Journal of Advances and Scholarly Researches in Allied Education
Vol. 19, Issue No. 2, March-2022, ISSN 2230-7540

To tackle the Identifiability issue and to Estimate the Misclassification Parameters

Karwanje Diwakar Prabhakarrao[1]*, Dr. Rishikant Agnihotri[2]

[1] Ph.d Scholar, Kalinga University, Raipur

[2] Professor, Department of Mathematics, Kalinga University, Raipur

Abstract - Cluster randomization studies have become more common in place of traditional trials that randomly assign participants one at a time when this method is impractical for theoretical, ethical, or practical reasons. In the setting of a complementary poison model with potentially misclassified data, we evaluate three interval estimators for binomial misclassification rates: one based on the wald statistic, another on the score statistic, and a third on the profile log-likelihood statistic. As a result of its improved power and lower type I error, the redesigned test comes highly recommended. Semiparametric testing of misclassification estimates Information on the parameters employed in g (x*, z) that underlie parametric models, misclassification, and model and identification-related problems

Keywords - Misclassification, Asymptotic Confidence Intervals, Simulation

---------------------------------x---------------------------------

INTRODUCTION

Statistics, identifiability is a property which a model must satisfy for precise inference to be possible. A model is identifiable if it is theoretically possible to learn the true values of this model's underlying parameters after obtaining an infinite number of observations from it. Mathematically, this is equivalent to saying that different values of the parameters must generate different probability distributions of the observable variables. Usually, the model is identifiable only under certain technical restrictions, in which case the set of these requirements is called the identification conditions.

A model that fails to be identifiable is said to be non-identifiable or unidentifiable: two or more parametrizations are observationally equivalent. In some cases, even though a model is non-identifiable, it is still possible to learn the true values of a certain subset of the model parameters. In this case we say that the model is partially identifiable. In other cases, it may be possible to learn the location of the true parameter up to a certain finite region of the parameter space, in which case the model is set identifiable.

Class prediction involves the use of statistical learning techniques to develop algorithms for classifying unknown samples through supervised learning on samples of known class. In assessing the performance of a classification algorithm, the goal is to estimate its ability to generalize, i.e., to predict the outcomes of samples not included in the data set used to train the classifier. The performance may be assessed on the basis of a number of different indices. For problems having a dichotomous outcome variable (e.g., positive or negative), the sensitivity, specificity, positive predictive value and negative predictive value are indices that may be of interest in addition to the overall prediction accuracy

Many health sciences issues and social science investigations use multivariate data because the data are often clustered or recorded longitudinally. Similarity between subjects within a cluster is more likely than similarity between subjects from different clusters. When many outcomes are assessed for the same person (a cluster), there is a high probability that they are all connected to one another. As opposed to cross-sectional data, which only collects information at a single moment in time, longitudinal data collects information on the same subjects over the course of numerous time periods, which naturally leads to correlations between the subjects' replies. In both cases, there is an association between the variables of interest because the observations under study share common characteristics. In fact, if an analyst fails to account for this sort of connection, it is possible that they may draw incorrect conclusions about the model's parameters.

When the outcome variable is assumed to follow a normal distribution, several statistical approaches exist for analyzing these types of data, whether they be clustered or longitudinal in nature. When the dependent variable is continuous, procedures for evaluating correlated data are well stocked to handle the situation. For the purposes of this research, correlated ordinal data is of primary interest. It is said that a categorical variable is ordinal if there exists a natural ordering among its several categories, i.e., breast cancer detection using a mammography diagnostic grading system depending on the patient's level of education. Once again, ordinal data is commonly employed in the social sciences for gauging views and perspectives. Strongly disagree, disagree, undecided, agree, and strongly agree are some of the possible responses to a question about one's opinion on a social problem.

LITERATURE REVIEW

Matth¨aus Kleindessner (2017) Ordinal distance information has recently replaced numeric distance measurements as preferred research setting for machine learning challenges. We call the binary results of distance comparisons liked (A, B) d(C, D) ordinal distance information (C, D). There are several machine learning and statistical issues for which it is not known how to approach a solution under these conditions. The standard method up to now has been to manually build an ordinal embedding of the data points in Euclidean space, which has its own set of problems. Given just ordinal data, we offer methods for the issues of medoid estimation, outlier detection, classification, and grouping. Both the lens depth function and the k-relative neighborhood graph are estimated from a given data set to produce these models. Our techniques are straightforward, much quicker than an ordinal embedding approach while avoiding some of its limitations, and readily parallelizable.

Daniel Fernandez (2019) Many psychological and psychiatric investigations gather and utilize ordinal variables. While continuous variable models are comparable to ordinal variable models, there are benefits to using a model built for ordinal data, such as avoiding "floor" and "ceiling" effects and not having to give scores (which might lead to score-sensitive outcomes in continuous models). The ordered stereotype model, created for modeling ordinal outcomes but less well-known than alternatives like linear regression and proportional chances models, is the topic of this research. This paper's goal is to evaluate the ordered stereotype model next to several other popular models utilized in the academic and professional communities. Using three, four, and five levels of ordinal categories and sample sizes of 100, 500, and 1000, this article evaluates the stereotype model in comparison to the proportional odd and linear regression models. This article also uses a simulation study to talk about the issue of considering ordinal replies as continuous. The program also includes the trend odds model. According to the results, three distinct models—an ordered stereotype model, a proportional chances model, and a trend odds model—were all adapted to the same real-world data set.

Haiyan Liu and Zhiyong Zhang (2017) Misclassification is a kind of measurement mistake in categorical data that occurs when the observed category does not match the underlying one. The literature is rich with studies and discussions on the measurement error in continuous data, particularly normally distributed data. However, in psychology, misclassification in a binary outcome variable has not yet received considerable attention. Using a Monte Carlo simulation analysis, we demonstrate that ignoring the misclassification results in significant biases in parameter estimations. We offer a model that incorporates false positive and false negative misclassification parameters to account for the impact of misclassification. In addition to providing information on the level of misclassification, such a model may estimate the underlying connection between the dependent and independent variables. The model is estimated using a maximum likelihood technique using a Newton-type approach. The performance of the new model is evaluated using simulation experiments, and its use is shown with real-world data. To facilitate its use, a corresponding R package is created.

Kent Riggs (2010) We examine the Wald interval, the score interval, and the profile log-likelihood interval as interval estimators for binomial misclassification rates in a supplementary Poisson model with potentially misclassified data. Through a simulation analysis, we examine the coverage and average width aspects of these intervals. The intervals' coverage may be subpar for low Poisson numbers and low misclassification rates. When compared to other intervals, the profile log-likelihood CI is generally shown to be superior due to its superior coverage and breadth qualities. Finally, we implement the CIs on a real-world data set consisting of traffic accident data with misclassified count data.

THE MODEL AND IDENTIFICATION

Here we show identification results for the regression function in models with misclassified regressors of the kind

$$\mathbb{E}[y - g(\tilde{Z})|\tilde{Z}] = 0$$

where g (·) is an unknown conditional expectation function.

So long as the random vector has $\tilde{z}$ dividable into$(x^*, \tilde{Z})$ where x^* Where z is an observable continuous random variable, $d_z \times 1$ Indeterminate veciral sequence. Contrary to what one may think, x^* we see *x*, inadvertently reclassified version of x^* (in academic parlance, a "surrogate"). The model holds if and only if the surrogate and an additional random variable, v (with attributes to be defined below), are both observed.

$$\mathbb{E}[y - g(\tilde{z})|\tilde{z}, x, v] = 0$$

The of non-differential measurement error states that, given the truth and the other covariates z, the conditional mean of y is unaffected by knowledge of x once x is known, and thus that the misclassification rates themselves are uninformative about the outcome of interest. The misclassification rates may provide insight into the responses via their correlation with the other explanatory variables in the model, making the conditional statement crucial. Bound, Brown, and Mathiowetz (2000) provide examples of when such an is likely to hold or not hold in their survey paper on measurement error. An analogous in the nonlinear setting with a convolution model for measurement error is that the error term in the outcome equation is conditionally mean independent of the measurement error in the mismeasured regressor.

Four key premises support the identification argument: There must be a dependency relationship between the unobserved regressor and the ILV and 1) identification of the model (I) when there is no misclassification, 2) limits on the degree to which misclassification can occur, 3) independence between the misclassification rates and the ILV conditional on the other regressors, and 4) no misclassification at all. For the sake of brevity, let's pretend that the ILV v only has two possible values, v_1 and v_2. This makes the arguments more understandable and allows for positive identification. The following comments describe these four presumptions plus one more. All the time, we will suppose that the econometrician uses an i.i.d.

Sample$\{y_i, x_i, z_i, v_i\}_{i=1}^{n}$. Let $z_a \in \mathbb{S}_z$ signify a non-essential part of the bolster of $\mathcal{Z}(\mathbb{S}_z)$ as a shortcut $\mathbb{p}(\cdot|z_a)$ mean to distinguish $\mathbb{p}(\cdot|z = z_a)$

THEOREM 1 Regarding the Presumptions 1-5, g (x^*, z_a) includes the rates of incorrect categorization η_0 (z_a) and η_1 (z_a) in the first model are accounted for.

Remark 1 Modifying to virtually universally hold $\mathcal{Z} \in \mathbb{S}_z$ (By affixing "a.e. $\mathbb{P}_z$ to Screens 2, 4, and 5, and finally the Full Regression Function g (·) and Misclassification Rates η_0 (·) and η_1 (·) are tracked down in model (1)

An appendix has the whole evidence, however here we will summarize its important points. Initially, we demonstrate that the regression function at is only meaningful if the misclassification rates are known (x^*, z_a) has been located. The second part of our method involves demonstrating the detection of the error rates in classification.

For the sake of argumentation, take into account

$$\underbrace{\mathbb{p}(x=1|z_a,v)}_{\eta_2(z_a,v)} = \sum_{s\in\{0,1\}} \mathbb{P}(x=1|x^*=s,z_a,v)\,\mathbb{P}(x^*=s|z_a,v)$$

Let $\eta_2^*(z_a,v) \equiv \mathbb{P}(x^*=1|z_a,v)$. Considering Premise No. 3, we may deduce

$$\eta_2^*(z_a,v) = \frac{\eta_2(z_a,v)-\eta_0(z_a)}{1-\eta_0(z_a)-\eta_1(z_a)}$$

In essence, the purpose η_2 (z_a, v) since the connection is unmistakable *{x, z, v}* be kept an eye on. This means that if the functions {η_0 (z_a), η_1 (z_a)} once those people are found, η^* 2 (z_a, v) also be able to be located.

The primary reasoning leads to the conclusion that if the misclassification rates {η_0 (z_a), η_1 (z_a)} when identified, keep in mind that we may write the identified instant in order to see the argument. $\mathbb{E}$ [y|z_a, v] as

$$\mathbb{E}[y|z_a,v] = g(0,z_a)\big(1-\eta_2^*(z_a,v)\big) + g(1,z_a)\eta_2^*(z_a,v)$$

Once the rates of misclassification are known, it is simple (a linear system of equations, in fact) to deduce using the variation in v

Last but not least, we demonstrate that, given a set of misclassification probabilities, the ILV and the directly observed moments guarantee identification up to a "probability flip." (η_0 (z_a), η_1 (z_a)),

$\big(\tilde{\eta}_0(z_a),\tilde{\eta}_1(z_a)\big) = \big(1-\eta_1(z_a), 1-\eta_0(z_a)\big)$ disproves these possibilities (since the $\tilde{\eta}_0(z_a)+\tilde{\eta}_1(z_a) > 1$ thus the rates of misclassification can be calculated.

Specifically, Appendix A.1 demonstrates that we can directly determine the misclassification rates as a function of the observed moments of w = *(y, x, xy)*

$$\eta_1(z_a) = \big(1-h_1\mathbb{E}(w|z_a,v_1),\mathbb{E}(w|z_a,v_2)\big) + h_0\big(\mathbb{E}(w|z_a,v_1),\mathbb{E}$$

$$\eta_0(z_a) = \big(1+h_1\mathbb{E}(w|z_a,v_1),\mathbb{E}(w|z_a,v_2)\big) + h_0\big(\mathbb{E}(w|z_a,v_1),\mathbb{E}$$

where the exact forms of the well-known smooth functions h1 (·) and h0 (·) are (41) and (40).

Based on the aforementioned misclassification rates, we can then $\eta_2^*(z,v)$ solving for g (x*, z) in terms of (6), which are then used to solve for g (x*, z) The exact form of these yields is given by (46) (see Appendix A.4), and it applies to the case of a smooth well-defined known function q (·).

$$g(1,z_a) = \frac{1}{2}q\big(\mathbb{E}(w|z_a,v_1)\mathbb{E}(w|z_a,v_2)\big) + \frac{1}{2}h_0\big(\mathbb{E}(w|z_a,v_1),\mathbb{E}($$

$$g(0,z_a) = \frac{1}{2}q\big(\mathbb{E}(w|z_a,v_1)\mathbb{E}(w|z_a,v_2)\big) - \frac{1}{2}h_0\big(\mathbb{E}(w|z_a,v_1),\mathbb{E}($$

Insightful and connecting the literature on estimation of endogenous regression models is the formula for the marginal effect implied by the two equations above:

$$g(1,z) - g(0,z) = \frac{\mathbb{E}(y|z_a,v_1)-\mathbb{E}(y|z_a,v_2)}{\mathbb{E}(x|z_a,v_1)-\mathbb{E}(x|z_a,v_2)}\,h_0\big(\mathbb{E}(w|z_a,v_1),\mathbb{E}$$

The first right-hand term is analogous to the Wald estimator of the marginal effect of *x* on *y* with v as the instrument. To account for the fact that the Wald-IV estimator does not identify the marginal effect when there is binary misclassification, the second term can be thought of as a correction term.

Furthermore, the marginal effect's shape implies that the model can be generalized to incorporate endogeneity of the true (unobserved) *x** in a regression setting where the errors are additive rather than multiplicative. Specifically, we can keep track of who performed the function $g^*(x^*, z)$

When

$$y = g^*(x^*, z) + \epsilon$$

Where $\mathbb{E}(\epsilon|z, v) = 0$ which is the standard meaning that *x** is both endogenous and incorrectly labeled. We need to impose the analog of (I), which is, in order to account for the error in measurement.

$$\mathbb{E}(y|x^*, z, x, v) = \mathbb{E}(y|x^*, z)$$

After looking over the Theorem 1 proofs, we can see that the function $g^*(x^*, z_a)$ is still recognizable in this model (a formal argument is included at the conclusion of the proof of Theorem 1 in Appendix A.2). Last but not least, with the corrections described in Theorem 1, the result can be extended to yield identification of the entire regression function $g*(x^*, \cdot)$.

PARAMETRIC MODELS

As a corollary of the preceding identification finding, parametric model identification is possible as well. The parametric binary choice model is an interesting specific instance The binary choice coefficient and the misclassification rates may be determined by appropriately modifying the identification result. Specifically, the model is

$$\mathbb{P}(y = 1|x^*, z, x, v) = \mathcal{F}\left(\theta_{10} + \theta_{20^{x^*}} + \theta_{30^{z}}\right)$$

where F (·) is a strictly growing function that is well known.

Lemma 1 Let's pretend that premises 10–14 are correct. And then for each $v_a \in \{v_1, v_2\}$

$$\sqrt{nh^{d_z}}\left(\hat{\mathbb{E}}(w|z_a, v_a) - \mathbb{E}(w|z_a, v_a)\right) \Rightarrow \mathcal{N}(0, V_a)$$

Were

$$V_a = \begin{bmatrix} Var(x|z_a, v_a) & Cov(x, y|z_a, v_a) & Cov(x, xy|z_a, v_a) \\ Cov(x, y|z_a, v_a) & Var(y|z_a, v_a) & Cov(x, xy|z_a, v_a) \\ Cov(x, xy|z, v) & Cov(y, xy|z_a, v_a) & Var(xy|z_a, v_a) \end{bmatrix}$$

the vectors, and

$$\sqrt{nh^{d_z}}\left(\hat{\mathbb{E}}(w|z_a, v_1) - \mathbb{E}(w|z_a, v_1)\right)$$

And

$$\sqrt{nh^{d_z}}\left(\hat{\mathbb{E}}(w|z_a, v_2) - \mathbb{E}(w|z_a, v_2)\right)$$

asymptotically free of one another.

It is a natural consequence of the Cramer-Wold device and Theorem in Bierens (1987). The "delta" technique allows us to reach this final conclusion.

Lemma 2 Take it for granted that Premises 1–5 and 10–14 are correct. Next, the predictors *gˆ (1, z) and gˆ (0, z)* weakly converge as follows, defined in (17) and (19) above:

$$\sqrt{nh^{d_z}}\left(\hat{g}(1, z_a)\, g(1, z_a)\right) \Rightarrow \mathcal{N}(0, \Omega_1)$$

$$\sqrt{nh^{d_z}}\left(\hat{g}(0, z_a)\, g(0, z_a)\right) \Rightarrow \mathcal{N}(0, \Omega_0)$$

and a slight impact

$$\sqrt{nh^{d_z}}\left(\left(\hat{g}(1, z_a) - \hat{g}(0, z_a)\right) - \left(g(1, z_a) - g(0, z_a)\right)\right) \Rightarrow \mathcal{N}(0, \Omega_M$$

Were

$$\Omega_1 = \frac{f_1\left(\mathbb{E}(w|z_a, v_1), \mathbb{E}(w|z_a, v_2), V(w|z_a, v_1)V(w|z_a\right.}{2\left[\left(\mathbb{E}(y|z_a, v_1) - \mathbb{E}(y|z_a, v_2)\right)\left(\mathbb{E}(x|z_a, v_1) - \mathbb{E}(x|z_a, v_2)\right)\right]^2(1}$$

$$\Omega_0 = \frac{f_0\left(\mathbb{E}(w|z_a, v_1), \mathbb{E}(w|z_a, v_2), V(w|z_a, v_1)V(w|z_a\right.}{2\left[\left(\mathbb{E}(y|z_a, v_1) - \mathbb{E}(y|z_a, v_2)\right)\left(\mathbb{E}(x|z_a, v_1) - \mathbb{E}(x|z_a, v_2)\right)\right]^2(1}$$

$$\Omega_M = \frac{f_M\left(\mathbb{E}(w|z_a, v_1), \mathbb{E}(w|z_a, v_2), V(w|z_a, v_1)V(w|z\right.}{2\left[\left(\mathbb{E}(y|z_a, v_1) - \mathbb{E}(y|z_a, v_2)\right)\left(\mathbb{E}(x|z_a, v_1) - \mathbb{E}(x|z_a, v_2)\right)\right]^2(1}$$

use in a constructive way $f_1(\cdot)$, $f_0(\cdot)$ and $f_M(\cdot)$ and where $V(w|z_a, v_k)$ identifies the vector's conditional variance-covariance matrix *(x, y, xy).*

The "Delta" approach, described for example by van der Vaart, is used to establish the theorem (1998). For their own sake, the asymptotic variances' denominator terms are interesting because they provide clarity on the connection the weak convergence result.

Estimating the average marginal effect is up next. Keep in mind that the marginal effect (conditional on z) can be consistently estimated by

$$\frac{\hat{g}(1, z) - \hat{g}(0, z)}{f(z|v_a)\mathbb{P}(v = v_a)} = \frac{\hat{\mathbb{E}}(y|z, v_1) - \hat{\mathbb{E}}(y|z, v_2)}{\hat{\mathbb{E}}(x|z, v_1) - \hat{\mathbb{E}}(x|z, v_2)} h_0\left(\hat{\mathbb{E}}(w|z, v_1), \hat{\mathbb{E}}(w\right.$$

In this article, we describe how to estimate the marginal impact by taking the mean over (a constant) z-support. The average marginal impact is what we're after in this subsection.

$$\beta_m = \mathbb{E}\left(l(z)\left(g(1,z) - g(0,z)\right)\right)$$

where *l(z)* is a fixed-function trimmer One reliable method for estimating this value is

$$\hat{\beta}_m \frac{1}{n}\sum_{i=1}^{n} l(z_i)\left(\hat{g}(1,z_i) - \hat{g}(0,z_i)\right)$$

together with a normal distribution as a function of $\sqrt{n}$ We leave out the specifics, but the rate can be obtained by checking the conditions for Theorem in Newey and McFadden (1994).

PARAMETRIC SPECIFICATIONS OF $g(x^*,z)$

When the regression function has a parametric specification, this model is a particularly useful special case. In this article, we focus on the binary choice model under the misclassification provided by (13) In this situation, you may continue with estimating in at least two different ways. First, we have a minimal distance estimator similar to Newey's (1994a) Example 2, and second,

Equation (13) $\mathbb{P}(y, x|z, v)$ provide the basis of $\mathbb{P}(y, x|z, v) =$
$\sum_{x^* \in \{0,1\}} \mathbb{P}\left(y(|x^*, z\right) \mathbb{P}\left(x(|x^*, z\right) \mathbb{P}\left(x^*|z, v\right)$

Where

$$\mathbb{P}\left(y(|x^*, z\right) = \mathcal{F}(\beta_1 + \beta_2 x^* + \beta_3 z)^y \left(1 - \mathcal{F}(\beta_1 + \beta_2 x^* + \beta_3 z)\right)^{1-y}$$

Misclassification probabilities are again not assumed to follow any particular functional shape $\mathbb{P}\left(x|x^*, z\right)$

and the likelihood $\mathbb{P}\left(x^*|z, v\right)$. Parameterizing these probabilities with the log-odds ratio ensures that our approximations will be between zero and one. Let $\lambda_k(\cdot)$ for k=1,2,3 indicate the probability as a logarithm $\mathbb{P}\left(x|x^*, z\right)$

$\mathbb{P}\left(x|x^* = 0, z\right)$, $\mathbb{P}\left(x|x^* = 1, z\right)$ and $\mathbb{P}\left(x^*|z, v\right)$ indicate the probability as a logarithm $L(x) = \frac{\exp(x)}{1+\exp(x)}$ Probabilities may be expressed in writing.

$$\mathbb{P}\left(x|x^* = 0, z\right) = \left(L(\lambda_1(z))\right)^x \left(1 - L(\lambda_1(z))\right)^{1-x}$$

$$\mathbb{P}\left(x|x^* = 1, z\right) = \left(L(\lambda_2(z))\right)^{1-x} \left(1 - L(\lambda_2(z))\right)^x$$

$$\mathbb{P}(x^*|z, v) = L(\lambda_3(z, v))$$

Above, we see that the likelihood (23) is a function of the parameters α = (β, λ) = (β, λ1 (·), λ2 (·), λ3 (·))

Moreover, we signify it with $\mathbb{p}$ *(y, x|z, v; α)*. In this context, the parameter is properly classified as a spatial $\mathcal{A} = B \times \Lambda$ where B defines a small set of $\mathbb{R}^{d_z+2}$ The space $\Lambda = \Pi_{k=1}^{3}\Lambda_k$ *where* Λ_1 and Λ_2 are collections of functions that are formally specified on the z and Λ_3 functions defined on top of the backing of *(z, v)* and satisfy $L(\lambda_1(z)) + L(\lambda_2(z)) < 1$ for any $\lambda_1 \in \Lambda_1$ *and* $\lambda_2 \in \Lambda_2$

With the goal of characterizing the sieve approximation of the spaces, Λ_j $j = 1, 2, 3,$ let $\{r_{j,m}\}_{m=1}^{\infty}$ to stand in for a collection of basic functions (like power, Fourier series, or splines) in Λ_j. Let $R_{j,k_n}(x) = (r_1(x), \ldots, r_{k_n}(x))$ denote a $k_n \times 1$ vector of fundamental operations and $\Pi_{j,n}$ conforming constants vector. Afterward, we establish a $\Lambda_{j,n} = \{R_{j,k_n}\Pi_{j,n} : |\Pi_{j,n}| \le c_n\}$ conforming constants vector. Afterward, we establish a *cn* → ∞ also, the sifting area $\mathcal{A}_n = \mathbf{B} \times \Lambda_n = \mathbf{B} \times \Lambda_{1,n} \times \Lambda_{2,n} \times \Lambda_{3,n.}$ See Mahajan (2004) and *Ai* and Chen (2007) for more information on the sieve's construction and the sequence of basic functions.

In order to estimate the infinite dimensional parameters of, the method of sieves is applied in a semi-parametric maximum likelihood framework.

The logarithm of the probability can be written as

$$\begin{aligned} l(w,\alpha) &= \ln \mathbb{P}(y, x|z, v; \alpha) \\ &= \ln\{(1 - \mathcal{F}(b_1 + b_3 z))^{1-y} \mathcal{F}(b_1 + b_3 z)^y \lambda_1(z)^x (1 - \lambda_1(z))^{1-x} (1 - \lambda_3(z, v)) \\ &\quad + (1 - \mathcal{F}(b_1 + b_2 + b_3 z))^{1-y} \mathcal{F}(b_1 + b_2 + b_3 z)^y \lambda_2(z)^{1-x} (1 - \lambda_2(z))^x \lambda_3(z, v)\} \end{aligned}$$

where w = (y, x, z, v) along with the parameter $\alpha \in \mathcal{A} = \mathbf{B} \times \Lambda.$ The Sieve Maximum Likelihood estimator is defined when we get data from a random sample on w:

$$\hat{\alpha} \equiv \left(\hat{b}, \hat{\lambda}\right) = \arg\max_{\mathcal{A}_n} \frac{1}{n}\sum_{i=1}^{n} l(w_i, \alpha)$$

So $\hat{\alpha}_n$ optimizing the sample log likelihood over the finite'sieve' space an is possible with commonly available software. According to Mahajan (2004), the proposed estimator is consistent, converges quickly, and is asymptotically normal $\hat{b}$ the semiparametric efficiency bound is met, and this is demonstrated.

It is also possible to estimate the parameters by building an estimator similar to Example 2 in Newey (1994a), with the caveat that the equation

$$\mathcal{F}^{-1}\left(g\left(0,z\right)\right)=\beta_1+\beta_3 z$$

can be implemented into an estimate of (β_1, β_3) in such a way that the gap between the left and right sides is minimized while keeping all other factors constant. Consequently, a straightforward substitute estimate for (β_1, β_2) is calculated using a least-squares-regression of $\mathcal{F}^{-1}\left(\hat{g}\left(0,z_i\right)\right)$ on z_i:

$$\left(\hat{\beta}_1,\hat{\beta}_3\right)=\arg\min_{(b_1,b_3)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_i\left(\mathcal{F}^{-1}\left(\hat{g}\left(0,z_i\right)\right)-b_1-b_3 z_i\right)^2$$

Where $\mathbb{I}_i$ has a predetermined number of cuts, and $\hat{g}\left(0,z_i\right)$) in which the non-parametric estimator is specified (19). β2 be approximated by $\mathcal{F}^{-1}\left(\hat{g}\left(1,z_i\right)\right)$ plus the projections $\left(\hat{\beta}_1,\hat{\beta}_3\right)$ which we've already acquired above.

TESTING FOR MISCLASSIFICATION

It's only logical to wonder whether there's a way to detect misclassification, seeing as how without it, estimate techniques may be simplified. While this work doesn't go into detail on how to test hypotheses in such models, we will look at how to develop a basic exclusion limit to check for misclassification. In particular, if there is no misclassification, then the expectation of the result conditional on *(x, z, v)* does not rely upon the ILV v for model (1), as mentioned above in the section on identification

Lemma 3 Take into account the model (1) based on the given (1–5) $\mathbb{P}\left(x^*=1|z_a,v\right)\in$ (0, 1). Then, $\eta_0\left(z_a\right)=\eta_1\left(z_a\right)=0$ *if and only if* $\mathbb{E}\left(y|x,z_a,v\right)=\mathbb{E}\left(y|x,z_a\right)$ *a.e.*

Proof. Simple inspection of the probability shape is at the heart of the proof $\mathbb{P}\left(x^*|x,z,v\right)$ assumed to be the case (1)-(5).

$$\begin{aligned}\mathbb{E}\left(y|x,z_a,v\right)&=\sum_{s\in\{0,1\}}g\left(s,z_a\right)\mathbb{P}\left(x^*=s|x,z_a,v\right)\\&=\sum_{s\in\{0,1\}}g\left(s,z_a\right)\frac{\mathbb{P}\left(x|x^*=s,z_a\right)\mathbb{P}\left(x^*=s|z_a,v\right)}{\mathbb{P}\left(x|z_a,v\right)}\\&=\{g\left(0,z_a\right)\left(\eta_0\left(z_a\right)x+\left(1-\eta_0\left(z_a\right)\right)\left(1-x\right)\right)\mathbb{P}\left(x^*=0|z_a,v\right)+\\&\quad g\left(1,z_a\right)\left(\eta_1\left(z_a\right)\left(1-x\right)+\left(1-\eta_1\left(z_a\right)\right)x\right)\mathbb{P}\left(x^*=1|z_a,v\right)\}\mathbb{P}\left(x|z_a,v\right)^{-1}\end{aligned}$$

in which Bayes' Rule led to a second equality (3). We begin by demonstrating the beginning symbol. Let's say ⇒ that

$$\mathbb{P}\left(x=1|z_a,v_2\right)=\eta_0\left(z_a\right)+\left(1-\eta_0\left(z_a\right)-\eta_1\left(z_a\right)\right)$$

$\eta_s\left(z_a\right)=0$ for $s\in\{0,1\}$, then $\mathbb{P}\left(x^*=1|z_a,v\right)=\mathbb{P}\left(x=1|z_a,v\right)$ and we get down to this as the conditional expectation

$$\begin{aligned}\mathbb{E}\left(y|x,z_a,v\right)&=\{g\left(0,z_a\right)\left(1-x\right)\mathbb{P}\left(x=1|z_a,v\right)+g\left(1,z_a\right)x\mathbb{P}\left(x=1|z_a,v\right)\}\mathbb{P}\left(x|z_a,v\right)^{-1}\\&=g\left(x,z_a\right)\end{aligned}$$

If you want to prove the converse, just keep in mind that if your conditional expectations are equal, then your actual outcomes will be equal as well.

$$\mathbb{E}\left(y|x,z_a,v_1\right)=\mathbb{E}\left(y|x,z_a,v_2\right)$$

which, after a bit of algebra, leads to the conclusion that

$$\left(g\left(1,z_a\right)-g\left(0,z_a\right)\right)\left(\mathbb{P}\left(x^*=1|x,z_a,v_1\right)-\mathbb{P}\left(x^*=1|x,z_a,v_2\right)\right)=0$$

The foregoing deduces, on the basis of, that

$$\mathbb{P}\left(x^*=1|x,z_a,v_1\right)-\mathbb{P}\left(x^*=1|x,z_a,v_2\right)=0$$

, which can also be expressed as

$$\left(\eta_1\left(z_a\right)\left(1-x\right)+\left(1-\eta_1\left(z_a\right)\right)x\right)\left\{\frac{\mathbb{P}\left(x^*=1|z_a,v_1\right)}{\mathbb{P}\left(x|z_a,v_1\right)}-\frac{\mathbb{P}\left(x^*=1|z_a,v_2\right)}{\mathbb{P}\left(x|z_a,v_2\right)}\right\}=0$$

The preceding expression simplifies to when *x* = *1.*

$$\left(1-\eta_1\left(z_a\right)\right)\left\{\frac{\mathbb{P}\left(x^*=1|z_a,v_1\right)}{\mathbb{P}\left(x=1|z_a,v_1\right)}-\frac{\mathbb{P}\left(x^*=1|z_a,v_2\right)}{\mathbb{P}\left(x=1|z_a,v_2\right)}\right\}=0$$

and the above

$$\frac{\mathbb{P}\left(x^*=1|z_a,v_1\right)}{\mathbb{P}\left(x=1|z_a,v_1\right)}-\frac{\mathbb{P}\left(x^*=1|z_a,v_2\right)}{\mathbb{P}\left(x=1|z_a,v_2\right)}=0$$

and based on the realization thatresult in our obtaining

$$\eta_0\left(z_a\right)\left(\frac{\mathbb{P}\left(x^*=1|z_a,v_1\right)-\mathbb{P}\left(x^*=1|z_a,v_2\right)}{\mathbb{P}\left(x^*=1|z_a,v_1\right)\mathbb{P}\left(x^*=1|z_a,v_2\right)}\right)=0$$

to the extent that holds, it follows that η_0 *(z_a) = 0*. An analogous defense for the situation where *x = 1* draws the conclusion that η_1 *(z_a) = 0* hence, the lack of measurement error may be inferred from the equality of conditional expectations.

Therefore, under the still-held a misclassification test can be based on a comparison of the conditional expectations for A rejection of the null hypothesis can also be interpreted as evidence against the identifying so it's important to keep those in mind.

Lemma 4 If we assume (1)– (5) and then apply the model (1), we get $\mathbb{P}\left(x^*=1|z_a,v\right)\in$ (0, 1). Let's assume there is no misclassification, so that η_0 *(za) =* η_1 *(z_a) = 0. For* x_a *∊ {0, 1} and va ∊ {*v_1*,* v_2*}* Provide an explanation for the statistical measure.

$$T(x_a, z_a, v_a) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{z_i - z_a}{h_n}\right) \mathbb{I}(x_i = x_a, v_i = v_a)}{\sum_{i=1}^{n} K\left(\frac{z_i - z_a}{h_n}\right) \mathbb{I}(x_i = x_a, v_i = v_a)} - \frac{\sum_{i=1}^{n} y_i K\left(\frac{z_i - z_a}{h_n}\right) \mathbb{I}(x_i = x_a)}{\sum_{i=1}^{n} K\left(\frac{z_i - z_a}{h_n}\right) \mathbb{I}(x_i = x_a)}$$

and let's pretend that 15-19 hold true with r = *x* and *r* = *(x, v).* Then,

$$\sqrt{nh^d}\, T(x_a, z_a, v_a) \Rightarrow \mathcal{N}(0, V_T^a)$$

Where

$$\begin{aligned} V_T^a \;=\; & \left(\mathbb{P}(x = x_a)\,\mathbb{P}(x = x_a, v = v_a)\, f(z_a|x_a, v_a)\, f(z_a|x_a)\right)^{-1} \int K(u)^2\, du \\ & \left\{ \begin{array}{c} [f(z_a|x_a)\,\mathbb{P}(x = x_a) - 2\mathbb{P}(x = x_a, v = v_a)]\, f(z_a|x_a, v_a)\, Var(y|z_a, x_a, v_a) \\ + [\mathbb{P}(x = x_a, v = v_a)\, f(z_a|x_a, v_a)]\, Var(y|z_a, x_a) \end{array} \right\} \end{aligned}$$

Theorem in Berens is the starting point for the proof (1987). The test may be easily implemented by use conventional kernel regression (with perhaps the use of the bootstrap to calculate standard errors). The suggested test often simplifies to testing for an exclusion constraint when considered in the context of other, parametric models. As an example, the following result (13) suggests that a direct test for the exclusion of *v* in the binary choice model may function as a test for misclassification. This test is easily implementable, since it may be conducted using any of the typical test statistics for testing such hypotheses in a maximum likelihood setting.

CONCLUSION

In statistics, identifiability is a necessary condition for inferential precision. If, given an unlimited number of observations, it is feasible to determine the real values of the model's underlying parameters, then we say that the model is identifiable. Parametric model identification follows naturally from the study of model and identification, parametric models, parametric specifications of g (x*, z), and testing for misclassification. It's reasonable to question whether misclassification can be detected, because without it, estimation procedures may be simplified. This paradigm has a special use when the regression function is parametric. The regression function for models including misclassified regressors

REFERENCES

1. Matth¨aus kleindessner (2017) “lens depth function and k-relative neighborhood graph: versatile tools for ordinal data analysis” journal of machine learning research 18 (2017) 1-52 submitted 2/16; revised 6/17; published 7/17

2. Daniel fernandez (2019) “a method for ordinal outcomes: the ordered stereotype model” received: 15 february 2019 revised: 23 may 2019 accepted: 6 july 2019 doi: 10.1002/mpr.1801

3. Yiqun zhang (2020) “an ordinal data clustering algorithm with automated distance learning” the thirty-fourth aaai conference on artificial intelligence (aaai-20)

4. Christophe biernacki, julien jacques, c. Keribin. A survey on model-based co-clustering: high dimension and estimation challenges. 2022. Ffhal-03769727f

5. J. K. Brittany t. Fasy fabrizio lecci, clement maria vincent rouvreau. The included gudhi is authored by clement maria dionysus by dmitriy morozov, michael kerber phat by ulrich bauer, and jan reininghaus. (2015). “tda: statistical tools for topological data analysis”. In: web site https://cran.r-project.org/web/ packages/tda/index.html.

6. Johnson, r. A. And d. W. Wichern (2013). “Applied multivariate statistical analysis: pearson new international edition”. In: pearson. Kroese, d. P. And j. C.c. Chan (2014). “Statistical modeling and computation”. In: new york, ny: springer new york.

7. Laboratory, s. U. C. G. (2015). “The stanford 3d scanning repository”. In: web site http://graphics.stanford.edu/data/3dscanrep/ bunny.

8. Müllner, d. (2013). “fastcluster: fast hierarchical, agglomerative clustering routines for r and python,” in: 2013, vol. 53, p. 18, 2013-05- 29 2013.

9. P. Y. Lum g. Singh, a. Lehman t. Ishkanov m. Vejdemo-johansson m. Alagappan et al. (2013). “Extracting insights from the shape of complex data using topology”. In: scientific reports, vol. 3.

10. Muller, p., quintana, f., jara, a., & hanson, t. (2015). Bayesian nonparametric data ¨ analysis. Springer series in statistics (. Pledger, s. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. Biometrics, 56, 434-442.

11. Pledger, s., & arnold, r. (2014). Clustering, scaling and correspondence analysis: unified pattern-detection models using mixtures. Computational statistics and data analysis, 71, 241-261.

12. Spiegelhalter, d. J., best, n. G., carlin, b. P., & linde, a. (2014). The deviance information criterion: 12 years on. Journal of the royal statistical society: series b (statistical methodology), 76(3), 485–493.

13. Gao, ruochu, "statistical analysis of correlated ordinal data: application to cluster

randomization trials" (2013). Electronic thesis and dissertation repository. 1696.

14. Fernández, daniel & arnold, richard. (2016). Model selection for mixture-based clustering for ordinal data. Australian & new zealand journal of statistics. 58. 10.1111/anzs.12179.

15. Haiyan Liu and Zhiyong Zhang, (2017), Logistic Regression with Misclassification in Binary Outcome Variables: Method and Software, Department of Psychology, University of Notre Dame; Zhiyong Zhang, Department of Psychology, University of Notre Dame

16. Kent Riggs, Dean Young & James Stamey (2011) Interval estimation for misclassification rate parameters in a complementary Poisson model, Journal of Statistical Computation and Simulation, 81:9, 1145-1156, DOI: 10.1080/00949651003762063

Corresponding Author

Karwanje Diwakar Prabhakarrao*

Ph.d Scholar, Kalinga University, Raipur