Information matrix tests for multinomial logit models

Dante Amengual[a], Gabriele Fiorentini[b], Enrique Sentana[a]

[a] *CEMFI, Madrid*
[b] *Università di Firenze and RCEA*

Abstract

We obtain simple and intuitive expressions for the information matrix test for the multinomial logit model that resemble a multivariate heteroskedasticity test à la White (1980) applied to the conditionally demeaned value of the outer product of the generalised residuals.

Keywords: Hessian matrix, Outer product of the score, Specification test, Unobserved heterogeneity

JEL Classification: C35, C25

Corresponding author: Enrique Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain. E-mail: `sentana@cemfi.es`

1. Introduction

White's (1982) information matrix (IM) test provides a general procedure for examining the correct specification of models estimated by maximum likelihood (ML). It directly assesses the IM equality, which states that the sum of the Hessian matrix and the outer product of the score vector should be zero in expected value when the estimated model is correctly specified. Chesher (1984) reinterpreted it as a score test against unobserved heterogeneity, a serious concern in microeconometric models as the parameters characterising objective functions or constraints often vary across agents. Not surprisingly, the IM test has been extensively studied for univariate probit and tobit models (see Horowitz (1994) and the references therein).

However, the IM test has not been derived for multinomial logit models. Polytomous choice models specify how the probabilities of mutually exclusive Bernoulli variables that make up a multinomial random variable $\boldsymbol{\xi} = (\xi_1, \dots, \xi_K)'$ of dimension K vary across observations as a function of L observed characteristics $\mathbf{z}$. Typically, they are parametrised as

$$p_k = \Pr(\xi_k = 1|\mathbf{z}) = F_k(\mathbf{z}; \boldsymbol{\beta}) \quad k = 1, ..., K, \tag{1}$$

where $\boldsymbol{\beta}$ is a finite vector of parameters. Since the distribution of $\boldsymbol{\xi}$ is necessarily multinomial, correct specification of (1) is equivalent to correct specification of the functional forms for $F_k(.;\boldsymbol{\beta})$.

There are two main categories of logit-type models for polytomous unordered selection:

1. *Conditional logit models* in which the probabilities depend on the choices' characteristics (for example, travel costs for transportation mode choice), but their effects are invariant across alternatives, so that $\boldsymbol{\beta}_k = \boldsymbol{\beta}\ \forall k$.
2. *Multinomial logit models* in which the probabilities depend on the choosers' characteristics (for example, education, age and gender for occupational choice), which are invariant across choices, while their effects are captured by $\boldsymbol{\beta}_k$'s that vary across alternatives.

We focus on the latter in the text because they are also popular in switching regime models for time series, but we exploit the close relationship between both specifications to explain how our results can be used to obtain the IM matrix test for the former in a supplemental appendix. Thus, we complement Mai, Frejinger and Bastin (2015), who apply the IM test to a variant of the conditional logit model for transportation mode choice originally introduced by McFadden (1974).[1]

The rest of the note is organised as follows. We derive our theoretical results in Section 2 and report the Monte Carlo exercises that look at the finite sample size and power of the test in Section 3. Finally, we present our conclusions and discuss some avenues for further research, relegating proofs, details about our simulations and MATLAB code to supplemental appendices.

2. Theoretical results

Consider the following parametrisation of the conditional probabilities in (1):

$$F_k(\mathbf{z};\boldsymbol{\beta}) = \frac{e^{\boldsymbol{\beta}_k'\mathbf{z}}}{\sum_{\ell=1}^{K} e^{\boldsymbol{\beta}_\ell'\mathbf{z}}}, \quad k = 1,\ldots,K, \tag{2}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_K')'$ is a vector that collects the K coefficient vectors of dimension L each. Naturally, $\sum_{k=1}^{K} p_k(\mathbf{z};\boldsymbol{\beta}) = 1$ for all $\mathbf{z}$ and $\boldsymbol{\beta}$. For identification purposes, we follow the usual practice of setting $\boldsymbol{\beta}_1 = \mathbf{0}$, so that the first category becomes the baseline one, thereby eliminating L elements of the score vector, $\mathbf{s}(\boldsymbol{\beta})$, and $L(L+1)/2$ of the Hessian matrix, $\mathbf{h}(\boldsymbol{\beta})$, without loss of generality because the ordering of the categories is arbitrary. In this respect, Lemma 1 in Amengual, Fiorentini and Sentana (2024a) implies that the IM test is numerically invariant to reparametrisations.

[1]See Chesher and Santos Silva (2002) for another generalisation of the conditional multinomial logit model that explicitly allows for individual heterogeneity in the model parameters.

Let $\mathbf{p}_r(\mathbf{z};\boldsymbol{\beta}) = [p_2(\mathbf{z};\boldsymbol{\beta}),\ldots,p_K(\mathbf{z};\boldsymbol{\beta})]'$ represent the vector of conditional probabilities of the $K-1$ non-normalised categories, and $\mathbf{u}_r(\boldsymbol{\xi}_r,\mathbf{z};\boldsymbol{\beta}) = [u_2(\xi_2,\mathbf{z};\boldsymbol{\beta}),\ldots,u_K(\xi_K,\mathbf{z};\boldsymbol{\beta})]' = \boldsymbol{\xi}_r - \mathbf{p}_r(\mathbf{z};\boldsymbol{\beta})$, with $\boldsymbol{\xi}_r = (\xi_2,\ldots,\xi_K)'$, the corresponding $K-1$ dimensional vector of what Gouriéroux, Monfort, Renault and Trognon (1987) called generalised residuals by analogy to OLS regressions. Finally, let $\hat{\boldsymbol{\beta}}_N = (\mathbf{0}',\hat{\boldsymbol{\beta}}'_{2N},\ldots,\hat{\boldsymbol{\beta}}'_{KN})' = (\mathbf{0}',\hat{\boldsymbol{\beta}}'_{rN})$ denote the ML estimator. Then, we can show that:

Proposition 1. *1) The score vector and Hessian matrix of model (2), which are of dimension $(K-1)L$, are given by*

$$\begin{aligned}\mathbf{s}_r(\boldsymbol{\xi},\mathbf{z};\boldsymbol{\beta}) &= \mathbf{u}_r(\boldsymbol{\xi},\mathbf{z};\boldsymbol{\beta})\otimes\mathbf{z}, &(3)\\ \mathbf{h}_r(\boldsymbol{\xi},\mathbf{z};\boldsymbol{\beta}) &= -\{diag[\mathbf{p}_r(\mathbf{z};\boldsymbol{\beta})] - \mathbf{p}_r(\mathbf{z};\boldsymbol{\beta})\mathbf{p}'_r(\mathbf{z};\boldsymbol{\beta})\}\otimes\mathbf{z}\mathbf{z}', &(4)\end{aligned}$$

respectively, so that the $K(K-1)L(L+1)/4$ IM influence functions are:

$$\mathbf{m}_r(\boldsymbol{\xi},\mathbf{z};\boldsymbol{\beta}) = vech[\mathbf{u}_r(\boldsymbol{\xi},\mathbf{z};\boldsymbol{\beta})\mathbf{u}'_r(\boldsymbol{\xi},\mathbf{z};\boldsymbol{\beta}) - \{diag[\mathbf{p}_r(\mathbf{z};\boldsymbol{\beta})] - \mathbf{p}_r(\mathbf{z};\boldsymbol{\beta})\mathbf{p}'_r(\mathbf{z};\boldsymbol{\beta})\}]\otimes vech(\mathbf{z}\mathbf{z}'). \quad (5)$$

2) Let $\overline{\mathbf{m}}_{rN}(\hat{\boldsymbol{\beta}}_N)$ denote the sample mean of $\mathbf{m}_r(\boldsymbol{\xi},\mathbf{z};\boldsymbol{\beta})$ evaluated at $\hat{\boldsymbol{\beta}}_N$, and define the joint covariance matrix of the IM influence functions and score

$$\begin{bmatrix}\mathcal{R}(\boldsymbol{\beta}) & \mathcal{U}(\boldsymbol{\beta})\\ \mathcal{U}'(\boldsymbol{\beta}) & \mathcal{I}(\boldsymbol{\beta})\end{bmatrix} = Var\left\{\begin{array}{c}\mathbf{m}_r(\boldsymbol{\xi}_r,\mathbf{z};\boldsymbol{\beta})\\ \mathbf{s}_r(\boldsymbol{\xi}_r,\mathbf{z};\boldsymbol{\beta})\end{array}\right\}. \quad (6)$$

Then, under correct specification, the IM test statistic

$$N\times\overline{\mathbf{m}}'_{rN}(\hat{\boldsymbol{\beta}}_N)[\mathcal{R}(\boldsymbol{\beta}_0) - \mathcal{U}(\boldsymbol{\beta}_0)\mathcal{I}^{-1}(\boldsymbol{\beta}_0)\mathcal{U}(\boldsymbol{\beta}_0)]^{-1}\overline{\mathbf{m}}_{rN}(\hat{\boldsymbol{\beta}}_N) \xrightarrow{d} \chi^2_{K(K-1)L(L+1)/4}. \quad (7)$$

If following Newey (1985) and Tauchen (1985) we regard the IM test as a moment test of the influence functions (5), it is clear that it is effectively testing the conditional mean independence of the conditionally demeaned outer product of the generalised residuals. Thus, it resembles a multivariate version of White's (1980) test for residual conditional heteroskedasticity, which in turn confirms Chesher's (1984) reinterpretation of the IM test as a score test for neglected unobserved heterogeneity.

One feasible version of IM test statistic (7) replaces the elements of (6) by their sample counterparts evaluated at $\hat{\boldsymbol{\beta}}_N$, which Chesher (1983) and Lancaster (1984) showed is numerically identical to NR^2 in the regression of 1 on $\mathbf{m}_r(\boldsymbol{\xi}_r,\mathbf{z};\hat{\boldsymbol{\beta}}_N)$ and $\mathbf{s}_r(\boldsymbol{\xi}_r,\mathbf{z};\hat{\boldsymbol{\beta}}_N)$. Given that this yields very noisy estimators of (6), we propose another feasible version of the IM test that evaluates the different elements of (6) by relying on the law of iterated expectations, with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}_N$ and unconditional expectations by sample averages. Our next result provides analytical expressions for the required conditional moments for $\mathbf{m}_{j\ell} = m_{jl}vech(\mathbf{z}\mathbf{z}')$, with $m_{jl} = u_j u_l$, and $\mathbf{s}_j = u_j\mathbf{z}$:

Proposition 2. *a) The relevant conditional variances and covariances required to compute $\mathcal{R}$ are:*

$$cov(\mathbf{m}_{j\ell},\mathbf{m}_{j'\ell'}) = E[E(m_{j\ell}m_{j'\ell'}|\mathbf{z})vech(\mathbf{z}\mathbf{z}')vech'(\mathbf{z}\mathbf{z}')], \quad \textit{where}$$

$$E(m_{jj}^2|\mathbf{z}) = p_j - 5p_j^2 + 8p_j^3 - 4p_j^4, \quad E(m_{j\ell}^2|\mathbf{z}) = p_j^2 p_\ell + p_j p_\ell^2 - 4p_j^2 p_\ell^2,$$
$$E(m_{jj}m_{j'j'}|\mathbf{z}) = -p_j p_{j'} + 2p_j^2 p_{j'} + 2p_j p_{j'}^2 - 4p_j^2 p_{j'}^2,$$
$$E(m_{jj}m_{j'\ell}|\mathbf{z}) = 2p_j p_{j'} p_\ell - 4p_j^2 p_{j'} p_\ell, \quad E(m_{j\ell}m_{j'\ell}|\mathbf{z}) = p_\ell p_j p_{j'} - 4p_\ell^2 p_j p_{j'}$$
$$E(m_{jj}m_{j\ell}|\mathbf{z}) = -p_j p_\ell + 4p_j^2 p_\ell - 4p_j^3 p_\ell \ \textit{and} \quad E(m_{j\ell}m_{j'\ell'}|\mathbf{z}) = -4p_j p_\ell p_{j'} p_{\ell'}.$$

b) In turn, the relevant conditional covariances required to compute $\mathcal{U}$ are:

$$E(\mathbf{m}_{j\ell}\mathbf{s}'_{j'}) = cov(\mathbf{m}_{j\ell}, \mathbf{s}_{j'}) = E[E(\mathbf{m}_{j\ell}u_j|\mathbf{z})vech(\mathbf{zz}')\mathbf{z}'], \quad \textit{where}$$
$$E(m_{jj}u_j|\mathbf{z}) = p_j - 3p_j^2 + 2p_j^3, \qquad E(m_{jj}u_{j'}|\mathbf{z}) = -p_j p_{j'} + 2p_j^2 p_{j'},$$
$$E(m_{j\ell}u_j|\mathbf{z}) = -p_j p_\ell + 2p_j^2 p_\ell \quad \textit{and} \quad E(m_{j\ell}u_{j'}|\mathbf{z}) = 2p_j p_\ell p_{j'}.$$

c) Finally, the information matrix is

$$\mathcal{I} = E(\mathbf{s}_j\mathbf{s}'_j) = Var(\mathbf{s}_j) = E\{[diag(\mathbf{p}_r) - \mathbf{p}_r\mathbf{p}'_r] \otimes \mathbf{zz}'\}. \tag{8}$$

It is important to mention that the IM test cannot be computed when the only regressor is a constant because in that case the score simplifies to $\mathbf{u}_r$ and the influence functions underlying the IM test have zero mean in the sample when evaluated at $\hat{\boldsymbol{\beta}}_N$. The same situation arises when the explanatory variables consist of an exhaustive set of dummy variables that in practice generate a partition of the observations because the coefficients of those dummies effectively correspond to a model which imposes that the probabilities are constant within each category but heterogeneous across categories. In both these cases, the multinomial logit model provides a perfect fit to the data. Nevertheless, as soon as at least one of the elements of $\mathbf{z}$ is a continuous random variable, the IM test can be computed.[2]

Composite likelihood:. A well-known property of multinomial logit models is that they continue to represent the relative probabilities of any subset of categories for those observations belonging to them. In particular, if we focus on the first and second categories only, we will end up with the following binary logit model:

$$p_2^b(\mathbf{z};\boldsymbol{\beta}_2) = \Pr(\xi_2 = 1|\xi_3 = \ldots = \xi_K = 0, \mathbf{z}) = \frac{e^{\boldsymbol{\beta}'_2\mathbf{z}}}{1 + e^{\boldsymbol{\beta}'_2\mathbf{z}}} = p_2(\mathbf{z};\boldsymbol{\beta}_2) \cdot \frac{1 + \sum_{\ell=2}^{K} e^{\boldsymbol{\beta}'_\ell\mathbf{z}}}{1 + e^{\boldsymbol{\beta}'_2\mathbf{z}}}$$

with the identification condition $\boldsymbol{\beta}_1 = \mathbf{0}$. Since this is true for any other category, a popular consistent estimation method for multinomial logit models obtains $\boldsymbol{\beta}_j$ from $K-1$ such binary logit models, in what is effectively a composite likelihood approach (see Lindsay (1988)). This yields computational gains at the cost of asymptotic efficiency. Nevertheless, the results in Proposition 1 apply to each of those conditional binary logit models as well, with the number of degrees of freedom becoming $L(L+1)/2$. For that reason, in Section 3 we study these binary IM tests too.

[2]The number of degrees of freedom might need to be adjusted in very special circumstances. For example, in a binary logit model with a constant and a single continuous explanatory variable, the IM test statistic will generally be distributed as a χ_1^2 when the slope coefficient is actually 0.

Unfortunately, the relationship between the IM test for the full model and the $K-1$ IM tests for the binary models is not straightforward because they are based on different subsets of observations. However, they all maintain not only the same distribution for the underlying choice shocks but also the independence of irrelevant alternatives assumption, which is precisely what guarantees the validity of the binary models.

3. Monte Carlo simulations

The asymptotic distribution of the IM test might not be very reliable in small samples. For that reason, we study its size and power properties in simulated samples of length $N = 125$, $N = 500$ and $N = 2,000$. To estimate the parameters for binary and multinomial logit models, we make use of the MATLAB toolbox available at https://www.spatial-econometrics.com/ (see LeSage and Pace (2009)).

3.1. Size properties

When assessing size, we generate $10,000$ samples under the null for each data generating process (DGP) we describe below. We then compare two asymptotically equivalent versions of the infeasible IM test statistic in (7): the Outer-Product-of-the Score version proposed by Chesher (1983) and Lancaster (1984) (OPS), and one that replaces the true parameter values $\boldsymbol{\beta}_0$ with their MLEs $\hat{\boldsymbol{\beta}}_N$ in the theoretical expressions of the conditional variances and covariances in Proposition 2 (CM). In all cases, we consider not only asymptotic critical values but also a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null, as proposed by Horowitz (1994).[3]

We simulate multinomial logit models with $K = 3$ and $K = 5$ categories, always including a constant and one or two continuous regressors. Details on the specific designs can be found in Supplemental Appendix C.1. Table 1 contains the rejection rates of the multinomial IM tests at the 1%, 5% and 10% significance levels. Panels A and B refer to models with three categories, with two and three explanatory variables, respectively, while Panels C and D to models with five categories.

The rejection rates using asymptotic critical values in the left subpanels of Table 1 confirm the need for finite sample size adjustments, especially for the OPS version of the IM test.[4] Still, the quality of the asymptotic approximation is much better when we use the

[3]Horowitz (1994) found that increasing the number of bootstrap samples beyond 99 had little effect on the results of his experiments.

[4]Given the number of replications, the 95% asymptotic confidence intervals for the Monte Carlo rejection probabilities under the null are (0.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1%, 5% and 10% levels.

theoretical expressions for the weighting matrix even in samples of size $N = 500$, although there is still a systematic overrejection of the null at the 1% level.

In contrast, the bootstrap-based rejection rates in the right subpanels of Table 1 give a completely different picture: sizes are very accurate and almost all Monte Carlo rejection rates fall within the relevant 95% confidence set, with the exceptions of the OPS version for $N = 125$ and $N = 500$, and the CM version when $N = 125$ in models with five categories (Panels C and D).

In Table A1 in the supplementary material we report the same figures but for the conditional binary logits mentioned at the end of section 2.[5] Not surprisingly, there is still massive overrejection of the OPS version of the tests that rely on asymptotic critical values. Interestingly, though, the overrejections of the CM test at the 1% level are more moderate, probably due to the smaller number of degrees of freedom of their asymptotic distribution. In any event, the parametric bootstrap corrects the size distortions for all the sample sizes we consider.

3.2. Power properties

We consider three types of alternatives, with Chesher's (1984) neglected heterogeneity interpretation of the IM test providing the motivation for the first two ones. Specifically, we consider a model in which the coefficients for one of the z's take different values in two equally sized subgroups of the population, while remaining homogeneous within subgroups. In addition, we consider another model in which the coefficients for one of the z's are randomly distributed as a multivariate Gaussian vector across individuals. Finally, we generate data from an nested logit model as an example of misspecification of the functional form F. Again, Supplemental Appendix C.1 contains the details of the specific designs.

We simulate 2,500 samples for each of these alternatives. Given our results in the previous subsection, we take an accept/reject decision by systematically relying on the bootstrap CM version of the IM test statistic, thereby ensuring that we carry out a feasible size adjustment.

In Panels A to C of Table 2 we report the results for DGP a to DGP c. As expected, power increases with the sample size N. In contrast, no clear pattern arises when increasing the number of explanatory variables. In particular, power seems to increase only for DGP a. The same comment applies when we move from three to five categories.

Finally, Table A2 in Supplemental Appendix C.2 reports the same figures for the three

[5]The corresponding results for models with five categories are available upon request.

binary logits implied by the models with three categories. As expected, the same pattern is obtained. More importantly, the IM test of the multinomial model is more powerful than the binary ones.

4. Conclusions and extensions

The IM test is a very simple diagnostic that empirical researchers estimating multinomial logit models should routinely report. In this respect, our main contributions are: (i) a simple interpretation of the influence functions underlying the test in relation to the conditional variances of the generalised residuals, (ii) simple to compute expressions for the asymptotic covariance matrices, (iii) Monte Carlo evidence showing that when the model is correctly specified these expressions substantially reduce size distortions, which are practically eliminated when combined with the parametric bootstrap, and that such size-adjusted tests have good power against several empirically relevant alternatives.

Our theoretical expressions can be used for related models. Specifically, in Supplemental Appendix B we exploit the fact that the conditional multinomial model mentioned in the introduction can be written as a multinomial logit model to explain how to derive the corresponding IM test.

The specification tests in this paper can also be extended in at least three empirically relevant directions. First, we could consider discrete Markov chains in which each column of the $K \times K$ transition matrix is a multinomial logit function of the explanatory variables $\mathbf{z}$. Given that a Markov chain is a collection of K separate multinomial logit models indexed by the value taken by the preceding multinomial variable $\boldsymbol{\xi}$ with coefficients which are variation-free, the IM influence functions will be the collection of IM influence functions for each of those K multinomial models. Second, we could study mixture models and switching regression models in which the probabilities of the mixture components or regimes are determined by another multinomial logit model. Given that the multinomial variable $\boldsymbol{\xi}$ becomes latent in those circumstances, as in Amengual, Fiorentini and Sentana (2024a), we would need to compute the conditional expected values of the outer product of the generalised residuals given the observable variables to obtain the IM test. Finally, we could combine the previous two extensions in a Markov switching regime model à la in Hamilton (1989), which would force us to rely on a smoother rather than a filter, as in Almuzara, Amengual and Sentana (2019). In Amengual, Fiorentini and Sentana (2024b,c,d) we are currently pursuing these three interesting research avenues.

Acknowledgements

We would like to thank Manuel Arellano and Andrew Chesher, as well as audiences at CEMFI and Università di Firenze for useful comments and suggestions. The comments of an associate editor and an anonymous referee have also improved the paper. Of course, the usual caveat applies. The first and third authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant PID2021-128963NB-I00, while the second one is thankful to MIUR through the PRIN project "High-dimensional time series for structural macroeconomic analysis in times of pandemic".

References

Almuzara, T., Amengual, D. and Sentana, E. (2019): "Normality tests for latent variables", *Quantitative Economics* 10, 981-1017.

Amengual, D., Fiorentini, G. and Sentana, E. (2024a): "The information matrix test for Gaussian mixtures", CEMFI Working Paper 2401.

Amengual, D., Fiorentini, G. and Sentana, E. (2024b): "Information matrix tests for discrete Markov chains with covariate-dependent transition probabilities", mimeo, CEMFI.

Amengual, D., Fiorentini, G. and Sentana, E. (2024c): "Information matrix tests for switching regression models", mimeo, CEMFI.

Amengual, D., Fiorentini, G. and Sentana, E. (2024d): "Information matrix tests for Markov switching models with covariate-dependent transition probabilities", mimeo, CEMFI.

Chesher, A. (1983): "The information matrix test: simplified calculation via a score test interpretation", *Economics Letters* 13, 45-48.

Chesher, A. (1984): "Testing for neglected heterogeneity", *Econometrica* 52, 865-872.

Chesher, A. and Santos Silva, J. M. C. (2002): "Taste variation in discrete choice models", *Review of Economic Studies* 69, 147-168.

Gouriéroux, C., Monfort, A., Renault, E. and Trognon, A. (1987): "Generalized residuals", *Journal of Econometrics* 34, 5-32.

Hamilton, J.D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica* 57, 357-384.

Horowitz, J. (1994): "Bootstrap-based critical values for the information matrix test", *Journal of Econometrics* 61, 395-411.

Lancaster, A. (1984): "The covariance matrix of the information matrix test", *Econometrica* 52, 1051-1053.

LeSage, J. and Pace, R.K. (2009): *Introduction to spatial econometrics*, CRC Press.

Lindsay, B. (1988): “Composite likelihood methods”, *Contemporary Mathematics* 80, 221-240.

Mai, T., Frejinger, E. and Bastin, F. (2015): “A misspecification test for logit based route choice models”, *Economics of Transportation* 4, 215-226.

McFadden, D. (1974): “Conditional logit analysis of qualitative choice behavior”, In P. Zarembka (Ed.), *Frontiers in econometrics*, pp. 104-142, Academic Press.

Newey, W.K. (1985): “Maximum likelihood specification testing and conditional moment tests”, *Econometrica* 53, 1047-70.

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

White, H. (1980): “A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity”, *Econometrica* 48, 817-838.

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

Table 1: Size properties: Multinomial IM tests

Panel A: Three categories, two explanatory variables: $\mathbf{z} = (1, z)'$ with $z \sim i.i.d.\ N(0,1)$

	Asymptotic critical values							Bootstraped critical values					
	OPS			CM				OPS			CM		
Sample size	10%	5%	1%	10%	5%	1%	Sample size	10%	5%	1%	10%	5%	1%
125	97.58	96.01	91.05	8.40	6.37	4.14	125	7.50	3.27	0.48	9.93	5.24	1.20
500	84.41	80.29	71.32	10.32	7.44	4.25	500	8.81	4.37	0.69	10.05	5.09	1.03
2,000	57.17	50.69	39.77	10.99	7.09	3.08	2,000	10.03	4.81	0.95	10.05	5.14	1.04

Panel B: Three categories, three explanatory variables: $\mathbf{z} = (1, z_1, z_2)'$ with $(z_1, z_2) \sim i.i.d.\ N(\mathbf{0}, \mathbf{I}_2)$

	Asymptotic critical values							Bootstraped critical values					
	OPS			CM				OPS			CM		
Sample size	10%	5%	1%	10%	5%	1%	Sample size	10%	5%	1%	10%	5%	1%
125	99.95	99.87	99.34	9.79	7.44	4.40	125	8.80	4.23	0.74	9.59	4.78	1.01
500	95.96	94.16	89.97	11.54	7.72	3.91	500	9.30	4.48	0.80	9.62	4.69	0.88
2,000	69.61	63.01	50.96	11.68	7.36	2.94	2,000	10.09	4.91	0.95	9.81	4.95	0.96

Panel C: Five categories, two explanatory variables: $\mathbf{z} = (1, z)'$ with $z \sim i.i.d.\ N(0,1)$

	Asymptotic critical values							Bootstraped critical values					
	OPS			CM				OPS			CM		
Sample size	10%	5%	1%	10%	5%	1%	Sample size	10%	5%	1%	10%	5%	1%
125	100.00	100.00	99.99	8.93	7.42	5.45	125	3.77	1.21	0.12	6.81	2.61	0.32
500	100.00	100.00	100.00	9.31	7.68	5.61	500	6.11	2.46	0.17	9.65	5.00	1.11
2,000	100.00	100.00	99.99	10.43	8.44	5.60	2,000	10.19	5.51	0.97	9.77	5.16	1.11

Panel D: Five categories, three explanatory variables: $\mathbf{z} = (1, z_1, z_2)'$ with $(z_1, z_2) \sim i.i.d.\ N(\mathbf{0}, \mathbf{I}_2)$

	Asymptotic critical values							Bootstraped critical values					
	OPS			CM				OPS			CM		
Sample size	10%	5%	1%	10%	5%	1%	Sample size	10%	5%	1%	10%	5%	1%
125	100.00	100.00	100.00	9.32	7.57	5.40	125	14.84	6.39	0.64	10.31	5.05	0.94
500	100.00	100.00	100.00	12.19	10.03	7.07	500	16.93	9.01	1.70	9.98	5.11	1.24
2,000	100.00	100.00	100.00	13.72	10.76	6.75	2,000	15.91	8.66	1.58	10.16	5.12	1.06

Notes: Monte Carlo rejection rates based on 10,000 replications. OPS refers to the version of the statistic proposed by Chesher (1983) and Lancaster (1984), while CM to the feasible version that makes use of the theoretical expressions in Proposition 2 replacing the true parameter values by their MLEs and unconditional expectations by sample averages. Rejection rates in the left subpanels are based on the asymptotic distribution in Proposition 1 while the right ones on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null. See Supplemental Appendix C.1 for details about the DGPs.

Table 2: Multinomial IM tests: Power properties

Panel A: Alternative hypothesis: Group heterogeneity in β_{i2}

	Three categories							Five categories					
	$L=2$			$L=3$				$L=2$			$L=3$		
Sample size	10%	5%	1%	10%	5%	1%	Sample size	10%	5%	1%	10%	5%	1%
125	29.32	18.16	4.36	32.64	18.88	4.84	125	21.04	8.12	1.00	20.96	9.28	1.32
500	45.40	26.44	5.48	55.64	38.72	10.92	500	67.96	38.92	6.80	98.00	83.64	20.40
2,000	74.32	58.76	20.48	97.52	94.28	58.60	2,000	90.88	71.20	21.96	100.00	100.00	97.40

Panel B: Alternative hypothesis: Heterogeneous Gaussian β_{i2}

	Three categories							Five categories					
	$L=2$			$L=3$				$L=2$			$L=3$		
Sample size	10%	5%	1%	10%	5%	1%	Sample size	10%	5%	1%	10%	5%	1%
125	63.16	40.52	6.48	63.36	40.40	7.80	125	9.96	3.44	0.52	16.24	7.24	1.32
500	99.48	98.16	59.76	80.88	71.72	38.88	500	41.80	26.16	8.96	42.96	25.40	5.80
2,000	100.00	100.00	99.80	100.00	100.00	93.96	2,000	99.84	98.44	95.52	99.96	99.72	71.92

Panel C: Alternative hypothesis: Nested logit

	Three categories							Five categories					
	$L=2$			$L=3$				$L=2$			$L=3$		
Sample size	10%	5%	1%	10%	5%	1%	Sample size	10%	5%	1%	10%	5%	1%
125	35.04	21.64	4.56	24.40	14.24	3.24	125	21.26	8.60	1.04	28.68	15.44	2.16
500	89.08	74.76	23.60	71.04	56.76	22.28	500	94.72	74.28	16.84	84.80	61.88	14.04
2,000	100.00	100.00	93.36	100.00	99.88	86.20	2,000	100.00	100.00	76.00	100.00	99.96	71.20

Notes: Monte Carlo rejection rates based on 2,500 replications. Results for the feasible version of the IM test that makes use of the theoretical expressions in Proposition 2 replacing the true parameter values by their MLEs and unconditional expectations by sample averages. Rejection rates are based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null. See Supplemental Appendix C.1 for details about the DGPs.