Detecting latent variable non-normality through the generalized Hausman test

Lucia Guastadisegni[1], Irini Moustaki[2], Vassilis Vasdekis[3], and Silvia Cagnone[4]

[1] University of Bologna,
lucia.guastadisegni2@unibo.it
[2] London School of Economics and Political Science,
i.moustaki@lse.ac.uk
[3] Athens University of Economics and Business,
vasdekis@aueb.gr
[4] University of Bologna,
silvia.cagnone@unibo.it

Abstract. This paper extends the generalized Hausman test to detect non-normality of the latent variable distribution in unidimensional IRT models for binary data. To build the test, we consider the estimator obtained from the two-parameter IRT model, that assumes normality of the latent variable, and the estimator obtained under a semi-nonparametric framework, that allows for a more flexible latent variable distribution. The behaviour of the test is evaluated through a simulation study. The results highlight the good performance of the test in terms of both Type I error rates and power with many items and large sample sizes.

Keywords: generalized Hausman test, SNP-IRT model, binary data

1 Introduction

One of the typical assumptions of latent variable models is the normal distribution of the latent variables. As shown in Ma and Genton (2010), this assumption is not always appropriate and misspecifying the form of the latent variable by assuming normality can result in large biases in parameter estimates. Several methods, that assume a different form for the latent variable, have been proposed within the generalized latent variable models (GLLVM) and Item Response Theory (IRT) framework. Some examples are the semi-parametric (Ma and Genton, 2010), the empirical histogram (Knott and Tzamourani, 2007), the Ramsey-curve (Woods, 2006) and the semi-nonparametric (SNP) (Gallant and Nychka, 1987, Woods and Lin, 2009, Irincheeva et al., 2012) methods.

Commonly information criteria are used to choose between a model where the latent variables are normal and a model where they have a more complex shape (Woods and Lin, 2009, Irincheeva et al., 2012). However, detecting non-normality of the latent variables through a statistical test remains an open issue.

Hausman (1978) proposes a specification test to detect failure of the orthogonality assumption in the regression model. The Hausman test can be applied also

in other contexts, to detect different types of model misspecification. The idea of the test is simple. It compares two different estimators that are consistent when the model is correctly specified and one is also efficient. In presence of model misspecification, only the inefficient estimator is consistent. The efficiency assumption simplifies the computation of the covariance matrix of the difference between the two estimators. However, this matrix can fail to be positive definite under model misspecification or in presence of small sample sizes. A generalized version of the Hausman (GH) test has been proposed by White (1982). In this case none of the estimators that result from different models need to be efficient and the covariance matrix involved in the test is always positive definite.

As far as we know, in the IRT context the classic Hausman test has been used only by Ranger and Much (2020) to detect misspecification of the item characteristic functions and local dependencies among items. In generalized linear mixed models (GLMM) for clustered data, a robust version of the Hausman test, similar to the one by White (1982), has been proposed by Bartolucci et al. (2017) when a discrete distribution for the random effects is assumed.

The objective of this work is to extend the GH test to detect non-normality of the latent variable distribution in unidimensional IRT models for binary data. To build the test, we consider the estimators resulting from two different models and estimation methods. The first model is the classical unidimensional IRT model for binary data based on the normality assumption of the latent variable, where we estimate the parameters using a maximum pairwise likelihood (PL) method. The PL method uses information from bivariate-order margins and belongs to the family of composite likelihood methods (Lindsay, 1988, Varin, 2008). It produces biased parameter estimates when the latent variable is not normally distributed. The second model is the unidimensional SNP-IRT model for binary data (Woods and Lin, 2009, Irincheeva et al., 2012), and we estimate the parameters using the quasi-maximum likelihood (ML) method. The choice of these estimators for the two models is motivated by the following reasons. First, both methods are consistent when the latent variable is normally distributed. Moreover, the quasi-ML method for the SNP_L model is consistent also under different distribution assumptions of the latent variable (Gallant and Tauchen, 1989, Irincheeva et al., 2012). These conditions on the consistency of the parameter estimators are required to correctly apply the Generalized Hausman test (White, 1982). Second, the maximum PL estimator is less efficient than the ML estimator. This implies that, also under normality of the latent variable distribution, the covariance matrix of the difference of the two estimators involved in the GH test is different from zero. This allows us to avoid numerical problems in the computation of the test.

The article is organized as follows. First, we review the classical and SNP-IRT model for binary data. Second, we introduce the GH test to detect non-normality of the latent variable distribution. Next, we present a Monte Carlo simulation study. Finally, we present some concluding remarks.

2 The classical and SNP-IRT model for binary data

Let us denote by $y_1, ..., y_p$ a set of observed binary variables/items, by n the number of individuals and by z the latent variable with density function $h(z)$.

For the classical IRT model, the response category probability for the i-th individual to the j-th item is modelled using a logistic model (measurement model)

$$P(y_{ij} = 1|z_i) = \pi_{ij}(z_i) = \frac{\exp(\alpha_{0j} + \alpha_{1j} z_i)}{1 + \exp(\alpha_{0j} + \alpha_{1j} z_i)}, \tag{1}$$

where α_{0j} is the item intercept and α_{1j} the item slope. In this model $h(z) = \phi(z)$, where $\phi(z)$ is the density of a standard normal.

For the SNP-IRT model, the response probability is the same as (1), where the latent variable has a SNP parametrization

$$h(z_i) = P_L^2(z_i)\phi(z_i) \qquad P_L(z_i) = \sum_{0 \le l \le L} a_i z_i^l, \tag{2}$$

$a_0, ..., a_L$ are the real coefficients of the polynomial $P_L(z_i)$ and L is the polynomial degree.

In order for $h(z)$ to be a density, the coefficients $a_0, ..., a_L$ of $P_L(z)$ should be chosen such that $\int h(z)dz = 1$. For this purpose, Gallant and Tauchen (1989) use a proportionality constant $1/\int P_L(z)^2\phi(z)dz$ and fix the constant term of the polynomial equal to 1. Alternatively, Irincheeva et al. (2012) and Woods and Lin (2009) use the parametrization proposed by Zhang and Davidian (2001), that imposes

$$1 = \int_R P_L^2(z)\phi(z)dz = E\{P_L^2(w)\} = a' E(\tilde{w}\tilde{w}')a = a' A a \tag{3}$$

with $w \sim N(0,1)$, $P_L(w) = a'\tilde{w}$, $\tilde{w} = (1, w, w^2, ..., w^L)$. The matrix A is positive definite by definition and $A = B'B$, where B is a positive definite matrix.

If $c = Ba$, equation (3) becomes $c'c = 1$ and $c = (c_1, ..., c_{L+1})'$. The elements of c can be represented using a polar coordinate transformation as $c_1 = \sin\varphi_1, c_2 = \cos\varphi_1 \sin\varphi_2, ..., c_L = \cos\varphi_1 \times \cos\varphi_{L-1} \sin\varphi_L, c_{L+1} = \cos\varphi_1 \cos\varphi_2 \times \cos\varphi_{L-1} \cos\varphi_L$, with angles $-\pi/2 < \varphi_t \le \pi/2$, $t = 1, ..., L$. The density of the latent variable in (2) can be expressed as

$$h(z|\boldsymbol{\varphi}, L) = (a'\tilde{\mathbf{z}})^2\phi(z), \tag{4}$$

where a can be obtained from c as $a = B^{-1}c$, $\tilde{\mathbf{z}} = (1, z, z^2, ..., z^L)'$ and $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_L)'$.

When $L = 1$, $P_L(z) = a_0 + a_1 z$, $a_0 = \sin\varphi_1$, $a_1 = \cos\varphi_1$. When $L = 0$ the distribution of the latent variable reduces to the normal one. In the following sections we indicate with SNP_1 the model for $L = 1$ and with SNP_0 the model for $L = 0$.

2.1 Pairwise estimator for the SNP_0 model

To implement the GH test, the parameters of the SNP_0 model are estimated with the pairwise method. The pairwise log-likelihood of the data, based on the bivariate

marginal densities $f(y_{ij}, y_{ik}, \boldsymbol{\theta})$, $j,k = 1,..p$ and $k > j$, is

$$pl_{SNP_0}(\mathbf{y},\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k>j} \ln f(y_{ij}, y_{ik}, \boldsymbol{\theta}) = \\ = \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k>j} \ln \int \Big[\pi_{ij}(z_i)^{y_{ij}}(1-\pi_{ij}(z_i))^{1-y_{ij}}\Big]\Big[\pi_{ik}(z_i)^{y_{ik}}(1-\pi_{ik}(z_i))^{1-y_{ik}}\Big]\phi(z_i)dz_i. \tag{5}$$

The pairwise log-likelihood is maximized with respect to $\boldsymbol{\theta}$, that includes the item intercepts and slopes. Under correct model specification, the maximum PL estimator $\tilde{\boldsymbol{\theta}}$ converges in probability to the true parameter value $\boldsymbol{\theta}_0$ and

$$\tilde{\boldsymbol{\theta}} \xrightarrow{p} N(\boldsymbol{\theta}_0, A^{-1}(\boldsymbol{\theta}_0)B(\boldsymbol{\theta}_0)A^{-1}(\boldsymbol{\theta}_0)), \tag{6}$$

where $A(\boldsymbol{\theta}) = E_y\Big[-\frac{\partial^2 pl_{SNP_0}(\mathbf{y},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big]$, $B = var\Big[\frac{\partial pl_{SNP_0}(\mathbf{y},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big]$ and $A(\boldsymbol{\theta}) \neq B(\boldsymbol{\theta})$ (Lindsay, 1988, Varin, 2008). These matrices can be estimated by their observed versions as

$$\hat{A}(\boldsymbol{\theta}) = -\sum_{i=1}^{n}\frac{\partial^2 pl_{SNP_0}(\mathbf{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \tag{7}$$

and

$$\hat{B}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\frac{\partial pl_{SNP_0}(\boldsymbol{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial pl_{SNP_0}(\boldsymbol{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}. \tag{8}$$

2.2 Quasi-ML estimator for the SNP_L model

The parameters of the SNP_L model, $L > 0$, are estimated with the quasi-ML method. The log-likelihood of the data is

$$l_{SNP_L}(\mathbf{y},\boldsymbol{\theta}) = \sum_{i=1}^{n}\ln f(\mathbf{y}_i,\boldsymbol{\theta}) = \\ = \sum_{i=1}^{n}\ln\int\prod_{j=1}^{p}\pi_{ij}(z_i)^{y_{ij}}(1-\pi_{ij}(z_i))^{1-y_{ij}}P_L^2(z_i)\exp\Big(-\frac{1}{2}z_i'z_i\Big)dz_i. \tag{9}$$

The integral in the log-likelihood $l(\mathbf{y},\boldsymbol{\theta})$ is approximated with the Gauss-Hermite quadrature, as in Woods and Lin (2009). The degree of the polynomial L is fixed and is not estimated by maximum likelihood. The log-likelihood function is maximized with respect to the unknown vector of parameter $\boldsymbol{\theta} = (\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\varphi})$ as follows

$$(\dot{\boldsymbol{\alpha}}_0, \dot{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\varphi}}) = argmax_{\boldsymbol{\theta}}\, l_{SNP_L}(\mathbf{y},\boldsymbol{\theta}). \tag{10}$$

For identifiability reasons, the item intercepts and slopes, that correspond to a latent variable that has mean 0 and variance 1, are rescaled as (Irincheeva et al., 2012)

$$\hat{\alpha}_{0j} = \dot{\alpha_{0j}} + \dot{\alpha_{1j}}\tilde{E}(Z) \qquad j = 1,...,p \tag{11}$$

$$\hat{\alpha}_{1j} = \alpha_{1j}^{\cdot}\sqrt{\tilde{V}(Z)} \qquad j = 1, ..., p, \tag{12}$$

where $\tilde{E}(Z)$ and $\tilde{V}(Z)$ are found given $\hat{\boldsymbol{\varphi}}$ and the SNP density of z. The final quasi-ML estimator is $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_0, \hat{\boldsymbol{\varphi}})$. Under normal, multi-modal and asymmetric distributions of the latent variables and if the regularity conditions A2-A6 of White (1982) are satisfied,

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} N(\boldsymbol{\theta}_{0*}, A^{-1}(\boldsymbol{\theta}_{0*})B(\boldsymbol{\theta}_{0*})A^{-1}(\boldsymbol{\theta}_{0*})), \tag{13}$$

where $\boldsymbol{\theta}'_{0*} = (\boldsymbol{\alpha}'_{00}, \boldsymbol{\alpha}'_{01}, \boldsymbol{\varphi}_*{}')$. $\boldsymbol{\alpha}_{00}$ and $\boldsymbol{\alpha}_{01}$ are the true parameter values for the item intercepts and slopes while $\boldsymbol{\varphi}_*$ is the value of $\boldsymbol{\varphi}$ that minimizes the Kullback-Leibler information criterion (White, 1982, Gallant and Tauchen, 1989, Irincheeva et al., 2012). $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$ are the expected Hessian and cross-product matrices, respectively. Their observed versions can be computed with the Delta method (Cramér, 1946) and are defined similarly to (7) and (8), where $pl_{SNP_0}(\mathbf{y}_i, \boldsymbol{\theta})$ is replaced by $l_{SNP_L}(\mathbf{y}_i, \boldsymbol{\theta})$.

3 The Generalized Hausman Test

In this section we present the GH test, derived by White (1982), applied to detect non-normality of the latent variable using the SNP-IRT model.

Let's denote by $\boldsymbol{\eta}$ the sub-vector of $\boldsymbol{\theta}' = (\boldsymbol{\alpha}'_0, \boldsymbol{\alpha}'_1, \boldsymbol{\varphi}')$ that includes the item intercepts $\boldsymbol{\alpha}_0$ and slopes $\boldsymbol{\alpha}_1$. $\boldsymbol{\eta}$ has dimension $2p \times 1$, where p is the number of items.

Consider the maximum PL estimator $\tilde{\boldsymbol{\theta}}_{SNP_0} = \tilde{\boldsymbol{\eta}}_{SNP_0}$ of a classic IRT model where the latent variable is normally distributed, that is the SNP_0 model.

Consider the quasi-ML estimator $\hat{\boldsymbol{\theta}}'_{SNP_L} = (\hat{\boldsymbol{\eta}}'_{SNP_L}, \hat{\boldsymbol{\varphi}}')$ of a SNP-IRT model with $L > 0$, where the sub-vector of parameter $\hat{\boldsymbol{\varphi}}$ has dimension $L \times 1$ and so $\hat{\boldsymbol{\theta}}_{SNP_L}$ has dimension $(2p+L) \times 1$. Following White (1982), under normality of the latent variable

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_{SNP_L} - \tilde{\boldsymbol{\eta}}_{SNP_0}) \xrightarrow{d} N(0, S(\boldsymbol{\eta}_0, \boldsymbol{\theta}_{0*})). \tag{14}$$

An estimator of $S(\boldsymbol{\eta}_0, \boldsymbol{\theta}_{0*})$ is

$$\begin{aligned}\hat{S}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L}) = \hat{A}^{\boldsymbol{\eta\varphi}}(\hat{\boldsymbol{\theta}}_{SNP_L})^{-1}\hat{B}(\hat{\boldsymbol{\theta}}_{SNP_L})\hat{A}^{\boldsymbol{\eta\varphi}}(\hat{\boldsymbol{\theta}}_{SNP_L})^{-1'} + \hat{A}(\tilde{\boldsymbol{\eta}}_{SNP_0})^{-1}\hat{B}(\tilde{\boldsymbol{\eta}}_{SNP_0})\hat{A}(\tilde{\boldsymbol{\eta}}_{SNP_0})^{-1'} - \\ - \hat{A}^{\boldsymbol{\eta\varphi}}(\hat{\boldsymbol{\theta}}_{SNP_L})^{-1}\hat{R}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})'\hat{A}(\tilde{\boldsymbol{\eta}}_{SNP_0})^{-1'} - \hat{A}(\tilde{\boldsymbol{\eta}}_{SNP_0})^{-1}\hat{R}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})\hat{A}^{\boldsymbol{\eta\varphi}}(\hat{\boldsymbol{\theta}}_{SNP_L})^{-1'},\end{aligned} \tag{15}$$

where the matrices $\hat{A}(\tilde{\boldsymbol{\eta}}_{SNP_0})$ and $\hat{B}(\tilde{\boldsymbol{\eta}}_{SNP_0})$, defined in formulas (7) and (8), have dimension $2p \times 2p$ and are evaluated at $\tilde{\boldsymbol{\eta}}_{SNP_0}$. $\hat{A}(\hat{\boldsymbol{\theta}}_{SNP_L})$ and $\hat{B}(\hat{\boldsymbol{\theta}}_{SNP_L})$ are the observed Hessian and cross-product matrix of dimension $(2p+L) \times (2p+L)$ for the SNP_L model, evaluated at $\hat{\boldsymbol{\theta}}_{SNP_L}$. The matrix $\hat{A}^{\boldsymbol{\eta\varphi}}(\hat{\boldsymbol{\theta}}_{SNP_L})^{-1}$ is obtained by deleting the last L row from the matrix $\hat{A}(\hat{\boldsymbol{\theta}}_{SNP_L})^{-1}$ and has dimension $2p \times (2p+L)$. The matrix $\hat{R}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})$ has dimension $2p \times (2p+L)$ and can be computed as

$$\hat{R}(\boldsymbol{\eta}_{SNP_0}, \boldsymbol{\theta}_{SNP_L}) = \sum_{i=1}^{n} \frac{\partial pl_{SNP_0}(\boldsymbol{y}_i, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial l_{SNP_L}(\boldsymbol{y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}, \tag{16}$$

where $pl_{SNP_0}(\boldsymbol{y}_i, \boldsymbol{\eta})$ is the pairwise log-likelihood for the individual i under the model SNP_0 and $l_{SNP_L}(\boldsymbol{y}_i, \boldsymbol{\theta})$ is the log-likelihood for the individual i under the model SNP_L.

The matrix in (16) is evaluated at $(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})$. We choose the maximum PL and the quasi-ML estimator for the two models to avoid that, under correct model specification, $\tilde{\boldsymbol{\eta}}_{SNP_0}$ and $\hat{\boldsymbol{\eta}}_{SNP_L}$ converge to the same covariance matrix, producing a $\hat{S}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})$ matrix in (15) with all entries close to 0.

Given the theoretical result in (14), the GH test is given by

$$GH = (\hat{\boldsymbol{\eta}}_{SNP_L} - \tilde{\boldsymbol{\eta}}_{SNP_0})' \hat{S}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})^{-1} (\hat{\boldsymbol{\eta}}_{SNP_L} - \tilde{\boldsymbol{\eta}}_{SNP_0}). \tag{17}$$

Under normality of the latent variable, the GH test is asymptotically distributed as a χ^2_{2p}, where $2p$ are the degrees of freedom, i.e. the number of parameters in $\boldsymbol{\eta}$.

However, the matrix $\hat{S}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})$ is often close to singularity and its inversion in formula (17) is numerically unstable.

Given the theoretical result in (14) and the quadratic form $(\hat{\boldsymbol{\eta}}_{SNP_L} - \tilde{\boldsymbol{\eta}}_{SNP_0})'(\hat{\boldsymbol{\eta}}_{SNP_L} - \tilde{\boldsymbol{\eta}}_{SNP_0})$, we consider the following test statistic (Ranger and Much, 2020)

$$GH_T = (\hat{\boldsymbol{\eta}}_{SNP_L} - \tilde{\boldsymbol{\eta}}_{SNP_0})'(\hat{\boldsymbol{\eta}}_{SNP_L} - \tilde{\boldsymbol{\eta}}_{SNP_0}). \tag{18}$$

Under normality of the latent variable

$$GH_T \sim \sum_{l=1}^{d} \lambda_l z_l^2, \qquad z_l \sim N(0,1), \tag{19}$$

where d is the rank of $S(\boldsymbol{\eta}_0, \boldsymbol{\theta}_{0*})$ and $\lambda_1, ..., \lambda_d$ are its non-zero eigenvalues.

It is possible to approximate the distribution in (19) as follows (Welch, 1938,Yuan and Bentler, 2010)

$$GH_T \sim a\chi^2_b. \tag{20}$$

The quantity a and b are defined as

$$a = \frac{\sum_{l=1}^{d} \lambda_l^2}{\sum_{l=1}^{d} \lambda_l} \tag{21}$$

and

$$b = \frac{(\sum_{l=1}^{d} \lambda_l)^2}{\sum_{l=1}^{d} \lambda_l^2}. \tag{22}$$

Since $S(\boldsymbol{\eta}_0, \boldsymbol{\theta}_{0*})$ can be consistently estimated by $\hat{S}(\tilde{\boldsymbol{\eta}}_{SNP_0}, \hat{\boldsymbol{\theta}}_{SNP_L})$ defined in (15), a and b can be consistently estimated substituting $\hat{\lambda}_1, ..., \hat{\lambda}_d$ in (21) and (22), where d is rank of $\hat{S}$ and $\hat{\lambda}_1, ..., \hat{\lambda}_d$ are its non-zero eigenvalues.

4 Simulation Study

4.1 Simulation design

In this section we study the performance of the GH_T test by a simulation study. The estimation of the SNP-IRT model is computationally expensive. Moreover, as the degree of the polynomial L increases ($L > 1$), the SNP_L model becomes more sensitive

to the choice of the initial values for all model parameters and the estimation results can be less reliable. Furthermore, in the data generating models we assume the latent variable distributed as mixtures of two normals, that can be well approximated with $L = 1$, as highlighted in Irincheeva et al. (2012). Thus, to implement the GH_T test, we consider the SNP_0 and the SNP_1 models. The optimization of the SNP_0 model is obtained with direct maximization using the function "optim" of the software R while, for the SNP_1 model, the function "nlminb", that makes use of the analytically computed gradient and Hessian matrix. For the SNP_1 model, initial values of the parameters $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ are the parameter estimates obtained with the SNP_0 model. In each data replication, for the φ_1 parameter, we sample 10 initial values from a sequence of values equally spaced by 0.1 in the interval $[-\frac{\pi}{2};\frac{\pi}{2}]$, i.e. the domain of φ_1, including the SNP_0 model as a subcase. Among the estimated SNP_1 models in each data replication, we select the one that corresponds to the maximum value of the log-likelihood function. All matrices involved in the GH_T test are computed numerically with the "NumDeriv" R package. Although assuming a SNP distribution for the latent variable is more computationally demanding than assuming the normal distribution, it has the great advantage that it is very flexible and produces accurate estimates in many situations.

We consider the following simulation conditions: number of items ($p = 4, 10, 20$) $\times$ sample size ($n = 500, 1000$) $\times$ test statistic (GH_T). In all the simulation scenarios, $R = 500$ replications are considered and $\alpha = 0.05$. Non-valid statistics, for example negative statistics, are excluded from the analysis. The Type I error rates and power of the GH_T test are computed as $\hat{p} = \sum_{l=1}^{N_v} \frac{I(GH_{T_l} \geq c)}{N_v}$, where N_v is the number of valid statistics out of the number of replications, I is an indicator function, GH_{T_l} is the value of the GH_T test statistic evaluated in the l-th replication. c is the theoretical asymptotic critical value corresponding to the $(1-\alpha)$th percentile of the $a\chi^2_b$ distribution for the GH_T test, where a and b are computed as in (21) and (22). The confidence interval (CI) of each rate $\hat{p}$ is computed as $\hat{p} \pm z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\alpha(1-\alpha)}{N_v}}$.

To evaluate the performance of the GH_T test, we consider three scenarios (SC), corresponding to three different distribution assumptions for the latent variable z in the data generating models. The general model is

$$\begin{aligned} &logit(\pi_{ij}) = \alpha_{0j} + \alpha_{1j} z_i \qquad i = 1, ..., n \qquad j = 1, 2, ..., p \\ &z \sim h(z) \end{aligned} \tag{23}$$

Item intercepts are randomly generated in the interval [-0.8; 1.12] while the item slopes in the interval [0.5; 1.5].

To study the Type I error rates of the GH_T test we consider the following scenario:

A $z \sim N(0,1)$

To study the power of the GH_T test we consider the following two scenarios:

B $z \sim 0.1N(-2, 0.25) + 0.9N(2,1)$,
where z has an overall mean equal to 1.6 and variance equal to 2.365.

C $z \sim 0.7N(-1.5, 0.6) + 0.3N(1.5, 0.5)$,

where z has an overall mean equal to -0.6 and variance equal to 2.217.
Under the distributional assumptions of the two scenarios **B** and **C**, the estimates of the quasi-ML parameters of the SNP_1 model are nearly unbiased (see the results on the bias of the parameters in scenario **B** reported in Irincheeva et al., 2012) while the maximum PL parameter estimates of the SNP_0 model are largely biased with respect to the true parameter values. This should result in a good GH_T test performance in terms of power.

4.2 Results

Table 1 reports the Type I error rates, mean and standard deviation of the theoretical(T) and empirical(E) distribution of the GH_T test for scenario **A**.

Table 1. Type I error rates, mean and standard deviation of the theoretical(T) and empirical(E) distribution of the GH_T test for scenario A, $p = 4, 10, 20$, $n = 500, 1000$

p	n	Distribution	Mean	SD	α
4	500	TD	2.01	2.00	0.050
		ED	1.61	1.81	**0.016**
	1000	TD	2.12	2.06	0.050
		ED	2.54	2.93	**0.086**
10	500	TD	3.44	2.62	0.050
		ED	2.89	2.64	**0.018**
	1000	TD	3.21	2.54	0.050
		ED	3.00	2.97	0.044
20	500	TD	3.48	2.64	0.050
		ED	3.44	3.15	0.056
	1000	TD	3.52	2.65	0.050
		ED	3.63	3.12	0.060

Note 1: Values in boldface indicate that the nominal level α is not included in their confidence interval

Overall, the GH_T test has good performance in terms of Type I error rates when the sample size is large and in general with many items. Moreover, the empirical distribution of the GH_T test approaches the theoretical one as the number of items and the sample size increase. Small differences can be found in terms of empirical and theoretical standard deviations, while the means of the two distributions are very similar under most conditions. Despite the good performance of the test with many items and large sample size in terms of Type I error rates, we observe an inconsistent pattern of results with 4 items and all sample sizes. In general, the estimation of the model parameters and the related information matrices, on which the GH_T test is based, is less accurate on small data sets. Indeed, few items and small sample sizes carry out less information than more items and large sample sizes. We should consider larger sample sizes to obtain Type I error rates of the GH_T test close to the nominal level α for 4 items, while $n = 1000$ is sufficient for 10 items and $n = 500$ for 20 items.

Table 2 presents the power of the GH_T test for scenarios **B** and **C**.

Table 2. Empirical power of the GH_T test for scenarios B and C, $p = 4, 10, 20$, $n = 500, 1000$

SC	p	n	power
B	4	500	0.53
		1000	0.86
	10	500	0.924
		1000	0.998
	20	500	0.99
		1000	0.998
C	4	500	0.796
		1000	0.92
	10	500	1
		1000	1
	20	500	0.986
		1000	1

The power of the GH_T test is high when the sample size is large and with 10 and 20 items. Moreover, it increases with the number of items and the sample size.

5 Conclusion

In this work, we extended the GH test to detect non-normality of the latent variable distribution in unidimensional IRT models for binary data. The GH test was obtained as the difference between the estimators of the classic IRT model for binary data and the SNP-IRT model, that allows for a more flexible shape of the latent variable distribution. To avoid the inversion of the covariance matrix of the difference between the parameter estimates, we considered an alternative form of this test, that we called GH_T test, and we evaluated its performance by means of a small simulation study.

The simulation study highlights that the GH_T test has good performance in terms of Type I error rates with many items and in particular for large sample sizes. For what concerns the power, the GH_T test has good performance with many items and large sample sizes. However, these are preliminary results. Further studies should include other distributions of the latent variables. Indeed, it would be interesting to study the behaviour of this test when the SNP approach performs less well in recovering the distribution of the latent variable, for example when it is skewed (Monroe, 2014). Moreover, the GH_T test presented in this work could be applied to IRT models for polytomous data, assuming the SNP representation of the latent variable distribution. Since these models involve a higher number of parameters, the additional issue, compared to binary data, could be the computational cost of the estimation process (Bartholomew et al., 2011).

The GH test could also be applied to detect other types of model violations, as local dependence or violation of the item characteristic function. In these cases, other

types of estimators consistent under model misspecification should be considered in order to apply the test.

Bibliography

Bartholomew, D. J., Knott, M., and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. Wiley, 3 edition.

Bartolucci, F., Bacci, S., and Pigini, C. (2017). Misspecification test for random effects in generalized linear finite-mixture models for clustered binary and ordered data. *Econometrics and Statistics*, 3:112–131.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2):363–390.

Gallant, A. R. and Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica*, 57(5):1091–1120.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Irincheeva, I., Cantoni, E., and Genton, M. G. (2012). Generalized linear latent variable models with flexible distribution of latent variables. *Scandinavian Journal of Statistics*, 39(4):663–680.

Knott, M. and Tzamourani, P. (2007). Bootstrapping the estimated latent distribution of the two-parameter latent trait model. *British Journal of Mathematical and Statistical Psychology*, 60(1):175–191.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, 80(1):221–239.

Ma, Y. and Genton, M. G. (2010). Explicit estimating equations for semiparametric generalized linear latent variable models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):475–495.

Monroe, S. L. (2014). *Multidimensional item factor analysis with semi-nonparametric latent densities*. PhD thesis, UCLA.

Ranger, J. and Much, S. (2020). Analyzing the fit of IRT models with the Hausman test. *Frontiers in Psychology*, 11.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, 92(1):1–28.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3-4):350–362.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Woods, C. M. (2006). Ramsay-curve item response theory (RC-IRT) to detect and correct for nonnormal latent variables. *Psychological Methods*, 11(3):253–270.

Woods, C. M. and Lin, N. (2009). Item response theory with estimation of the latent density using Davidian curves. *Applied Psychological Measurement*, 33(2):102–117.

Yuan, K.-H. and Bentler, P. M. (2010). Two simple approximations to the distributions of quadratic forms. *British Journal of Mathematical and Statistical Psychology*, 63(2):273–291.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57(3):795–802.