}$ is not *KL*-invariant under endogeneity. The error term η_i is here a natural choice, see (5) in Example 2.3. Because η_i is not observed, the estimate $s_i = \hat{\eta}_i$, i.e. the residuals from fitting (5), may be used. This is a sorting score based on instruments, here \mathbf{x}_i^* .

3.3 Models for endogeneity

We now discuss the Chow test and its power when associated with commonly used models of endogeneity (see, e.g., Vella, 1992). In these models, endogeneity is most often introduced through an unobserved variable, u_i , in the mean function, i.e.

$$E_{y|\mathbf{x},z,u}(y_i|\mathbf{x}_i, z_i, u_i) = h(\mathbf{x}_i' \boldsymbol{\beta} + \gamma z_i + u_i), \quad (12)$$

or with a random coefficient specification

$$E_{y|\mathbf{x},z,u}(y_i|\mathbf{x}_i, z_i, u_i) = h(\mathbf{x}_i' \boldsymbol{\beta} + \gamma z_i + u_i z_i), \quad (13)$$

see, e.g., Garen (1984). Such specifications may be completed with a model for z_i ,

$$z_i^* = \mathbf{x}_i^{*'} \boldsymbol{\delta} + \eta_i, \quad (14)$$

with either $z_i = z_i^*$ (continuous treatment case) or z_i^* is a latent variable and $z_i = \mathbb{I}(z_i^* > 0)$, where $\mathbb{I}(\cdot)$ is the indicator function, and $z_i = 1$ corresponds to treated individuals. Here, $E_{\eta|\mathbf{x}^*}(\eta_i|\mathbf{x}_i^*) = 0$. Then, a standard assumption for introducing endogeneity is

$$E_{u|\eta}(u_i|\eta_i) = \phi + \lambda\eta_i, \quad (15)$$

where ϕ is an intercept.

3.3.1 The random intercept model

If u_i and z_i are independent, then ignoring u_i can lead to a parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}', \gamma)'$ which is either KL -invariant for z_i (i) or not (ii). This independence assumption is stronger than $\lambda = 0$, and by Proposition 3.2 z_i is exogenous for $\boldsymbol{\theta}$ as soon as (i) holds.

Under (ii), the Chow test associated with z_i as the sorting score has power against the presence of the heterogeneity u_i , while under (i), the Chow test associated with z_i as a sorting score has the right size under exogeneity. By Proposition 2.1 (i) occurs, for instance, with distributions from the LEF with a well-specified conditional expectation $E_{y|z}(y_i|z_i)$. Examples where discarding an independent heterogeneity u_i in (12) does not lead to a misspecification of $E_{y|z}(y_i|z_i)$ includes the linear regression model but also the Poisson model with a canonical link (see Section 4.2 below), and the Bernoulli model with a probit link and normal u_i (this case is further discussed in Sections 4.3 and 5).

When u_i and z_i are dependent, e.g. $\lambda \neq 0$, then parameter $\boldsymbol{\theta}$ is in general not KL -invariant for z_i , in which case the Chow test has power when the data is sorted with respect to z_i . However, there are particular situations, as illustrated in Example 2.3, where $\lambda \neq 0$ and the parameter $\boldsymbol{\theta}$ is KL -invariant for z_i . Then, instruments are needed to perform a Chow test of exogeneity.

3.3.2 The random coefficient model

For the random coefficient model (13), ignoring the presence of u_i generally leads to a non KL -invariant parameter $\boldsymbol{\theta}$ for z_i , even when u_i and z_i are independent. The only exception of which we are aware is the linear regression case. Except such cases, using the Chow test associated with z_i as a sorting score is a test with power for the mere presence of a

random coefficient. To obtain a test using z_i as a sorting score only with power against endogeneity, the random coefficient specification must be taken into account. This implies that assumptions on the distribution of u_i are needed (rendering the Chow test sensitive to such assumptions). In such cases, a closed form for the likelihood is generally not available, but simulated maximum likelihood techniques can be used to estimate the parameters (see Gouriéroux and Monfort, 1996).

3.3.3 Dichotomous z_i

In the above discussion, we have assumed that the sorting with respect to z_i provides two samples where θ is estimable. This is not always the case, however, an example of which arises when z_i is a dummy variable; then a Chow test based on $z_i \geq 0.5$ is not possible since γ cannot be estimated. We will need to include a few individuals with $z_i = 1$ in the $z_i = 0$ group and vice versa. This is not a very efficient solution and a better sorting is obtained with the propensity score $\Pr(z_i = 1 | \mathbf{x}_i^*)$; see de Luna and Johansson (2000) and de Luna and Johansson (2001).

4 Monte Carlo study

In this section, we consider the case where the endogenous variable z_i is continuous and study the small sample performance of the Chow test (8) based on the setting described in (12)-(15), with the sorting scores $s_i = z_i, \eta_i, \hat{\eta}_i, z_i\eta_i$ and $z_i\hat{\eta}_i$, where $\hat{\eta}_i$ is the OLS estimated residual.¹¹

More precisely, six setups are used: One model where y_i is continuous; one where y_i is a count and one where y_i is dichotomous. For those, we consider (i) a random intercept specification (12) where the possibly endogenous variable z_i is log-normal, i.e.,

$$z_i = \exp(\delta_0 + \delta_1 x_i^* + \eta_i), \quad (16)$$

and (ii) a random coefficient setup (13) where the possibly endogenous variable z_i is generated as

$$z_i = \delta_0 + \delta_1 x_i^* + \eta_i. \quad (17)$$

¹¹The results of a simulation study where z_i is a dummy variable were reported in de Luna and Johansson (2000).

In both setups, $\eta_i \sim N(0, 0.5^2)$, $x_i^* = U[0, 1]$, $\delta_0 = 1$, $\delta_1 = -1$. Further, $u_i = \lambda\eta_i + \nu_i$, $\nu_i \sim N(0, 0.3^2)$ in (12) and (13) and $\gamma = 0.5$. Parameter λ varies within the interval $[-0.75, 0.75]$ with increment 0.25. We set $\beta = -0.3$ and $\beta = -\gamma \exp(0.5 + 0.5^2/2)$ for (12) and (13) setups, respectively.¹² The number of observations is fixed at $n = 200, 400$ and 600 . To perform the Chow test, we split the sample into two halves, i.e. c is the median of the respective sorting scores. Throughout, the number of replications is 1000.

The random intercept setup in (12) and (16) can be considered as an inclusion of a log-normal wage or income in the mean function and this can be suspected to be correlated with unobservables. For this setup, the parameters are *KL*-invariant for z_i if $\lambda = 0$ while non-invariant for $\lambda \neq 0$.

For the random coefficient model where the link function $h(\cdot)$ is the identity, then again $\lambda = 0$ provides the size of the test against endogeneity. On the other hand, for the non-linear models (non-linear link), ignoring the random coefficient leads to a misspecification of $E_{y|z}(y_i|z_i)$ and, therefore, the parameters are not *KL*-invariant even when $\lambda = 0$. In these cases, the size of the tests against endogeneity is obtained with $u_i \equiv 0$.

Vella (1992, 1993) gives conditions (e.g., u_i and η_i bivariate normal) for asymptotic optimality of a t -test of $\rho = 0$, based on the estimation of the model

$$E_{y|z,u}(y_i|z_i, u_i) = h(\beta + \gamma z_i + \rho \hat{\eta}_i), \quad (18)$$

with maximum likelihood. Here, $\hat{\eta}_i$ are the residuals from fitting either (16) or (17) with OLS. This test corresponds to the familiar Hausman (1978) test where h is the identity function, and the test by Smith and Blundell (1986) for the Tobit model. For comparison, we include this t -test in the simulation study, although using the QMLE covariance matrix described in Section 2. To our knowledge, a similar Hausman test for non-linear link models with random coefficients is not available. However, we use a t -test of $\rho = 0$ based on an estimation of the model

$$E_{y|z,u}(y_i|z_i, u_i) = h(\beta + \gamma z_i + \rho \hat{\eta}_i z_i), \quad (19)$$

with maximum likelihood, to have a ground for comparison for the Chow test. Garen (1984) shows the OLS estimator of ρ to be consistent. The

¹²This choice of intercept is such that for the count data regression model, $E_{y|z,u}(y_i|z_i, u_i) \approx \exp(0)$ under H_0 .

proposed t -test is, hence, consistent for the linear model. Including $\widehat{\eta}_i z_i$ in the mean leads to heteroscedastic errors, however.

4.1 Linear model

In this first situation, the data are generated using

$$y_i = \beta + \gamma z_i + \omega_i + \varepsilon_i, \quad (20)$$

where ε_i is $N(0, 0.3^2)$, $\omega_i = u_i$ and $\omega_i = u_i z_i$ correspond to the random intercept and random coefficient models, respectively.

Parameters β and γ are estimated with OLS and neglecting ω_i , i.e. a QMLE is utilized. The results from the Monte Carlo study are given in Table 1.¹³

Let us first comment on the random intercept model. Here, z_i is endogenous for γ if $\lambda \neq 0$. We can see that the only time we obtain a size significantly¹⁴ different from the nominal size 5% for the Chow test is with $n = 200$ and $s_i = z_i$. The t -test is most powerful in this case. The power of the Chow test obtained with $s_i = \widehat{\eta}_i$ and $s_i = \eta_i$ is similar. As might have been expected, the power is also higher with $s_i = \widehat{\eta}_i$ than with $s_i = z_i$. Notice that (as discussed in Example 3.1) there is power with $s_i = z_i$ because $E_{u|z}(u_i|z_i)$ is non linear in z_i .

For the random coefficient setup, we can first observe that the sizes of the different Chow tests are slightly too large, however and decrease with n . The t -test is too large for all experiments.¹⁵ The Chow test using z_i as a sorting score displays the best power.

4.2 Count data model

We let the response variable y_i be a count, and $y_i|z_i, \omega_i$ be Poisson distributed with the expectation and variance equal to

$$E_{y|z,\omega}(y_i|z_i, \omega_i) = \exp(\beta + \gamma z_i + \omega_i), \quad (21)$$

¹³For both model setups, the bias and power of the test are, as expected, symmetric with respect to λ and, therefore, only results for $\lambda \geq 0$ are presented.

¹⁴Values which are not significantly different of the nominal size of 5% are those within the interval [3.6%, 6.4%]. This is valid for all tables in this article.

¹⁵However, in non-reported experiments with larger sample sizes ($n = 800, 1200, 1600$), the nominal size was obtained.

Table 1: Linear model (20) with random intercept or random coefficient: power and size of the Chow test (three different sorting scores) and a t -test. The percentage bias of the OLS estimator of γ is also presented.

λ	Random intercept				Random coefficient			
	0.00	0.25	0.50	0.75	0.00	0.25	0.50	0.75
$n = 200$								
$z_i \leq c$	7.20	22.10	61.50	88.90	7.00	67.30	99.50	100.00
$\eta_i \leq c$ and $z_i \eta_i \leq c$	5.90	34.90	86.10	99.20	6.20	50.10	97.30	99.90
$\hat{\eta}_i \leq c$ and $z_i \hat{\eta}_i \leq c$	5.50	34.30	85.60	99.20	5.50	51.30	98.10	99.90
t	5.40	72.00	99.70	100.00	9.20	83.10	99.90	100.00
BIAS	0.09	-16.68	-33.46	-50.23	-0.15	-17.99	-35.83	-53.67
$n = 400$								
$z_i \leq c_i$	5.60	36.00	85.60	99.50	6.40	93.00	100.00	100.00
$\eta_i \leq c$ and $z_i \eta_i \leq c$	5.60	61.70	99.40	100.00	6.60	82.40	100.00	100.00
$\hat{\eta}_i \leq c$ and $z_i \hat{\eta}_i \leq c$	5.40	61.60	99.10	100.00	6.60	82.20	100.00	100.00
t	5.40	95.00	100.00	100.00	6.60	97.40	100.00	100.00
BIAS	0.01	-16.44	-32.88	-49.33	-0.01	-18.30	-36.58	-54.86
$n = 600$								
$z_i \leq c$	5.20	48.60	96.90	100.00	5.80	98.60	100.00	100.00
$\eta_i \leq c$ and $z_i \eta_i \leq c$	5.90	77.20	99.90	100.00	5.70	96.10	100.00	100.00
$\hat{\eta}_i \leq c$ and $z_i \hat{\eta}_i \leq c$	6.00	77.50	99.90	100.00	5.20	95.50	100.00	100.00
t	5.30	98.90	100.00	100.00	7.20	99.80	100.00	100.00
BIAS	0.04	-16.33	-32.71	-49.09	0.18	-18.54	-37.25	-55.97

where $\omega_i = u_i$ and $u_i z_i$ in the random intercept and random coefficient models, respectively.

The model is estimated ignoring ω_i and using a QMLE under the assumption that y_i is Poisson distributed. The results from the simulations are given in Tables 2 and 3.

For the random intercept setup, we can see from Table 2 that the power of the tests and the bias of the Poisson maximum likelihood estimator are almost symmetric with respect to λ . We can also see that the sizes of the t -test and the Chow test using either η_i or $\hat{\eta}_i$ as sorting scores are of the same magnitude and slightly too large. Nonreported results with sample sizes $n = 800$ and 1000 , gave correct sizes however. The t -test has the best power. The Chow test with either η_i or $\hat{\eta}_i$ displays similar power, while it yields the lowest power when using $s_i = z_i$.

In the random coefficient model (Table 3), the bias of the Poisson maximum likelihood estimator and the power of the test is larger for λ positive than for λ negative. We can also see that the size of the tests against endogeneity (more precisely $u_i \equiv 0$) is at the nominal level, except for the t -test. However, with $n = 800, 1000, 1200$ and 1400 , the t -test yielded sizes 6.3, 6.8, 5.1 and 5.8 percent, respectively. The t -test (which uses instruments) was most powerful and the Chow test has the largest power when using z_i as a sorting score, i.e. without instruments.

4.3 Discrete choice model

The data is here generated as

$$y_i = \mathbb{I}(\beta + \gamma z_i + \omega_i + \varepsilon_i > 0), \quad (22)$$

where ε_i is $N(0, 0.3^2)$ and $\omega_i = u_i$ and $u_i z_i$ in the random intercept and random coefficient models, respectively. The parameters are estimated with a probit maximum likelihood, ignoring ω_i , i.e. a QMLE is used.

The results from the simulations are provided in Tables 4 and 5. We note that the bias (here defined as $BIAS = \hat{\gamma}/\hat{\beta} - \gamma/\beta$)¹⁶ from the misspecified (endogeneity ignored) probit maximum likelihood estimator is generally small. For the random coefficient specification (Table 5), the

¹⁶Since we can only estimate the parameters up to scale (i.e. β/σ and γ/σ where $\sigma = \sqrt{\sigma_\varepsilon^2 + \sigma_\omega^2}$), σ_ε^2 and σ_ω^2 are the variances of ε_i and ω_i , conditional on z_i) in the discrete choice model.

Table 2: Count data random intercept model (21): power and size of the Chow test (three different sorting scores) and the t -test. The percentage bias of the Poisson ML estimator is also presented.

λ	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
$n = 200$							
$z_i \leq c$	37.60	21.30	8.70	6.50	10.60	20.20	36.80
$\eta_i \leq c$	51.90	31.60	13.60	7.80	13.90	32.30	57.80
$\hat{\eta}_i \leq c$	52.70	31.00	12.10	7.50	13.40	31.00	57.60
t	89.10	60.90	21.90	6.00	26.30	66.10	92.20
BIAS	41.12	26.47	13.38	0.22	-10.91	-21.57	-32.01
$n = 400$							
$z_i \leq c$	63.20	36.00	12.80	5.10	14.60	35.10	63.80
$\eta_i \leq c$	82.90	51.00	17.50	6.00	20.20	55.30	83.00
$\hat{\eta}_i \leq c$	83.30	51.60	17.70	6.20	20.30	54.90	83.20
t	99.40	86.70	35.10	6.50	38.70	88.20	98.90
BIAS	38.62	24.52	12.18	0.43	-10.23	-20.70	-30.26
$n = 600$							
$z_i \leq c$	81.50	51.30	17.50	5.50	19.80	50.30	80.40
$\eta_i \leq c$	95.70	70.10	24.90	6.30	27.00	70.90	94.70
$\hat{\eta}_i \leq c$	96.00	70.20	24.50	6.40	28.40	71.10	94.50
t	100.00	97.00	52.90	6.80	54.30	94.80	99.80
BIAS	37.38	23.96	11.68	0.59	-10.11	-19.90	-29.21

Table 3: Count data random coefficient model (21): power and size of the Chow test (three different sorting scores) and the t -test. The percentage bias of the Poisson ML estimator is also presented.

λ	-0.75	-0.50	-0.25	0.00	$u_i \equiv 0$	0.25	0.50	0.75
$n = 200$								
$z_i \leq c$	57.80	32.50	12.40	7.70	7.50	25.50	65.40	92.60
$z_i \eta_i \leq c$	41.10	22.10	9.70	6.30	5.90	13.50	32.70	60.80
$z_i \hat{\eta}_i \leq c$	40.40	22.50	10.40	6.60	5.70	12.70	31.20	59.70
t	84.40	56.80	21.60	7.70	5.90	39.10	85.60	98.40
BIAS	45.55	27.86	5.94	-21.62	-3.27	-59.02	-106.35	-166.65
$n = 400$								
$z_i \leq c$	86.90	57.30	18.20	7.00	5.00	41.70	90.10	99.60
$z_i \eta_i \leq c$	69.10	38.40	12.80	6.70	6.50	18.10	52.50	88.90
$z_i \hat{\eta}_i \leq c$	68.60	39.40	12.80	6.80	6.80	18.20	53.00	88.20
t	99.20	83.30	36.40	8.80	6.60	59.50	97.30	100.00
BIAS	45.98	28.19	5.75	-22.52	-3.35	-61.00	-112.02	-177.59
$n = 600$								
$z_i \leq c$	97.20	74.60	24.60	7.70	6.10	56.60	98.10	99.90
$z_i \eta_i \leq c$	85.30	52.90	17.90	6.80	5.20	23.30	71.10	97.40
$z_i \hat{\eta}_i \leq c$	84.60	53.50	18.80	7.10	6.60	23.50	71.50	97.10
t	99.90	95.10	47.40	9.20	7.40	73.90	99.80	100.00
BIAS	48.09	29.88	6.96	-21.74	-2.64	-60.77	-113.28	-180.87

bias is not a monotonous function of λ ; when $\lambda > 0$, the bias is small and negative; for $\lambda < 0$ the bias is positive and quite large.

In Table 4, the sizes of the tests are at the nominal level. The t -test is as expected (Vella, 1992) the one with the best power. Furthermore, sorting with respect to η_i or $\hat{\eta}_i$ makes little difference. The power of the test when sorting with respect to z is quite low, especially when the correlation is positive.

In Table 5 (random coefficient model), the size of the tests against endogeneity ($u_i \equiv 0$) is at the nominal level, except for the t -test. When $n = 800, 1000, 1200$ and 1400 we obtain the sizes 7.7, 6.4, 6.5 and 6.4 percent, respectively. Hence, the t -test does not seem to be consistent in this setting. Among the Chow tests, sorting with respect to z_i yields the best power. The power against the presence of a random coefficient ($\lambda = 0$) is largest with the Chow test together with z_i as the sorting score.¹⁷

4.4 Discussion

In practice, it is typically unknown if a random intercept or a random coefficient is closest to reality. Ideally, we want a test for the exogeneity of a treatment z_i which is as robust to various misspecifications as possible. For instance, t -tests (Hausman type) are sensitive to distributional specifications as shown in de Luna and Johansson (2000)¹⁸ and they need instrumental variables. In the above Monte Carlo study, we have further noticed that the size of the t -test (by including $z_i \hat{\eta}_i$) may be too large when individual heterogeneity enters as a random coefficient.¹⁹

When using a Chow test, the presence of a random coefficient should, in most situations, be accounted for, since ignoring it most often leads to non KL -invariant parameters. If the random coefficient can be integrated

¹⁷Since u_i and ε_i are independent normal with expectations 0 and variances σ_u^2 and σ_ε^2 , respectively, we have that $E_{y|z}(y_i|z_i) = \Phi\left(\frac{(\beta + \gamma z_i)/\sqrt{\sigma_u^2 z_i^2 + \sigma_\varepsilon^2}}{\sqrt{\sigma_u^2 z_i^2 + \sigma_\varepsilon^2}}\right)$ and hence, the Chow test based on any sorting will have power.

¹⁸In this paper, simulation results show that the t -test can have empirical sizes not equal to the nominal level, for instance when the errors in the selection equation are skewed (e.g. χ^2 distributed).

¹⁹In unreported experiments for the random coefficient setup, a Hausman test for the presence of $\hat{\eta}_i$ yielded correct sizes, although with much lower power than for the Chow tests.

Table 4: Discrete choice random intercept model (22): power and size of the Chow test (three different sorting scores) and the t -test. The percentage bias of the ratio estimate ($\hat{\gamma}/\hat{\beta}$) of the probit ML estimator is also presented.

λ	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75
$n = 200$							
$z_i \leq c$	51.20	21.90	8.80	6.50	7.30	9.60	10.90
$\eta_i \leq c$	86.40	48.10	15.60	6.20	13.00	38.60	66.20
$\hat{\eta}_i \leq c$	86.40	47.20	14.20	5.80	12.70	37.90	65.50
t	100.00	93.00	37.20	6.20	28.60	73.30	94.40
BIAS	-0.68	-1.65	-0.90	0.51	2.03	3.38	4.58
$n = 400$							
$z_i \leq c$	74.40	35.60	9.80	6.00	7.80	11.70	13.90
$\eta_i \leq c$	99.10	81.10	24.80	7.10	23.70	69.50	93.50
$\hat{\eta}_i \leq c$	99.50	79.30	23.60	6.10	23.40	69.10	93.50
t	100.00	99.90	65.10	6.10	51.10	95.20	99.90
BIAS	1.55	-0.57	-0.70	0.24	1.37	2.51	3.50
$n = 600$							
$z_i \leq c$	90.10	49.20	11.10	4.70	7.20	12.00	17.20
$\eta_i \leq c$	99.90	93.10	34.90	5.10	34.20	84.00	98.80
$\hat{\eta}_i \leq c$	99.90	93.00	33.90	5.60	31.80	83.70	98.70
t	100.00	100.00	79.70	5.40	67.20	99.80	100.00
BIAS	2.44	-0.19	-0.55	0.09	1.17	2.18	3.08

Table 5: Discrete choice random coefficient model (22): power and size of the Chow test (three different sorting scores) and the t -test. The percentage bias of the ratio estimate ($\widehat{\gamma}/\widehat{\beta}$) of the probit ML estimator is also presented.

λ	-0.75	-0.50	-0.25	0.00	$u_i \equiv 0$	0.25	0.50	0.75
$n = 200$								
$z_i \leq c$	98.50	89.40	52.70	15.40	6.20	12.40	44.80	80.80
$z_i \eta_i \leq c$	92.20	69.10	28.00	9.30	5.80	18.40	60.00	91.00
$z_i \widehat{\eta}_i \leq c$	92.00	67.60	28.00	9.00	6.00	20.10	58.10	91.10
t	100.00	98.50	75.80	16.20	8.10	26.50	78.20	98.20
BIAS	60.46	28.46	13.13	4.36	-0.03	-0.77	-4.39	-7.37
$n = 400$								
$z_i \leq c$	100.00	99.60	84.90	24.10	4.80	18.00	70.90	97.60
$z_i \eta_i \leq c$	99.90	96.00	57.30	13.00	5.30	29.40	85.90	99.90
$z_i \widehat{\eta}_i \leq c$	99.90	95.20	56.70	13.40	5.70	30.00	86.10	99.70
t	100.00	100.00	94.80	23.40	6.60	37.70	96.10	100.00
BIAS	50.48	27.18	12.75	4.40	0.06	-0.58	-4.11	-6.99
$n = 600$								
$z_i \leq c$	100.00	100.00	94.90	36.50	5.40	24.10	86.60	99.50
$z_i \eta_i \leq c$	100.00	99.90	74.70	17.70	6.20	43.70	95.90	100.00
$z_i \widehat{\eta}_i \leq c$	100.00	99.60	73.80	17.00	6.20	43.70	95.70	100.00
t	100.00	100.00	99.10	29.70	7.50	46.90	99.50	100.00
BIAS	49.90	27.18	12.86	4.44	0.06	-0.48	-3.92	-6.83

out (typically distributional assumptions on u_i must be made) then a Chow test against endogeneity may be performed without instruments, by sorting with respect to z_i .

5 Empirical illustration

Johansson and Palme (1996) examine the effect of economic incentives on absenteeism. Implications of compensating wage differentials and efficiency wage hypotheses are discussed. The data set is a sample of 1,967 blue collar workers (1,045 women and 922 men) obtained from the 1981 Swedish Level of Living Survey. The dependent variable is the number of days each individual is absent per year.

In their model, the demand for absent time of person i , $t_i^a = \alpha + \gamma c_i + \delta \mu_i + x_i' \beta + \varepsilon_i$, is a latent variable that is not directly observable, where c_i is the cost of being absent, μ_i is virtual income and x_i are socioeconomic variables. In Johansson and Palme (1996), an individual is assumed to be absent, $I_i = 1$, if $t_i^a > 0$ and at work, $I_i = 0$, if $t_i^a \leq 0$. Then, y_i is defined as the number of days person i is absent during a year and the parameter $\theta = (\alpha, \gamma, \delta, \beta)'$ is estimated with QMLE under the assumption that y_i is binomial with parameters $N = 365$ and

$$\pi_i = \Phi(\alpha + \gamma c_i + \delta \mu_i + \mathbf{x}_i' \beta), \quad (23)$$

where $\Phi(\cdot)$ is the normal distribution function, i.e.

$$\ln L = \sum_{i=1}^n y_i \ln \pi_i + (N - y_i) \ln(1 - \pi_i), \quad (24)$$

is maximized with respect to θ . In the case of unobserved heterogeneity (in the form of a normally distributed random intercept, independent of the regressors) (23) corresponds to the conditional expectation $E(y_i | c_i, \mu_i, \mathbf{x}_i)$, and, hence, the QMLE is consistent.

The cost of being absent is $c_i = (1 - r)w_i$, where w_i is the net wage and r is the share of the income the worker receives when absent. Since the cost partly depends on the income, we can expect it to be endogenous. There are at least four reasons why unobserved preference heterogeneity may be correlated with the individual's wage rate. First, a worker with strong preferences for work absence may choose a job allowing her to be

absent from work at the cost of, *ceteris paribus*, a lower wage rate (see e.g. Allen, 1981). Second, since jobs differ in the cost of absenteeism for the employer, it may be profitable for the employer to pay some employees more, i.e. an efficiency wage, in order to elicit them to have low work absence rates (see, e.g., Shapiro and Stiglitz, 1984; Weiss, 1985). Third, when there are economic returns to on-the-job training, then individuals with strong preferences for work absence will, everything else equal, earn less. Finally, it is a well known empirical fact that, on average, workers with bad health have a higher work absence rate than workers without health problems (see, e.g., Broström, Johansson and Palme, 1998). For some jobs, it is reasonable to assume that workers with bad health are less productive and, therefore, earn less than workers with a good health status.²⁰

From the above discussion, it is reasonable to assume that γ may be individual specific and negatively correlated with the cost of being absent, hence $\pi_i = \Phi(\alpha + \gamma_i c_i + \delta \mu_i + \mathbf{x}'_i \boldsymbol{\beta})$. We sort the data with respect to c_i and estimate model (23) on two subsets such that each subset has an equal number of observations. The result from the exercise is given in Table 6. We note that the parameter estimates from the two samples are quite different, especially for variables correlated with c_i and the parameter for c_i itself. The Chow test is significant at the five percent level. Therefore, we can reject *KL*-invariance of $\boldsymbol{\theta}$ for c_i in this case. This lack of invariance may be due to the fact that the variable c_i should be included non-linearly in (23), or due to a random coefficient for c_i (either correlated with c_i or not). In order to test for selectivity (i.e. that individuals with low γ_i are the ones with the highest cost) we expand (23) to account for the possibility of an independent and normally distributed random coefficient with mean zero and variance σ_u^2 , i.e. we use

$$\pi_i = \Phi \left((\alpha + \gamma c_i + \delta \mu_i + \mathbf{x}_i \boldsymbol{\beta}) / \sqrt{\sigma_u^2 c_i^2 + 1} \right),$$

together with the log-likelihood function (24). With a *t*-test, we cannot reject that $\sigma_u = 0$.²¹ A Chow test based on the same sub-samples as previously yields $W = 33.14$ (p -value = 0.060), i.e. the test does not clearly reject the hypothesis of nonselectivity. The empirical evidence,

²⁰The relation between wages and absenteeism has been studied in several theoretical as well as empirical studies (see Brown and Sessions, 1996, for an overview).

²¹This result might be due to the random coefficient being correlated with c_i .

therefore, is here slightly ambiguous if no distributional assumptions are made. However, if distributional assumptions are made, i.e. the inverse of the Hessian is used for the covariance matrix instead of the QMLE covariance matrix, then we obtain $W = 3100.5$ (in the random coefficient setup), i.e. strong evidence against nonselectivity of c_i .

From Table 6, we can see that the parameter estimates for μ_i and c_i have the correct sign for the sample with larger costs and the wrong sign for the sample with lower costs. Hence, neglecting the presence of endogeneity/selectivity would (as expected) reduce the estimated effect of economic incentives on work absence.

6 Conclusion

We have proposed a general test of exogeneity which, in some cases, is robust against distributional misspecification. Indeed, by defining the concept of KL -invariance of a parameter, we give conditions under which the exogeneity hypothesis can be tested without making strong distributional assumptions. A test against endogeneity can be carried out without instrumental variables, by using the (supposedly) endogenous variable as a sorting score.

Based on the results of this paper, we propose the following two strategies for the practitioner, depending on whether (i) we know that the parameters of the model are KL -invariant for z_i only if the variable z_i is exogenous or (ii) the model might be KL -invariant in z_i , even under endogeneity (see Example 2.3).

In both situations we suggest to start by performing a Chow test by sorting with respect to z_i . If the KL -invariance in z_i is not rejected, then: if (i) holds, proceed as if z_i were exogenous, while if (ii) holds, proceed with a sorting based on instruments and perform another Chow test. If, on the other hand, the former Chow test rejected KL -invariance, then z_i is not exogenous and measures for dealing with this must be taken. The lack of KL -invariance for z_i indicates a misspecification of the conditional expectation (see Remark 2.2): for instance, this non-invariance may be due to the presence of a random coefficient. Thus, if the model is expanded to include an independent random coefficient, then this independence (no selectivity) can also be tested with a Chow test by sorting the data with respect to z_i .

Table 6: Parameter estimates of the binomial regression model, for the two samples sorted with respect to c . The QML covariance matrix is being used to estimate standard errors (s.e.) and to calculate the Chow test statistic W .

c_i Parameters	0.40 – 1.28		1.28 – 25.03	
	Est.	Est./s.e.	Est.	Est./s.e.
CONST	-3.82	-6.44	-2.63	-7.82
Contracted working time	0.19	2.91	0.02	0.62
μ_i	-0.00	-1.41	0.00	0.60
c_i	0.10	0.41	-0.03	-1.42
Local unemployment rate	-0.07	-1.52	-0.08	-2.04
FEMALE	0.21	2.28	0.27	2.22
SINGLE	0.02	0.16	0.17	1.33
DIVORCED	-0.09	-0.87	0.35	2.45
DISABLED	0.47	5.31	0.44	4.86
AGE	0.01	3.91	0.01	2.57
Important to be on time	0.15	1.85	-0.01	-0.12
Punch clock	0.14	1.75	0.05	0.54
Risk at the workplace	0.05	1.63	0.03	1.05
Number of children				
< 6 years	-0.10	-1.26	-0.03	-0.54
7 years–16 years	-0.09	-1.62	-0.02	-0.32
Principal components: working conditions				
1:st	0.14	2.46	0.13	2.38
2:nd	-0.15	-2.17	-0.12	-1.48
3:rd	0.01	0.14	0.03	0.19
Principal components: health conditions				
1:st	0.12	1.21	-0.28	-3.13
2:nd	-0.31	-2.25	-0.19	-1.62
3:rd	0.21	1.90	0.11	0.82
W	33.02 (p-value 0.045)			

References

- Allen, S. G. (1981). "Compensation, Safety, and Absenteeism: Evidence from the Paper Industry", *Industrial and Labor Relations Review* **34**, 77 – 87.
- Broström, G., Johansson, P. and Palme, M. (1998). "Assessing the Effect of Economic Incentives on Incidence and Duration of Work Absence", *Working Paper Series in Economic and Finance* **228**, Stockholm School of Economics.
- Brown, S. and Sessions, J. G. (1996). "The Economics of Absence: Theory and Evidence", *Journal of Economic Surveys* **10**, 23 – 53.
- Chow, G. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica* **28**, 591 – 605.
- de Luna, X. and Johansson P. (2000). "Testing Exogeneity in Cross-section Regression by Sorting Data", *Working Paper* **2000:2**, Swedish Office of Labour Market Policy Evaluation (IFAU).
- de Luna, X. and Johansson P. (2001). "Graphical Diagnostics of Endogeneity", *Ume Economic Studies* **553**, Ume University.
- Engle, R. F., Hendry, D. F. and Richard J.-F. (1983). "Exogeneity", *Econometrica* **51**, 277 – 304.
- Garen, J. (1984). "The returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable", *Econometrica* **52**, 1199 – 1218.
- Gouriéroux, C. and Montfort, A. (1995). *Statistics and Econometric Models*, Cambridge University Press. Cambridge.
- Gouriéroux, C. and Montfort, A. (1996). *Simulation-based Econometric Methods*, Oxford University Press. Oxford.
- Gouriéroux, C., Monfort, A. and Trognon, A. (1984). "Pseudo Maximum Likelihood Methods: Theory", *Econometrica* **52**, 681 – 700.
- Hausman, J. A. (1978). "Specification Test in Econometrics", *Econometrica* **46**, 1251 – 1271.

- Heckman, J. J., Ichimura, H. and Todd, P. (1998). "Matching as an Econometric Evaluation Estimator", *Review of Economic Studies* **65**, 261 – 294.
- Holland, P. (1986). "Statistics and Causal Inference", with discussion, *Journal of the American Statistical Association* **81**, 945 – 970.
- Imbens, G. W. (1997) "Book Review: The Foundation of Econometric Analysis", *Journal of Applied Econometrics* **12**, 91 – 94.
- Johansson, P. and Palme, M. (1996). "Do Economic Incentives Affect Worker Absence? Empirical Evidence Using Swedish Data", *Journal of Public Economics* **59**, 195 – 218.
- Koopmans, T. C. (1950). "When is an Equation System Complete for Statistical Purposes", in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans, John Wiley and Sons. New York.
- Kullback, S. and Leibner, R. A. (1951). "On Information and Sufficiency", *Annals of Mathematical Statistics* **22**, 79 – 86.
- Rosenbaum, P. R. (1995). *Observational Studies*, Springer Series in Statistics, Springer Verlag. New York.
- Rosenbaum, P. R. and Rubin, D. B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effect", *Biometrika* **70**, 41 – 55.
- Shapiro, C. and Stiglitz, J. (1984). "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review* **74**, 433 – 444.
- Smith, J. S. and Blundell, R. (1986). "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply", *Econometrica* **54**, 679 – 685.
- Strotz, R. H. and Wold, H. O. A (1960). "Recursive Vs. Nonrecursive Systems: An Attempt at Synthesis", *Econometrica* **28**, 417 – 427.
- Vella, F. (1992). "Simple Tests for Sample Selection Bias in censored and Discrete Choice Models", *Journal of Applied Econometrics* **7**, 413 – 421.

- Vella, F. (1993). "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors", *International Economic Review* **34**, 441 – 456.
- Weiss, A. (1985). "Absenteeism and Wages", *Economic Letters* **19**, 277 – 279.
- White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models", *Econometrica* **50**, 1 – 25.
- Wooldridge, J. M. (1991). "Specification Testing and Quasi-maximum-likelihood", *Journal of Econometrics* **48**, 29 – 55.