Journal of Econometrics 184 (2015) 111–123

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing for time-invariant unobserved heterogeneity in generalized linear models for panel data[☆]

Francesco Bartolucci[a], Federico Belotti[b], Franco Peracchi[b,c,*]

[a] University of Perugia, Italy

[b] University of Rome Tor Vergata, Italy

[c] EIEF, Italy

ARTICLE INFO

Article history:
Received 14 May 2013
Received in revised form
21 April 2014
Accepted 2 September 2014
Available online 16 September 2014

JEL classification:
C12
C33
C35

Keywords:
Generalized linear models
Longitudinal data
Fixed-effects
Hausman-type tests
Self-reported health
Health and Retirement Study

ABSTRACT

Recent literature on panel data emphasizes the importance of accounting for time-varying unobservable individual effects, which may stem from either omitted individual characteristics or macro-level shocks that affect each individual unit differently. In this paper, we propose a simple specification test of the null hypothesis that the individual effects are time-invariant against the alternative that they are time-varying. Our test is an application of Hausman (1978) testing procedure and can be used for any generalized linear model for panel data that admits a sufficient statistic for the individual effect. This is a wide class of models which includes the Gaussian linear model and a variety of nonlinear models typically employed for discrete or categorical outcomes. The basic idea of the test is to compare two alternative estimators of the model parameters based on two different formulations of the conditional maximum likelihood method. Our approach does not require assumptions on the distribution of unobserved heterogeneity, nor it requires the latter to be independent of the regressors in the model. We investigate the finite sample properties of the test through a set of Monte Carlo experiments. Our results show that the test performs well, with small size distortions and good power properties. We use a health economics example based on data from the Health and Retirement Study to illustrate the proposed test.

1. Introduction

A distinctive feature of panel data modeling is the treatment of unobserved heterogeneity, which is typically interpreted as the effect of unobservable factors on the outcome of interest. The simplest way of dealing with this form of heterogeneity is to include in the model time-invariant unobservable individual (i.e., unit-specific) effects. Assuming that these effects are constant over time, however, may be difficult to justify in certain applications. For example, Stowasser et al. (2011) convincingly argue that the dynamic pattern of self-reported health status can be better modeled by introducing a latent time-varying individual-specific health component. Clearly, biased parameter estimates may result if the individual effects are assumed to be time-invariant when in fact they are not. This is especially true in the case of long panels.

Linear panel data models with time-varying individual effects have been studied, among others, by Holtz-Eakin et al. (1988), Chamberlain (1992) and Ahn et al. (2001, 2013) in a large n and small T framework, and by Bai (2009), Bonhomme and Manresa (2012) and Kneip et al. (2012) in a large n and large T framework; see Ahn et al. (2013) for a detailed review of this literature.

On the other hand, only a few studies have tried to relax the assumption of time-invariant individual effects in nonlinear settings. For example, Heiss (2008) proposes a limited dependent variable model with time-varying effects which are assumed to follow a first-order autoregressive process with parameters that are common across sample units, while Bartolucci and Farcomeni (2009) present a multivariate extension of the dynamic logit model based on time-varying individual effects which are assumed to follow a time-homogeneous Markov chain for every sample unit. Although

☆ We thank Bill Greene, Pravin Trivedi, the Editor, an Associated Editor and two anonymous referees for their constructive suggestions. We also thank participants at the 2012 Annual Health Econometrics Workshop and the 2013 Italian Congress of Econometrics and Empirical Economics, and seminar participants at the University of Padua (Department of Statistical Sciences) and the Max Planck Institute of Economics for useful comments. We are grateful to Florian Heiss for allowing us to use his `arldv` Stata package.

* Correspondence to: Department of Economics and Finance, University of Rome Tor Vergata, via Columbia 2, 00133 Rome, Italy. Tel.: +39 06 7259 5934; fax: +39 06 2040 219.

E-mail address: franco.peracchi@uniroma2.it (F. Peracchi).

http://dx.doi.org/10.1016/j.jeconom.2014.09.002
0304-4076/

the specification in Heiss (2008) is parsimonious (it uses only one additional parameter with respect to a standard random-effects model) and perhaps more easily justifiable in many applications, the discrete approach adopted by Bartolucci and Farcomeni (2009) results in a model that is more flexible and tends to fit the data better; see Bartolucci et al. (2011) for more detailed comments. Unlike the linear case, however, both approaches are computationally demanding. Further, the first approach requires strong parametric assumptions on the distribution of the random effects. Therefore, practitioners may find it useful to carry out a preliminary test for the presence of time-invariant unobserved heterogeneity before estimating this type of models.

In this paper, we present a simple test for the null hypothesis of time-invariant individual effects in generalized linear models (GLMs) for panel data. This class of models is quite broad and includes the Gaussian linear model and a variety of nonlinear models typically employed for discrete or categorical outcomes, such as logit, probit, Poisson and negative binomial regression models. The basic idea of the test is to compare two alternative estimators of the model parameters based on two different formulations of the conditional maximum likelihood method. It extends to GLMs with canonical link the suggestion by Wooldridge (2010, p. 325) of comparing the fixed-effects and the first-difference estimators as a way of formally testing violations of strict exogeneity.

Because our test is a pure specification test[1] based on the comparison of two alternative estimators of the same parameter vector, we refer to it as a Hausman-like test. Unlike the standard version of the Hausman test (Hausman, 1978), however, we compare estimators that are both inconsistent under the alternative. In fact, as pointed out by Ruud (1984), what matters for a specification test to have power is that it is based on estimators that diverge under the alternative (that is, their difference converges in probability to a nonzero limit), and that the sampling variance of their difference is sufficiently small. We show that, since our alternative estimators depend on different functions of the data, they generally converge in probability to different points in the parameter space when the individual effects are time-varying. Thus, our test has power against a variety of alternatives resulting in time-varying individual effects, such as omitted time-varying regressors, failure of functional form assumptions, and general misspecification of the systematic part of the model. Clearly, when the inconsistency of both estimators is the same, as in the case of a panel with only two waves, our test has no power.

It is worth emphasizing three features of our test. First, it does not require assumptions on the distribution of unobserved heterogeneity, nor it requires the latter to be independent of the regressors in the model. Second, it can be easily implemented using standard statistical software, as the test statistic is a simple quadratic form involving the difference of the parameter estimates and consistent estimates of their asymptotic variances and covariance.[2] Third, it does not require assumption on how time-invariant regressors enter the model, as the conditional likelihood function does not depend on them.

The remainder of this paper is organized as follows. Section 2 introduces our test in the case of a linear panel data model and analyzes its power properties in this simple setting. Section 3 presents our general statistical framework for the test. Section 4 investigates the small sample properties of the proposed test through a set of Monte Carlo experiments. Section 5 provides an empirical illustration based on data from the Health and Retirement Study. Finally, Section 6 offers some conclusions.

[1] A pure specification test one that places little structure on the alternative hypothesis; see Cox and Hinkley (1974) and Ruud (1984) for a detailed discussion.

[2] We implemented the proposed test in a series of R and Stata functions which are available from the corresponding author upon request.

2. The test in the case of linear panel data models

Consider a balanced panel where n units, drawn at random from a given population, are observed for T periods. For each sample unit $i = 1, \ldots, n$, we denote by $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ the vector of observations on the outcome of interest and by $\boldsymbol{X}_i$ the matrix of observations on k time-varying regressors. The tth row of $\boldsymbol{X}_i$ is denoted by $\boldsymbol{x}_{it} = (x_{it1}, \ldots, x_{itk})'$.

Under the null hypothesis of time-invariant unobserved heterogeneity, our model for the data is the standard linear panel data model

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{it} + \epsilon_{it}, \quad i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{1}$$

where α_i is a time-invariant unobservable individual effect and the error vector $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$ is assumed to be mean independent of $\boldsymbol{X}_i$. Note that, at this stage, we make no other assumption on the ϵ_{it}, so they may be heteroskedastic or serially correlated for a given i. Under our set of assumptions, a consistent estimator of $\boldsymbol{\beta}$ is the fixed-effects (FE) estimator

$$\hat{\boldsymbol{\beta}}_1 = \left(\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1} \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{y}}_i,$$

with $\tilde{\boldsymbol{X}}_i = \boldsymbol{L}\boldsymbol{X}_i$ and $\tilde{\boldsymbol{y}}_i = \boldsymbol{L}\boldsymbol{y}_i$, where $\boldsymbol{L}$ is the $T \times T$ symmetric idempotent matrix that transforms a vector into deviations from the time average of its elements. An alternative consistent estimator of $\boldsymbol{\beta}$ is the first-difference (FD) estimator

$$\hat{\boldsymbol{\beta}}_2 = \left(\sum_{i=1}^{n} \Delta\boldsymbol{X}_i'\Delta\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{n} \Delta\boldsymbol{X}_i'\Delta\boldsymbol{y}_i,$$

where $\Delta\boldsymbol{X}_i = \boldsymbol{P}\boldsymbol{X}_i$, $\Delta\boldsymbol{y}_i = \boldsymbol{P}\boldsymbol{y}_i$ and $\boldsymbol{P}$ is the $(T-1) \times T$ matrix that transforms a vector into first differences. Both estimators may be regarded as OLS estimators based on different transformations of the original data. Since we allow the ϵ_{it} to be heteroskedastic or serially correlated, neither estimator is efficient under the null hypothesis,[3] although both are consistent.

2.1. The test statistic

To test the null hypothesis of time-invariant unobserved heterogeneity we propose a Hausman-type test based on the difference $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2$ between the FE and the FD estimators. In fact, comparing the FE and FD estimators via a Hausman test is mentioned by Wooldridge (2010, p. 325) as one way to formally detect violations of strict exogeneity,[4] although he does not study in detail the power properties of the test and its possible generalization to nonlinear models.

Under the null hypothesis of time-invariant unobserved heterogeneity,

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta} \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}\boldsymbol{0} \\ \boldsymbol{0}\end{pmatrix}, \begin{bmatrix}\boldsymbol{V}_1 & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{12}' & \boldsymbol{V}_2\end{bmatrix}\right).$$

This implies that the asymptotic null distribution of $\sqrt{n}\hat{\boldsymbol{\delta}} = \sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)$ is Gaussian with mean zero and variance $\boldsymbol{V}_0 = \boldsymbol{V}_1 + \boldsymbol{V}_2 - \boldsymbol{C}_{12} - \boldsymbol{C}_{12}'$. A consistent estimator of $\boldsymbol{V}_1$ is

$$\widehat{\boldsymbol{V}}_1 = \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i'\hat{\boldsymbol{\epsilon}}_{i1}\hat{\boldsymbol{\epsilon}}_{i1}'\tilde{\boldsymbol{X}}_i\right) \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1}, \tag{2}$$

[3] The FE estimator is more efficient when the errors in (1) are homoskedastic and serially uncorrelated, while the FD estimator is more efficient when they follow a random walk.

[4] It follows that our test has power against a broad class of alternatives resulting in endogeneity, such as time-varying individual effects, omitted time-varying regressors, failure of functional form assumptions and general misspecification of the systematic part of the model.

with $\hat{\boldsymbol{\epsilon}}_{i1} = \tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}}_1$, a consistent estimator of $\boldsymbol{V}_2$ has the same form as $\boldsymbol{V}_1$ with $\tilde{\boldsymbol{X}}_i$ replaced by $\Delta\boldsymbol{X}_i$ and $\hat{\boldsymbol{\epsilon}}_{i1}$ replaced by $\hat{\boldsymbol{\epsilon}}_{i2} = \Delta\boldsymbol{y}_i - \Delta\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_2$, while a consistent estimator of $\boldsymbol{C}_{12}$ is

$$\widehat{\boldsymbol{C}}_{12} = \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\hat{\boldsymbol{\epsilon}}_{i1}\hat{\boldsymbol{\epsilon}}_{i2}'\Delta\boldsymbol{X}_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\Delta\boldsymbol{X}_i'\Delta\boldsymbol{X}_i\right)^{-1}.$$

Therefore, our test statistic is

$$\hat{\xi} = n\,\hat{\boldsymbol{\delta}}'\,\widehat{\boldsymbol{V}}_0^-\,\hat{\boldsymbol{\delta}}, \tag{3}$$

where $\widehat{\boldsymbol{V}}_0 = \widehat{\boldsymbol{V}}_1 + \widehat{\boldsymbol{V}}_2 - \widehat{\boldsymbol{C}}_{12} - \widehat{\boldsymbol{C}}_{12}'$ and $\widehat{\boldsymbol{V}}_0^-$ denotes a generalized inverse of $\widehat{\boldsymbol{V}}_0$.[5] By construction, $\widehat{\boldsymbol{V}}_0$ is guaranteed to be non-negative definite. The asymptotic null distribution of $\hat{\xi}$ as $n \to \infty$ is χ^2 with number of degrees of freedom equal to the rank of $\boldsymbol{V}_0$ which, in the "regular case" when $\boldsymbol{V}_0$ is positive definite, is just equal to the number k of time-varying regressors. We can therefore test the null hypothesis in the usual way and compute an asymptotic p-value measuring the strength of the evidence provided by the data against this hypothesis. Note that our test is valid even when the errors in (1) are heteroskedastic or serially correlated.

2.2. Power of the test

For our test to have power, the FE and FD estimators must converge in probability to different points in the parameter space under the alternative. Since the two estimators are exactly the same when either $T = 2$ or $\boldsymbol{x}_{it} = \boldsymbol{x}_t$ for all i, our test has power only when $T \geq 3$ and some of the regressors vary over both i and t. Further, since the two estimators are consistent when unobserved heterogeneity is uncorrelated with the time-varying regressors, our test has power only when unobserved heterogeneity is correlated with the time-varying regressors.

In this section we study the inconsistency of the FE and FD estimators when unobserved heterogeneity is time-varying in order to draw conclusions about the power of the proposed test. For simplicity, we focus on the case of a single observed regressor x_{it}, so $y_{it} = \alpha_{it} + \beta x_{it} + \epsilon_{it}$, and we assume that

$$x_{it} = \phi\alpha_{it} + (1-\phi^2)^{1/2}z_{it}, \tag{4}$$

where the ϵ_{it} and the z_{it} are independently and identically distributed (i.i.d.), independently of the α_{it}, with zero mean and unit variance. Thus, x_{it} has zero mean and unit variance, and its covariance with α_{it} is proportional to ϕ. When $\phi = 0$, x_{it} and α_{it} are uncorrelated.

Denoting by $u_{it} = \alpha_{it} + \epsilon_{it}$ the composite error term in model (1) and letting $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iT})'$ and $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iT})'$, one may express the FE and the FD estimators as

$$\hat{\beta}_1 = \beta + \frac{\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{u}}_i}{\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{x}}_i}, \qquad \hat{\beta}_2 = \beta + \frac{\sum_{i=1}^{n}\Delta\boldsymbol{x}_i'\Delta\boldsymbol{u}_i}{\sum_{i=1}^{n}\Delta\boldsymbol{x}_i'\Delta\boldsymbol{x}_i},$$

where $\tilde{\boldsymbol{x}}_i = \boldsymbol{L}\boldsymbol{x}_i$, $\Delta\boldsymbol{x}_i = \boldsymbol{P}\boldsymbol{x}_i$, with $\tilde{\boldsymbol{u}}_i$ and $\Delta\boldsymbol{u}_i$ defined accordingly. As $n \to \infty$, we have that

$$\operatorname{plim}\hat{\beta}_1 - \beta = \frac{\mathbb{E}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{u}}_i}{\mathbb{E}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{x}}_i}, \qquad \operatorname{plim}\hat{\beta}_2 - \beta = \frac{\mathbb{E}\Delta\boldsymbol{x}_i'\Delta\boldsymbol{u}_i}{\mathbb{E}\Delta\boldsymbol{x}_i'\Delta\boldsymbol{x}_i},$$

so $\operatorname{plim}(\hat{\beta}_1 - \hat{\beta}_2)$ is just the difference between these two expressions. Moreover,

$$\mathbb{E}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{u}}_i = \phi\tilde{\tau},$$
$$\mathbb{E}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{x}}_i = \phi^2\tilde{\tau} + (1-\phi^2)(T-1),$$
$$\mathbb{E}\Delta\boldsymbol{x}_i'\Delta\boldsymbol{u}_i = \phi\Delta\tau,$$
$$\mathbb{E}\Delta\boldsymbol{x}_i'\Delta\boldsymbol{x}_i = \phi^2\Delta\tau + 2(1-\phi^2)(T-1),$$

where $\tilde{\tau} = \sum_{t=1}^{T}\mathbb{E}(\alpha_{it} - \bar{\alpha}_{it})^2$, $\Delta\tau = \sum_{t=2}^{T}\mathbb{E}(\alpha_{it} - \alpha_{i,t-1})^2$, and we use the fact that $\sum_{t=1}^{T}\mathbb{E}(z_{it} - \bar{z}_{it})^2 = T - 1$ and $\sum_{t=2}^{T}\mathbb{E}(z_{it} - z_{i,t-1})^2 = 2(T-1)$ because the z_{it} are i.i.d. with unit variance. Thus,

$$\operatorname{plim}\hat{\beta}_1 - \beta = \frac{\phi\tilde{\tau}}{\phi^2\tilde{\tau} + (1-\phi^2)(T-1)} \tag{5}$$

and

$$\operatorname{plim}\hat{\beta}_2 - \beta = \frac{\phi\Delta\tau}{\phi^2\Delta\tau + 2(1-\phi^2)(T-1)}. \tag{6}$$

This shows that our test has no power when $\phi = 0$, because in this case both estimators are consistent, nor when $\phi = \pm 1$, because in this case both converge to $\beta \pm 1$. Note that if $\alpha_{it} - \alpha_{i,t-1}$ is stationary then $\Delta\tau$ is proportional to $T - 1$, which in turn implies that the inconsistency of the FD estimator does not depend on T.[6]

To get sharper results we need to be more specific about the time-series properties of the individual effects. We first consider the case of individual effects that are independent across sample units and follow a stationary AR(1) process parameterized as

$$\alpha_{it} = \begin{cases} \upsilon_{i1}, & t = 1, \\ \rho\alpha_{it-1} + (1-\rho^2)^{1/2}\upsilon_{it}, & t = 2, \ldots, T, \end{cases} \tag{7}$$

where the υ_{it} are i.i.d. with zero mean and unit variance, independently of the ϵ_{it} and the z_{it}. Note that $\rho = 1$ here represents the case where the individual effects are time-invariant, while $\rho = 0$ represents the case where they follow a white-noise. Appendix A.1 shows that, under (7), $\tilde{\tau} = T - 1 - 2\sum_{t=1}^{T-1}[1 - (t/T)]\rho^t$ and $\Delta\tau = 2(T-1)(1-\rho)$. If $\rho = 1$ then $\tilde{\tau} = \Delta\tau = 0$, whereas if $\rho = 0$ then $\tilde{\tau} = T - 1$ and $\Delta\tau = 2(T-1)$. Thus, $\hat{\beta}_1$ converges in probability to β if $\rho = 1$ and to $\beta + \phi$ if $\rho = 0$. If $-1 < \rho < 1$, then $\tilde{\tau} = (T-1)(1-\tilde{\rho})$, where

$$\tilde{\rho} = \frac{2\rho}{(T-1)(1-\rho)}\left(1 - \frac{1}{T}\,\frac{1-\rho^T}{1-\rho}\right).$$

Thus

$$\operatorname{plim}\hat{\beta}_1 - \beta = \phi\,\frac{1-\tilde{\rho}}{1-\tilde{\rho}\phi^2}. \tag{8}$$

Since $\tilde{\rho}$ increases with T, the inconsistency of $\hat{\beta}_1$ also increases with T.

As for $\hat{\beta}_2$, we have

$$\operatorname{plim}\hat{\beta}_2 - \beta = \phi\,\frac{1-\rho}{1-\rho\phi^2}, \tag{9}$$

which does not depend on T. Notice that (9) has the same form as (8) with $\tilde{\rho}$ replaced by ρ.[7] Since $\hat{\beta}_2$ also converges in probability to β if $\rho = 1$ and to $\beta + \phi$ if $\rho = 0$, our test has no power in these two cases. On the other hand, the fact that $\hat{\beta}_1$ and $\hat{\beta}_2$ behave very differently as functions of T when $T \geq 3$ is the source of the increasing power of our test as T increases. Fig. 1 shows the relationship between $\operatorname{plim}(\hat{\beta}_1 - \hat{\beta}_2)$ and ρ for $\phi = 0.50$ and different values of T ($T = 3, 5, 10$). It is interesting to note that this relationship is inversely U-shaped, with evidence of an asymmetric behavior for low and high values of ρ.

5 Generalized inverses are not unique, but Holly and Monfort (1986) show that test statistics of the form (3) are invariant to the choice of generalized inverse.

6 A similar result was noted by Wooldridge (2010, pp. 322–323) for the case when x_{it} is weakly dependent and $\Delta x_{it}\Delta u_{it}$ is stationary.

7 We thank an Associate Editor for pointing out this interesting result.

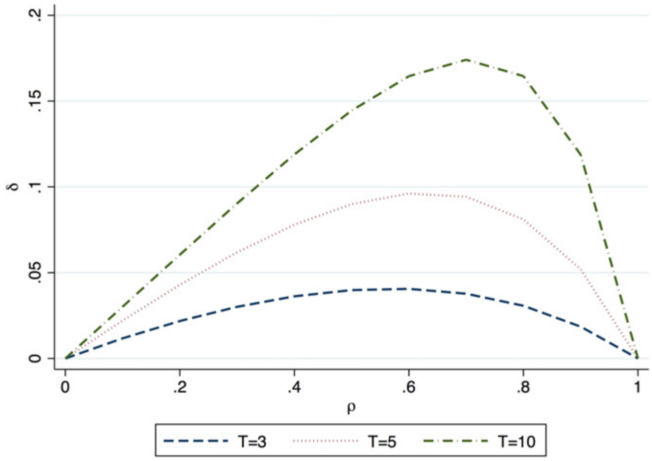

Fig. 1. Behavior of the difference δ between the inconsistency of the FE and the FD estimators as a function of the autocorrelation coefficient ρ in (7) for $\phi = 0.50$ and different values of T.

In Appendix A.1 we also consider the case when α_{it} follows a pure random walk

$$\alpha_{it} = \begin{cases} v_{i1}, & t = 1, \\ \alpha_{it-1} + v_{it}, & t = 2, \dots, T, \end{cases}$$

where the v_{it} are i.i.d., independently of ϵ_{it} and z_{it}. In this case $\tilde{\tau} = (T^2 - 1)/6$ and $\Delta\tau = T - 1$, so our test has power approaching 1 as $n \to \infty$ for any T.

Finally, in the "interactive fixed-effects" case considered by Bai (2009), $\alpha_{it} = \lambda_i f_t$ with

$$f_t = \begin{cases} v_1, & t = 1, \\ \rho f_{t-1} + (1-\rho)^{1/2} v_t, & t = 2, \dots, T, \end{cases} \tag{10}$$

where $|\rho| < 1$ and the λ_i and the v_t are i.i.d., independently of the ϵ_{it} and the z_{it}.[8] The main difference with respect to the AR(1) case is that f_t is common to all units. How this "macro" factor impacts on the ith micro-unit depends on the value of λ_i. Appendix A.2 shows that, as $n \to \infty$ and $T \to \infty$, $\text{plim}\hat{\beta}_1 - \beta = \phi$ while $\text{plim}\hat{\beta}_2 - \beta$ is exactly the same as (9). We conclude that our test has no power when $\rho = 0$, that is, when the f_t are independent over time.

3. Generalization to nonlinear panel data models

In this section we extend the testing approach illustrated in Section 2 to nonlinear panel data models based on a GLM formulation (McCullagh and Nelder, 1989). Our test compares two alternative estimators of the model parameters based on two different formulations of the conditional maximum likelihood (CML) method. The first is the standard CML estimator which, under the assumption that the unobservable individual effects are time-invariant, conditions on a sufficient statistic for α_i, such as the sum y_i^+ of the outcomes observed for the ith unit over the T periods. The second is a pairwise version of the CML estimator based on pairs of consecutive outcomes, which conditions on their sum over the two periods. The basis for this extension is the fact that, under the additional assumption of Gaussian errors in (1), these two CML estimators respectively coincide with the FE and FD estimators in Section 2.

3.1. Likelihood-based justification

Under the additional assumption that the errors in model (1) are Gaussian and serially uncorrelated with constant variance σ_ϵ^2, the joint density of $\boldsymbol{y}_i$ (conditional on $\boldsymbol{X}_i$) is

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i) = \left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\right)^T \exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^T (y_{it} - \alpha_i - \boldsymbol{\beta}'\boldsymbol{x}_{it})^2\right],$$

whereas the density of $y_i^+ = \sum_{t=1}^T y_{it}$ is

$$f(y_i^+|\boldsymbol{X}_i) = \frac{1}{\sqrt{2\pi T\sigma_\epsilon^2}} \exp\left[-\frac{1}{2T\sigma_\epsilon^2}\left(y_i^+ - T\alpha_i - \sum_{t=1}^T \boldsymbol{\beta}'\boldsymbol{x}_{it}\right)^2\right].$$

Thus, the density of $\boldsymbol{y}_i$ conditional on y_i^+ (and $\boldsymbol{X}_i$) is equal to

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_i^+) = \frac{\sqrt{2\pi T\sigma_\epsilon^2}}{(\sqrt{2\pi\sigma_\epsilon^2})^T} \exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^T (\tilde{y}_{it} - \tilde{\boldsymbol{\beta}}'\boldsymbol{x}_{it})^2\right], \tag{11}$$

and depends only on $\boldsymbol{\beta}$, not on α_i. The corresponding conditional log-likelihood is equal to $L_1(\boldsymbol{\beta}) = \sum_{i=1}^n L_{1i}(\boldsymbol{\beta})$, where $L_{1i}(\boldsymbol{\beta})$ is proportional to $-\sum_{t=1}^T (\tilde{y}_{it} - \tilde{\boldsymbol{\beta}}'\boldsymbol{x}_{it})^2$. Maximizing $L_1(\boldsymbol{\beta})$ gives the full conditional maximum likelihood (FCML) estimator, which coincides with the FE estimator $\hat{\boldsymbol{\beta}}_1$. If we allow the errors in (1) to be heteroskedastic, serially correlated or Gaussian, then the FCML estimator is still consistent and asymptotically normal, but its asymptotic variance has the "sandwich form" and may be estimated consistently by $\widehat{\boldsymbol{V}}_1 = \widehat{\boldsymbol{H}}_1^{-1}\widehat{\boldsymbol{S}}_{11}\widehat{\boldsymbol{H}}_1^{-1}$, where the matrix $\widehat{\boldsymbol{H}}_1 = n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i$ is proportional to minus the Hessian of the log-likelihood $L_1(\boldsymbol{\beta})$ and the matrix $\widehat{\boldsymbol{S}}_{11} = n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i'\hat{\boldsymbol{\epsilon}}_{i1}\hat{\boldsymbol{\epsilon}}_{i1}'\tilde{\boldsymbol{X}}_i$ is proportional to the outer product of the likelihood score $\partial L_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ evaluated at $\hat{\boldsymbol{\beta}}_1$.

On the other hand, putting $T = 2$ in (11), we have

$$f(y_{i,t-1}, y_{it}|\boldsymbol{x}_{i,t-1}, \boldsymbol{x}_{it}, y_{i,t-1} + y_{it}) = \frac{1}{\sqrt{\pi\sigma_\epsilon^2}} \exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{h=t-1}^t (\tilde{y}_{iht} - \boldsymbol{\beta}'\tilde{\boldsymbol{x}}_{iht})^2\right],$$

where $\tilde{y}_{iht} = y_{ih} - 0.5(y_{it} + y_{i,t-1})$ and $\tilde{\boldsymbol{x}}_{iht} = \boldsymbol{x}_{ih} - 0.5(\boldsymbol{x}_{it} + \boldsymbol{x}_{i,t-1})$, for $h = t-1, t$. The corresponding pairwise conditional log-likelihood is $L_2(\boldsymbol{\beta}) = \sum_{i=1}^n L_{2i}(\boldsymbol{\beta})$, where $L_{2i}(\boldsymbol{\beta})$ is proportional to $-\sum_{t=2}^T (\Delta y_{it} - \boldsymbol{\beta}'\Delta\boldsymbol{x}_{it})^2$. Maximizing $L_2(\boldsymbol{\beta})$ gives the pairwise conditional maximum likelihood (PCML) estimator $\hat{\boldsymbol{\beta}}_2$, which is equivalent to the FD estimator. If we allow the errors in (1) to be heteroskedasticity, serially correlated or Gaussian, then the PCML estimator is still consistent and asymptotically normal, but its asymptotic variance has the "sandwich form" and may be estimated consistently by $\widehat{\boldsymbol{V}}_2 = \widehat{\boldsymbol{H}}_2^{-1}\widehat{\boldsymbol{S}}_{22}\widehat{\boldsymbol{H}}_2^{-1}$, where the matrix $\widehat{\boldsymbol{H}}_2 = n^{-1}\sum_{i=1}^n \Delta\boldsymbol{X}_i'\Delta\boldsymbol{X}_i$ is proportional to minus the Hessian of the log-likelihood $L_2(\boldsymbol{\beta})$ and the matrix $\widehat{\boldsymbol{S}}_{22} = n^{-1}\sum_{i=1}^n \Delta\boldsymbol{X}_i'\hat{\boldsymbol{\epsilon}}_{i2}\hat{\boldsymbol{\epsilon}}_{i2}'\Delta\boldsymbol{X}_i$ is proportional to the outer product of the likelihood score $\partial L_2(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ evaluated at $\hat{\boldsymbol{\beta}}_2$.

The Hausman-like test statistic based on the difference between these two estimators has the same form as the statistic in (3). If $\hat{\boldsymbol{\beta}}_1$ is asymptotically efficient, then we may use as weighting matrix a generalized inverse of $\widehat{\boldsymbol{V}}_0 = \widehat{\boldsymbol{V}}_2 - \widehat{\boldsymbol{V}}_1$, otherwise we use a generalized inverse of $\widehat{\boldsymbol{V}}_0 = \boldsymbol{D}_k\widehat{\boldsymbol{W}}_0\boldsymbol{D}_k'$, where $\boldsymbol{D}_k = [\boldsymbol{I}_k, -\boldsymbol{I}_k]$ and

$$\widehat{\boldsymbol{W}}_0 = \begin{bmatrix} \widehat{\boldsymbol{W}}_{11} & \widehat{\boldsymbol{W}}_{12} \\ \widehat{\boldsymbol{W}}_{12}' & \widehat{\boldsymbol{W}}_{22} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{H}}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \widehat{\boldsymbol{H}}_2 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\boldsymbol{S}}_{11} & \widehat{\boldsymbol{S}}_{12} \\ \widehat{\boldsymbol{S}}_{12}' & \widehat{\boldsymbol{S}}_{22} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{H}}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \widehat{\boldsymbol{H}}_2 \end{bmatrix}^{-1}, \tag{12}$$

with $\widehat{\boldsymbol{S}}_{12} = n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i'\hat{\boldsymbol{\epsilon}}_{i1}\hat{\boldsymbol{\epsilon}}_{i2}'\Delta\boldsymbol{X}_i$. The expression for $\widehat{\boldsymbol{V}}_0$ reported above coincides with that proposed in Section 2 but is of more general validity.

[8] For ease of exposition, we discuss the case where α_{it} has a factor structure with only one factor but results can be easily extended to the case of more than one factor.

3.2. Generalized linear models

Now assume that, under the null hypothesis of time-invariant unobserved heterogeneity, the conditional distribution of y_{it} given $\boldsymbol{X}_i$ belongs to the linear exponential family with density function of the form

$$f(y_{it}|\boldsymbol{X}_i) = f(y_{it}|\boldsymbol{x}_{it}) = \exp\left[\frac{y_{it}\eta_{it} - b(\eta_{it})}{\gamma} + c(y_{it}, \gamma)\right],$$

where η_{it} is a parameter that varies both across sample units and over time depending on the regressors and the time-invariant individual effect, $\gamma > 0$ is a dispersion parameter treated here as known, $b(\cdot)$ is a known, strictly convex and twice differentiable function, and $c(\cdot, \gamma)$ is a known function. An important property of GLMs is that the conditional mean and variance of y_{it} given $\boldsymbol{X}_i$ and α_i are respectively equal to $\mu_{it} = b'(\eta_{it})$ and $\sigma^2_{it} = \gamma\, b''(\eta_{it})$. We further assume that $\mu_{it} = h(\alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{it})$, where $h(\cdot)$ is the inverse link function.

To ensure the existence of a conditional likelihood, we restrict the inverse link function to be canonical, $h(\cdot) = b'(\cdot)$, in which case $\eta_{it} = \alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{it}$. For example, $h(\cdot)$ is the logit transformation for the binomial regression model, the log transformation for the Poisson regression model, and the identity function for the Gaussian linear model. The existence of the conditional likelihood depends on the structure of the model and is not guaranteed in general.

If the inverse link function is canonical and the y_{it} are independent conditional on $\boldsymbol{X}_i$, then the logarithm of the joint density of $\boldsymbol{y}_i$ is equal to

$$\ln f(\boldsymbol{y}_i \mid \boldsymbol{X}_i) = \alpha_i y_i^+ + \boldsymbol{\beta}' \sum_{t=1}^T \boldsymbol{x}_{it} y_{it} - \sum_{t=1}^T b(\alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{it}) + \sum_{i=1}^T c(y_{it}).$$

This log-density is the sum of two terms: the first is $\alpha_i y_i^+ - \sum_{t=1}^T b(\alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{it})$, which depends only on $y_i^+ = \sum_{t=1}^T y_{it}$ and on α_i (and also on $\boldsymbol{\beta}$ and $\boldsymbol{X}_i$), the second is $\boldsymbol{\beta}' \sum_{t=1}^T \boldsymbol{x}_{it} y_{it} + \sum_{i=1}^T c(y_{it})$, which does not depend on y_i^+ and α_i. We conclude that y_i^+ is sufficient for α_i, so the density $f(\boldsymbol{y}_i \mid \boldsymbol{X}_i, y_i^+)$ of $\boldsymbol{y}_i$ conditional on y_i^+ (and $\boldsymbol{X}_i$) depends only on $\boldsymbol{\beta}$, not on α_i; see Chamberlain (1980), Diggle et al. (2002), and Sartori and Severini (2004). Note that the conditional likelihood approach eliminates α_{it} but also any time-invariant regressor originally included in the model. The resulting FCML estimator of $\boldsymbol{\beta}$, again denoted by $\hat{\boldsymbol{\beta}}_1$, maximizes the full conditional log-likelihood

$$L_1(\boldsymbol{\beta}) = \sum_{i=1}^n L_{1i}(\boldsymbol{\beta}),$$

where $L_{1i}(\boldsymbol{\beta})$ is equal (up to an additive constant) to the logarithm of $f(\boldsymbol{y}_i \mid \boldsymbol{X}_i, y_i^+)$.

The PCML estimator, denoted again by $\hat{\boldsymbol{\beta}}_2$, maximizes instead the pairwise conditional log-likelihood function

$$L_2(\boldsymbol{\beta}) = \sum_{i=1}^n L_{2i}(\boldsymbol{\beta})$$

where $L_{2i}(\boldsymbol{\beta})$ is equal (up to an additive constant) to the logarithm of $f(y_{i,t-1}, y_{it}|\boldsymbol{x}_{i,t-1}, \boldsymbol{x}_{it}, y_{i,t-1}+y_{it})$, the density of an adjacent pair of outcomes conditional on the sufficient statistic $y_{i,t-1}+y_{it}$ for α_i. When the inverse link is canonical, this conditional density again depends only on $\boldsymbol{\beta}$, not on α_i. If $T = 2$, then $L_1(\boldsymbol{\beta}) = L_2(\boldsymbol{\beta})$ so $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ coincide.

Under the null hypothesis of time-invariant individual effects, $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are both consistent for the true value $\boldsymbol{\beta}_0$ of $\boldsymbol{\beta}$ provided that the assumed conditional mean $\mu_{it} = h(\alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{it})$ is correctly specified (Gourieroux et al., 1984). This allows the y_{it} to be dependent or "clustered" conditional on $\boldsymbol{X}_i$. It also allows for over- or under-dispersion (for example, the conditional variance of y_{it} in a count-data model may be greater than μ_{it}). In all these cases, $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are asymptotically normal but their asymptotic variance has the "sandwich form". Further,

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0\end{pmatrix} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}_0),$$

where $\boldsymbol{W}_0$ has exactly the same form as (12) and its elements may be consistently estimated by $\widehat{\boldsymbol{W}}_{11} = \widehat{\boldsymbol{H}}_1^{-1}\widehat{\boldsymbol{S}}_{11}\widehat{\boldsymbol{H}}_1^{-1}$, $\widehat{\boldsymbol{W}}_{22} = \widehat{\boldsymbol{H}}_2^{-1}\widehat{\boldsymbol{S}}_{22}\widehat{\boldsymbol{H}}_2^{-1}$ and $\widehat{\boldsymbol{W}}_{12} = \widehat{\boldsymbol{H}}_1^{-1}\widehat{\boldsymbol{S}}_{12}\widehat{\boldsymbol{H}}_2^{-1}$, with

$$\widehat{\boldsymbol{H}}_p = -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 L_{pi}(\hat{\boldsymbol{\beta}}_p)}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}'}, \quad p = 1, 2,$$

and

$$\widehat{\boldsymbol{S}}_{pq} = \frac{1}{n}\sum_{i=1}^n \frac{\partial L_{pi}(\hat{\boldsymbol{\beta}}_p)}{\partial\boldsymbol{\beta}}\frac{\partial L_{qi}(\hat{\boldsymbol{\beta}}_q)}{\partial\boldsymbol{\beta}'}, \quad p, q = 1, 2.$$

Under the alternative hypothesis of time-varying individual effects, neither estimator is generally consistent for $\boldsymbol{\beta}$. Further, being based on different functions of the data when $T > 2$, $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ will generally converge to different points in the parameter space. In fact, as pointed out by Varin et al. (2011) and Xu and Reid (2012), $\hat{\boldsymbol{\beta}}_2$ is more robust to violations of the assumption of time-invariant unobserved heterogeneity than $\hat{\boldsymbol{\beta}}_1$, as it only requires this assumption to be satisfied for the two-dimensional conditional likelihood quantities.

The above results suggest a test that rejects the null hypothesis of time-invariant unobserved heterogeneity for large values of the statistic $\hat{\xi}$ defined in (3), namely a quadratic form in the difference $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2$ with weighting matrix $\widehat{\boldsymbol{V}}_0 = \boldsymbol{D}_k\widehat{\boldsymbol{W}}_0\boldsymbol{D}_k'$, where the elements of $\boldsymbol{W}_0$ have been defined above in terms of the matrices $\widehat{\boldsymbol{H}}_p$, $p = 1, 2$, and $\widehat{\boldsymbol{S}}_{pq}$, $p, q = 1, 2$. As for the linear case, $\widehat{\boldsymbol{V}}_0$ is guaranteed to be non-negative definite, and the resulting Hausman-like test is valid even when the outcomes observed for the ith unit are "clustered" or exhibit over- or under-dispersion (Cameron and Trivedi, 2005). If, as it may happen, the asymptotic variance matrix $\boldsymbol{V}_0$ is singular, then the asymptotic null distribution of the test statistic is χ^2 with a number of degrees of freedom equal to the rank of $\boldsymbol{V}_0$.

To illustrate our results, in the remainder of this section we provide more details for some commonly used panel data GLMs in which the dispersion parameter is known, namely the binary logit model, the ordered logit model and the Poisson regression model. We refer the reader to Hausman et al. (1984) and Wooldridge (2010) for a detailed discussion of CML estimation of other GLMs, such as the exponential and gamma models for continuous nonnegative outcomes and the negative binomial (type I) model for discrete outcomes.

3.3. Examples

In the binary logit case, y_{it} can take only two values, 0 or 1. The full conditional log-likelihood $L_1(\boldsymbol{\beta})$ is based on

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_i^+) = \frac{\prod_{t=1}^T \exp(\boldsymbol{\beta}'\boldsymbol{x}_{it}y_{it})}{\sum_{\boldsymbol{d}_i\in\mathcal{D}_{i+}}\prod_{t=1}^T \exp(\boldsymbol{\beta}'\boldsymbol{x}_{it}d_{it})}, \tag{13}$$

where $\mathcal{D}_{i+}$ consists of all T-dimensional vectors $\boldsymbol{d}_i = (d_{i1}, \dots, d_{iT}')$ whose elements d_{it} are equal to 0 or 1 and add up to $y_i^+ = 0, \dots, T$. The pairwise conditional log-likelihood $L_2(\boldsymbol{\beta})$ is instead based on

$$f(y_{i,t-1}, y_{it}|\boldsymbol{x}_{i,t-1}, \boldsymbol{x}_{it}, y_{i,t-1}+y_{it}=1) = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{i,t-1}y_{i,t-1}+\boldsymbol{\beta}'\boldsymbol{x}_{it}y_{it})}{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{i,t-1}d_{i,t-1}+\boldsymbol{\beta}'\boldsymbol{x}_{it}d_{it})}, \quad (14)$$

where $d_{i,t-1}$ and d_{it} are equal to 0 or 1 and add up to 1.

In the ordered logit case, y_{it} can take any integer value from 0 to $J-1$. Let $y_{it}^{(j)}$ denote the binary indicator obtained by dichotomizing the ordinal outcome y_{it} at value j, that is, $y_{it}^{(j)} = 1\{y_{it} > j-1\}$, $j = 1, \ldots, J-1$. Under the assumption that the unknown parameter vector is the same for all $y_{it}^{(j)}$, Baetschmann et al. (2011) show that the FCML estimator maximizes

$$L_1(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{J-1} \ln f(\boldsymbol{y}_i^{(j)}|\boldsymbol{X}_i, y_{i+}^{(j)}),$$

where $f(\boldsymbol{y}_i^{(j)}|\boldsymbol{X}_i, y_{i+}^{(j)})$ has the same form as in (13) with $\boldsymbol{y}_i^{(j)} = (y_{i1}^{(j)}, \ldots, y_{iT}^{(j)})'$ and $y_{i+}^{(j)} = 1, 2, \ldots, T-1$. The PCML estimator maximizes instead

$$L_2(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{J-1}\sum_{t=2}^{T} \ln f(y_{i,t-1}^{(j)}, y_{it}^{(j)}|\boldsymbol{X}_i, y_{i,t-1}^{(j)}+y_{it}^{(j)}=1),$$

where $f(y_{i,t-1}^{(j)}, y_{it}^{(j)}|\boldsymbol{X}_i, y_{i,t-1}^{(j)}+y_{it}^{(j)}=1)$ has the same form as in (14).

In the Poisson regression case, y_{it} can take any integer value $0, 1, 2, \ldots$. The full conditional log-likelihood $L_1(\boldsymbol{\beta})$ is based on

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_i^+) = \frac{\left(\sum_{t=1}^{T} y_{it}\right)!}{\prod_{t=1}^{T} y_{it}!} \prod_{t=1}^{T}\left[\frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{it})}{\sum_{t=1}^{T}\exp(\boldsymbol{\beta}'\boldsymbol{x}_{it})}\right]^{y_{it}}.$$

As pointed out by Cameron and Trivedi (2005), this conditional log-likelihood is proportional to the concentrated log-likelihood obtained by substituting $\hat{\alpha}_i = \sum_{t=1}^{T} y_{it}/\sum_{t=1}^{T}\exp(\boldsymbol{\beta}'\boldsymbol{x}_{it})$ in the unconditional log-likelihood. The pairwise conditional log-likelihood is instead based/on

$$f(y_{i,t-1}, y_{it}|\boldsymbol{x}_{i,t-1}, \boldsymbol{x}_{it}, y_{i,t-1}+y_{it}) = \frac{(y_{i,t-1}+y_{it})!}{y_{i,t-1}!y_{it}!}\left[\frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{i,t-1})}{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{i,t-1})+\exp(\boldsymbol{\beta}'\boldsymbol{x}_{it})}\right]^{y_{i,t-1}} \times \left[\frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{it})}{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{i,t-1})+\exp(\boldsymbol{\beta}'\boldsymbol{x}_{it})}\right]^{y_{it}}.$$

4. Monte Carlo evidence

We now present some Monte Carlo evidence about the size and power properties of the proposed test for four commonly used GLMs, namely binary logit, ordered logit, Poisson regression and the Gaussian linear model. For the latter, we also compare our test with the Hausman-type test proposed by Bai (2009).

4.1. Setup

For binary and ordered logit, we generate the outcome of interest as $y_{it} = \sum_{j=1}^{J-1} 1\{y_{it}^* > \omega_j\}$, where $1\{A\}$ is the indicator of the event A, the ω_j are fixed thresholds, $J \geq 2$ is the number of outcome categories, and y_{it}^* is a continuous latent variable that obeys the linear model

$$y_{it}^* = \alpha_{it} + \beta x_{it} + \epsilon_{it}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \quad (15)$$

with x_{it} a scalar regressor and the ϵ_{it} i.i.d. as standard logistic. We use $\omega_1 = 0$ for binary logit ($J = 2$) and $\omega_j = -2, -0.75, 0.75, 2$ for ordered logit with $J = 4$ categories. For Poisson regression, we define the mean as $\lambda_{it} = \exp(\alpha_{it} + \beta x_{it})$.

For all DGPs, the individual effects α_{it} follow a stationary AR(1) process parameterized as in (7), where the υ_{it} are i.i.d. as standard Gaussian.[9] As for the autoregressive coefficient ρ, we consider a set of eleven equally-spaced values ranging from 0 (purely random individual effects) to 1 (time-invariant individual effects). To allow for dependence between the individual effects and the regressor, we generate x_{it} according to (4), where the z_{it} are i.i.d. as standard Gaussian.

Since the FCML and the PCML estimators are both inconsistent for β under model misspecification, we consider the following design. In the baseline scenario, we assume no correlation between the individual effect and the regressor ($\phi = 0$). We also set $\beta = 1$, implying a low regression R^2 (≈ 0.19) for the latent model (15).[10] We consider two departures from the baseline:

(i) $\phi = 0$ and $\beta = 2$, that is, no correlation between x_{it} and α_{it} but a higher latent regression R^2 (≈ 0.48);
(ii) $\phi = 0.50$ and $\beta = 1$, that is, positive correlation between x_{it} and α_{it}.

For the Poisson and Gaussian regression models, the FCML and the PCML estimators are consistent provided that the regressor and the individual effect are uncorrelated. Since our test has no power in this case, our baseline scenario for the Poisson and the Gaussian models has a low degree of correlation between the individual effect and the regressor ($\phi = 0.10$), which increases to $\phi = 0.50$ in our second scenario. In both scenarios, we set $\beta = 1$.

For each value of ρ and each scenario, we investigate the behavior of tests of asymptotic level equal to 5% for two different sample sizes ($n = 1000$ and 4000) and three different panel lengths ($T = 3, 5$ and 10). These sample sizes and the panel lengths are selected with an eye to the empirical illustration in Section 5. We ran a total of $11 \times 2 \times 3 \times 3 = 198$ experiments in the case of the logit and ordered logit models, and $11 \times 2 \times 3 \times 2 = 132$ experiments in the case of the Poisson and Gaussian models. The Monte Carlo size and power of our test are obtained using 1000 replications of each experiment.

4.2. Results

Tables 1 and 2 present the size of our test for all models considered, along with the mean and standard deviation (SD) of the test statistic under different scenarios. The size distortion is always very small and not statistically different from zero, but the test exhibits a slight tendency to over-reject.

Results for the power of our test are presented separately for each model, in tabular form in Tables 3–6 and graphically in Figs. 2–5. As expected, our test has no power when either $\rho = 0$ (no persistence) or $\rho = 1$ (time-invariant individual effects). Although in the Poisson and Gaussian cases it has power only when the regressor and the individual effect are correlated, in the logit and ordered logit cases it also has power when they are uncorrelated provided that $T > 3$ and ρ is away from 0 and 1.

[9] We also ran all the described experiments assuming a discrete distribution for α_{it}. In particular we have used a three-state first-order homogeneous Markov chain with zero mean and unit variance. Results are similar, so they are not reported although they are available upon request.

[10] Since the individual effects have unit variance and the ϵ_{it} have variance equal to $\pi^2/3$, if $\beta = 1$ the latent model (15) has regression R^2 equal to $\beta^2/(1+\beta^2+\pi^2/3) = 0.189$. If $\beta = 2$, then $R^2 = 0.482$.

Table 1
Size analysis for binary and ordered logit models.

T	Binary logit						Ordered logit					
	$n = 1000$			$n = 4000$			$n = 1000$			$n = 4000$		
	Mean	SD	Size	Mean	SD	Size	Mean	SD	Size	Mean	SD	Size
	$\phi = 0,\ \beta = 1$											
3	1.05	1.50	0.060	0.98	1.42	0.045	1.03	1.36	0.058	0.99	1.32	0.044
5	1.05	1.48	0.054	1.06	1.48	0.054	0.98	1.46	0.047	1.01	1.33	0.049
10	1.06	1.60	0.053	0.99	1.45	0.055	0.98	1.34	0.043	0.88	1.28	0.040
	$\phi = 0,\ \beta = 2$											
3	1.01	1.40	0.052	1.05	1.55	0.053	1.03	1.56	0.057	1.01	1.41	0.052
5	1.01	1.41	0.051	1.04	1.47	0.051	0.99	1.42	0.040	0.97	1.36	0.049
10	1.05	1.44	0.050	0.95	1.39	0.048	0.98	1.37	0.048	0.94	1.34	0.048
	$\phi = 0.50,\ \beta = 1$											
3	1.03	1.43	0.044	1.06	1.51	0.063	1.04	1.40	0.052	1.01	1.41	0.046
5	0.93	1.25	0.042	1.04	1.53	0.052	0.96	1.42	0.050	1.05	1.40	0.059
10	1.06	1.48	0.057	0.93	1.28	0.044	0.97	1.34	0.043	0.92	1.26	0.046

Table 2
Size analysis for Poisson and Gaussian regression models.

T	Poisson						Gaussian					
	$n = 1000$			$n = 4000$			$n = 1000$			$n = 4000$		
	Mean	SD	Size	Mean	SD	Size	Mean	SD	Size	Mean	SD	Size
	$\phi = 0.10,\ \beta = 1$											
3	1.00	1.32	0.052	1.04	1.43	0.050	0.98	1.39	0.051	0.99	1.44	0.054
5	1.01	1.49	0.051	1.00	1.38	0.044	0.93	1.26	0.039	1.00	1.42	0.045
10	1.05	1.57	0.062	1.01	1.44	0.046	0.97	1.41	0.046	0.95	1.35	0.045
	$\phi = 0.50,\ \beta = 1$											
3	0.98	1.34	0.043	1.05	1.46	0.055	0.98	1.39	0.051	0.99	1.44	0.054
5	1.04	1.42	0.061	0.92	1.34	0.033	0.93	1.26	0.039	1.00	1.42	0.045
10	1.06	1.51	0.055	1.01	1.56	0.054	0.97	1.41	0.046	0.95	1.35	0.045

Table 3
Power analysis for the binary logit model.

T	$\phi = 0,\ \beta = 1$						$\phi = 0,\ \beta = 2$						$\phi = 0.50,\ \beta = 1$					
	$n = 1000$			$n = 4000$			$n = 1000$			$n = 4000$			$n = 1000$			$n = 4000$		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
$\rho = 0.20$																		
3	1.00	1.40	0.049	1.19	1.55	0.073	1.03	1.41	0.052	1.08	1.43	0.057	1.08	1.64	0.063	1.41	1.99	0.104
5	1.20	1.61	0.070	1.93	2.29	0.172	1.17	1.53	0.068	2.02	2.39	0.166	1.41	2.15	0.094	2.35	2.99	0.216
10	1.66	2.06	0.126	4.46	3.99	0.451	1.73	2.03	0.126	4.65	3.93	0.478	2.31	2.70	0.205	5.97	4.72	0.595
$\rho = 0.40$																		
3	1.03	1.40	0.053	1.43	1.77	0.102	1.08	1.44	0.058	1.34	1.67	0.091	1.31	2.01	0.097	1.92	2.48	0.165
5	1.64	1.99	0.120	3.69	3.57	0.360	1.58	1.92	0.109	3.68	3.44	0.384	2.37	3.17	0.212	5.81	5.05	0.554
10	3.60	3.34	0.371	12.65	6.76	0.950	3.52	3.15	0.377	12.36	6.38	0.941	6.37	5.32	0.614	21.18	9.62	0.994
$\rho = 0.60$																		
3	1.07	1.48	0.057	1.47	1.85	0.096	1.07	1.41	0.050	1.37	1.76	0.092	1.45	2.29	0.107	2.28	2.82	0.214
5	1.87	2.10	0.151	4.54	3.95	0.461	1.77	2.07	0.136	4.51	3.79	0.474	3.18	3.82	0.289	8.90	6.41	0.768
10	5.45	4.06	0.577	20.34	8.46	0.995	5.13	3.82	0.554	18.98	7.87	0.991	11.94	7.64	0.888	42.77	14.32	1.000
$\rho = 0.80$																		
3	1.02	1.53	0.055	1.26	1.68	0.075	1.03	1.48	0.047	1.15	1.56	0.047	1.37	2.05	0.096	1.87	2.52	0.145
5	1.50	1.87	0.112	3.15	3.16	0.319	1.49	1.89	0.100	3.06	2.95	0.315	2.80	3.36	0.248	7.31	5.65	0.690
10	4.82	3.78	0.516	17.38	7.80	0.988	4.61	3.63	0.503	16.01	7.07	0.983	13.45	8.26	0.917	47.93	14.52	1.000

In line with the asymptotic behavior of the difference between the FE and the FD estimators analyzed in Section 2.2 (see especially Fig. 1), the profile of the power of our test as a function of ρ is always inversely U-shaped, with evidence of an asymmetric behavior for low and high values of ρ.

As for dependence on the panel length T, the power is always very low for short panels ($T = 3$) but increases rapidly with T. Apart from marginal differences, this behavior is common across scenarios and types of model, and is especially evident for the largest sample size ($n = 4000$). This is again in line with the asymptotic behavior observed for the difference between the FE and the FD estimators, and is consistent with the discussion in Varin (2008), according to which the ML estimator based on a slightly misspecified pairwise log-likelihood may be closer to the

Table 4
Power analysis for the ordered logit model.

T	$\phi=0,\ \beta=1$						$\phi=0,\ \beta=2$						$\phi=0.50,\ \beta=1$					
	$n=1000$			$n=4000$			$n=1000$			$n=4000$			$n=1000$			$n=4000$		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
$\rho=0.20$																		
3	0.97	1.36	0.042	1.21	1.68	0.079	1.02	1.44	0.046	1.30	1.77	0.082	1.18	1.75	0.069	1.55	2.15	0.120
5	1.36	1.82	0.090	2.58	2.84	0.238	1.40	1.82	0.088	3.19	3.26	0.320	1.66	2.29	0.119	3.09	3.49	0.296
10	2.28	2.57	0.203	6.08	4.64	0.630	2.89	2.98	0.290	8.49	5.42	0.798	3.08	3.38	0.302	9.15	5.80	0.812
$\rho=0.40$																		
3	1.04	1.41	0.059	1.53	2.00	0.118	1.09	1.49	0.050	1.70	2.02	0.125	1.52	2.25	0.116	2.73	3.20	0.248
5	1.97	2.40	0.167	4.86	4.05	0.500	2.09	2.37	0.184	6.09	4.58	0.633	3.30	3.69	0.315	9.39	6.35	0.810
10	4.87	4.03	0.526	16.65	7.59	0.986	6.31	4.61	0.644	22.01	8.81	0.998	10.65	6.96	0.846	38.37	12.44	1.000
$\rho=0.60$																		
3	1.05	1.46	0.061	1.61	2.08	0.120	1.08	1.48	0.048	1.78	2.07	0.143	1.73	2.42	0.138	3.42	3.70	0.332
5	2.25	2.57	0.202	5.96	4.48	0.609	2.39	2.57	0.219	7.19	5.03	0.699	4.98	4.72	0.483	15.43	8.27	0.966
10	7.32	5.11	0.721	26.45	9.55	1.000	9.16	5.65	0.832	33.51	10.74	1.000	21.15	10.13	0.986	79.84	18.81	1.000
$\rho=0.80$																		
3	1.04	1.49	0.057	1.32	1.82	0.097	1.03	1.48	0.049	1.37	1.69	0.091	1.49	2.13	0.100	2.63	3.19	0.240
5	1.80	2.20	0.149	4.11	3.69	0.420	1.78	2.13	0.136	4.73	3.91	0.510	4.22	4.27	0.435	12.74	7.37	0.917
10	6.43	4.72	0.648	23.01	8.76	0.999	7.73	5.11	0.743	27.87	9.73	1.000	23.89	10.97	0.996	89.08	20.18	1.000

Table 5
Power analysis for the Poisson model.

T	$\phi=0.10,\ \beta=1$						$\phi=0.50,\ \beta=1$					
	$n=1000$			$n=4000$			$n=1000$			$n=4000$		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
$\rho=0.20$												
3	1.10	1.57	0.062	1.32	2.06	0.088	1.95	2.57	0.166	3.64	3.81	0.368
5	1.23	1.78	0.076	1.80	2.46	0.136	3.87	3.89	0.398	11.21	7.24	0.865
10	1.61	2.12	0.125	3.14	3.21	0.308	8.38	6.23	0.750	30.07	11.51	0.999
$\rho=0.40$												
3	1.16	1.57	0.066	1.61	2.29	0.123	3.39	3.77	0.334	8.25	6.13	0.732
5	1.82	2.26	0.151	3.39	3.44	0.335	10.34	7.34	0.810	34.71	14.66	0.997
10	3.00	3.36	0.289	9.16	5.98	0.810	31.88	13.55	0.999	116.93	27.24	1.000
$\rho=0.60$												
3	1.30	1.82	0.086	1.71	2.31	0.128	4.10	4.31	0.397	10.97	7.14	0.843
5	2.10	2.58	0.196	4.28	4.00	0.434	16.91	10.36	0.941	55.90	20.01	1.000
10	4.24	3.88	0.436	14.33	7.57	0.951	60.59	22.41	1.000	222.99	51.84	1.000
$\rho=0.80$												
3	1.27	1.72	0.081	1.56	2.11	0.124	3.55	3.80	0.336	9.02	6.48	0.772
5	1.63	2.09	0.122	3.27	3.46	0.326	14.80	9.30	0.911	48.18	19.70	0.999
10	4.18	3.93	0.423	13.08	7.53	0.929	65.21	26.07	0.999	230.17	58.36	1.000

true parameter value than the FCML estimator. In our case, if $T > 2$ both types of conditional likelihood are misspecified when ρ is different from one, but the pairwise conditional likelihood is "less misspecified" than the full conditional likelihood, the difference between the two likelihoods increasing as T increases.

As for the role of the slope parameter β (or equivalently the regression R^2) in the latent linear model (15), the power of the test increases going from $\beta = 1$ to $\beta = 2$ in the ordered logit case, but not in the binary logit case. This behavior reflects the fact that, unlike an ordered outcome with more than two categories, a binary outcome is completely uninformative about scale.

We conclude this section by briefly summarizing the main results of the comparison with the test proposed by Bai (2009).[11] Our evidence shows that Bai's test has very small power when the individual effects are correlated with the regressor and follow the AR(1) process in (7), especially for $\rho < 0.4$. The reason is the fact that Bai's interactive fixed-effects estimator is consistent for β as both $n \to \infty$ and $T \to \infty$ when the individual effects have a factor structure, but not when T is small or when the individual effects display a different and less restrictive pattern of temporal dependence. In fact, our Monte Carlo results show that, when the time-varying individual effects follow an AR(1) process, Bai's interactive fixed-effects estimator has a finite sample bias that is too close to that of the FE estimator for his test to have power. On the other hand, as shown in Appendix A.2, our test has power in the case of a linear model with interactive fixed effects when the common factors are persistent, but has no power when they are independent over time. Bai's test is instead powerful in both cases.

5. Empirical illustration

In our empirical illustration we consider the same example analyzed by Heiss (2008). The outcome of interest is the self-rated health status (SRHS) of older Americans, recorded on a 5-point ordered scale (poor, fair, good, very good, excellent). The data are from the University of Michigan Health and Retirement

11 Detailed tabulations are available upon request.

Table 6
Power analysis for the Gaussian linear model.

T	$\phi = 0.10,\ \beta = 1$						$\phi = 0.50,\ \beta = 1$					
	$n = 1000$			$n = 4000$			$n = 1000$			$n = 4000$		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
$\rho = 0.20$												
3	1.17	1.76	0.061	1.43	1.91	0.099	3.34	3.56	0.318	9.49	5.88	0.835
5	1.58	2.13	0.122	3.38	3.33	0.344	11.49	6.70	0.891	43.08	13.33	1.000
10	3.21	3.24	0.313	10.21	6.25	0.859	39.15	12.16	1.000	155.02	24.71	1.000
$\rho = 0.40$												
3	1.38	2.01	0.092	2.18	2.57	0.193	7.35	5.45	0.718	24.92	9.65	1.000
5	2.71	3.04	0.262	8.02	5.46	0.757	34.06	11.57	1.000	133.61	23.49	1.000
10	8.54	5.58	0.785	31.68	11.16	1.000	133.70	22.40	1.000	534.52	45.58	1.000
$\rho = 0.60$												
3	1.42	2.06	0.089	2.37	2.72	0.216	9.25	6.11	0.834	32.31	11.03	1.000
5	3.30	3.43	0.311	10.46	6.29	0.852	50.37	13.98	1.000	199.21	28.67	1.000
10	13.35	7.05	0.942	50.90	14.18	1.000	227.94	28.68	1.000	911.66	58.88	1.000
$\rho = 0.80$												
3	1.23	1.84	0.073	1.73	2.22	0.138	6.20	4.92	0.628	20.54	8.83	0.996
5	2.46	2.90	0.229	7.10	5.11	0.697	38.51	12.26	1.000	152.23	25.19	1.000
10	11.84	6.59	0.919	44.63	13.33	1.000	222.26	27.79	1.000	887.22	57.72	1.000

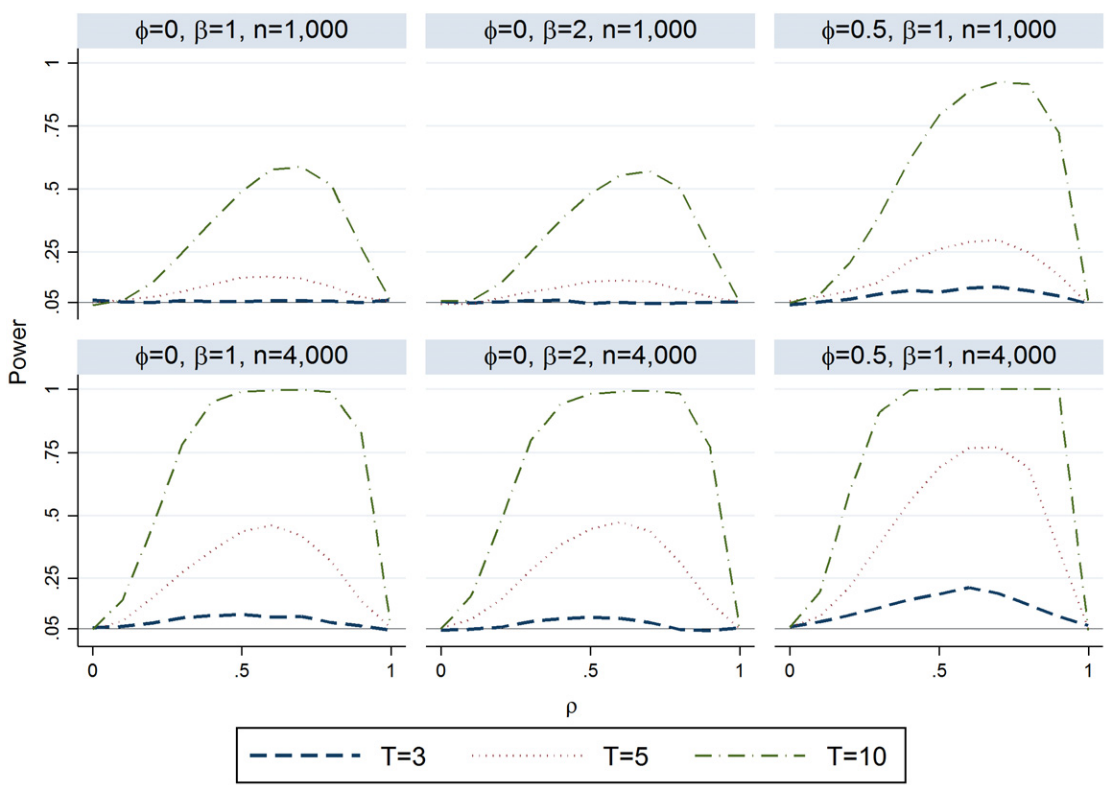

Fig. 2. Power curves of the test for the binary logit model.

Study (HRS), a longitudinal survey of the US population aged 50 and older.

As an alternative to a conventional ordered logit model with time-invariant individual effects, Heiss (2008) proposes a model that includes, in addition to both time-varying and time-invariant exogenous regressors, a set of time-varying unobservable individual effects. The time-varying individual effect, interpreted as an individual's unobserved "true" health, is assumed to be independent of the regressors and to follow an AR(1) process parameterized as in (7). Heiss (2008) argues that such a model is much more plausible than other models in the literature, as it is better able to capture the pattern of slowly declining autocorrelation exhibited by SRHS.[12] His estimate of the autoregressive parameter ρ indicates that the individual effects are highly persistent ($\hat{\rho} = 0.9439$, with an asymptotic standard error of 0.0128), but he provides no formal test of the null hypothesis that they are time-invariant. In fact, since ρ lies in the closed interval $[-1, 1]$, the hypothesis of time-invariant individual effects ($\rho = 1$) is on the boundary of the

[12] The direct competitor to the approach in Heiss (2008) would be a random-effects model with state dependence. However, state dependence is not very convincing in this context as it implies that the simple perception of own health affects future true health status.

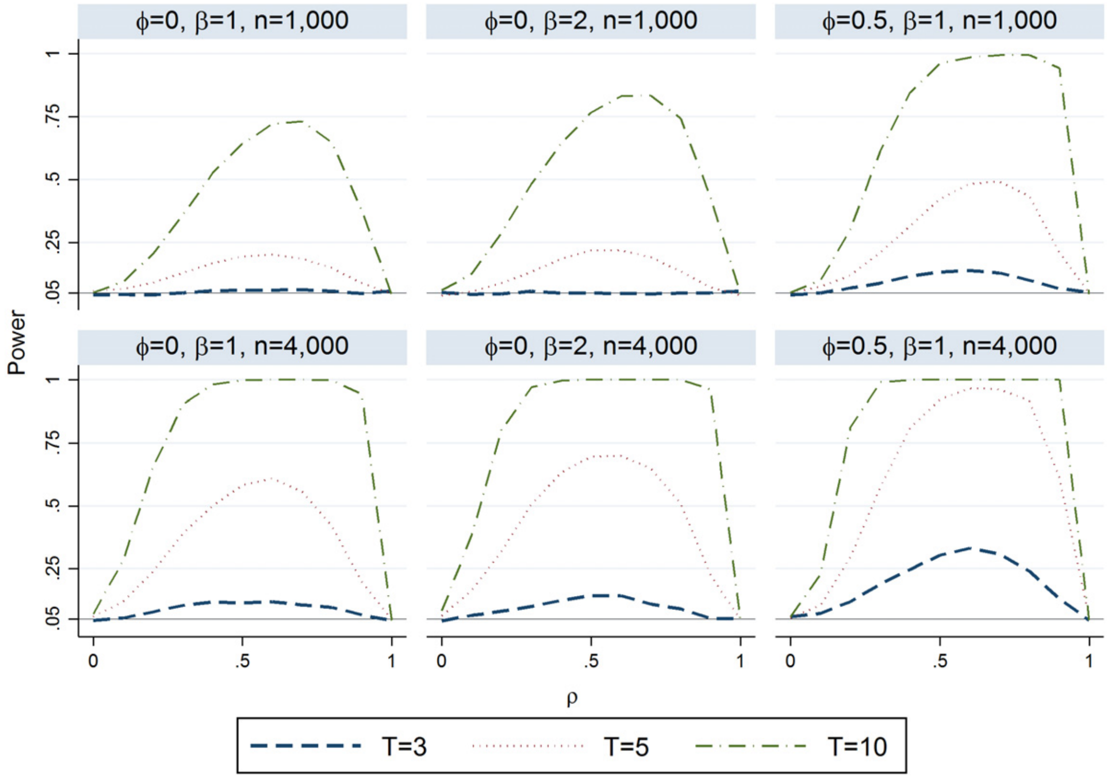

Fig. 3. Power curves of the test for the ordered logit model.

parameter space, so standard hypothesis testing procedures do not apply. On the other hand, no difficulty arises with our approach.

Our working sample from the HRS consists of a balanced panel of $n = 4094$ respondents observed for all ten available waves from 1992 to 2010 ($T = 10$).[13] Table 7 presents definitions and summary statistics for all the variables considered. It is worth noting that the cross-sectional dimension (n) and the time-series dimension (T) of our sample exactly match our Monte Carlo experiments.

Our test compares the FCML and the PCML estimators of an ordered logit model with fixed effects, as described in Section 3.3. We also consider a version of the test based on FCML and PCML estimators of a binary logit model with fixed effects, where the binary outcome is equal to one if SRHS is good or better, and to zero otherwise. For each model we consider two different specifications. The first (Model M1) includes as time-varying regressors only the body mass index (BMI) and a quadratic age spline with a single knot at age 65, which historically has been the normal retirement age in the USA. The second (Model M2) adds to Model M1 a set of wave dummies. We include no constant term and no time-invariant regressor because, under the CML method, they are eliminated from the conditional log-likelihood along with the time-invariant individual effects.

Table 8 presents the parameter estimates which are used to compute our test statistic. The top panel shows the FCML estimates for the two different specifications of each model, while the central panel shows the PCML estimates. The bottom panel shows the value of our test statistic and its p-value based on an asymptotic χ^2 distribution. Since the p-value is always lower than 1%, our test strongly rejects the hypothesis of time-invariant unobserved heterogeneity. In line with our simulations, the test statistic is larger for the ordered logit model, especially when time dummies are included. These results lend formal support to a modeling strategy that allows for time-varying unobserved heterogeneity.

To complete our empirical illustration, we estimate an AR(1) random-effects logit model similar to that considered by Heiss (2008), except for a different specification of the age effects and the inclusion of BMI as an additional time-varying regressor. We include as time-invariant regressors the same socio-demographic variables considered by Heiss (2008), namely indicator for gender, race and educational attainments. Table 9 shows the estimates obtained for both the binary and the ordered logit models under our two model specifications.[14] Our estimates are very similar to those obtained by Heiss (2008). In particular, our estimates of ρ are always very close to his estimate.

6. Conclusions

This paper proposes a computationally convenient Hausman-like specification test for the null hypothesis of time-invariant unobserved heterogeneity in GLMs for panel data against the alternative of time-varying unobserved heterogeneity of unspecified form. The test is based on the comparison of alternative estimators obtained by maximizing, respectively, a full and a pairwise conditional likelihood function.

The finite-sample properties of the proposed test are investigated via a set of Monte Carlo experiments. Our results suggest

[13] We employ the RAND HRS Data File (Version L), a user-friendly version of the data produced by the RAND Center for the Study of Aging.

[14] We estimated the model using the `arldv` Stata package kindly provided by Florian Heiss. The likelihood of this model does not have a closed-form solution, so numerical integration is necessary. We used the sequential Gauss–Legendre quadrature method proposed by Heiss (2008), with 50 integration points. To eliminate convergence issues, we also dropped 37 individuals (370 observations) with BMI values greater than the 99.9th percentile.

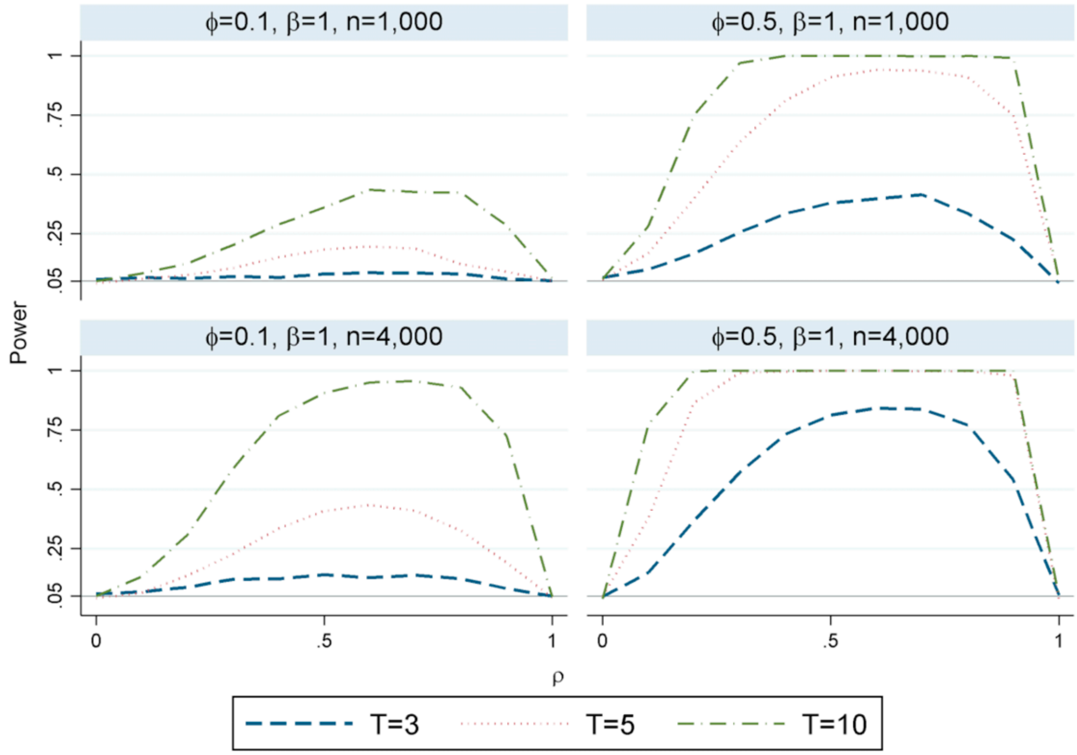

Fig. 4. Power curves of the test for the Poisson regression model.

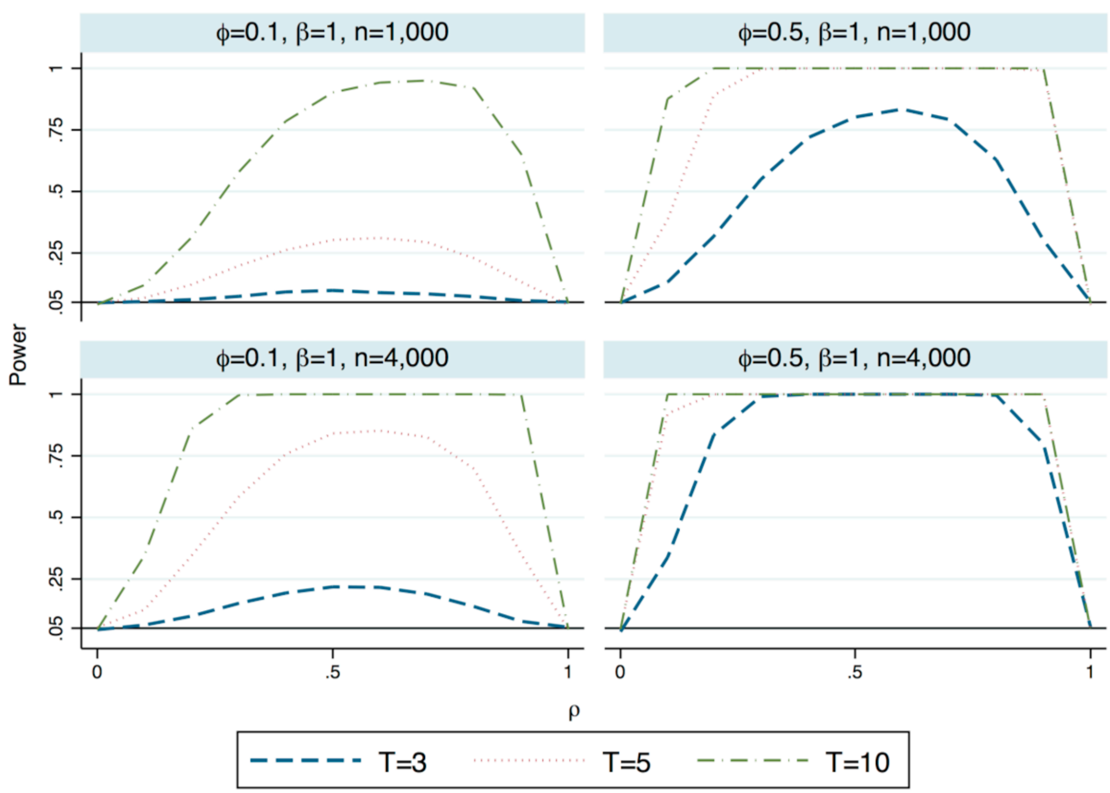

Fig. 5. Power curves of the test for the Gaussian linear model.

that the test generally performs well, showing small size distortions and good power properties for sample sizes common in economic applications.

Our test is attractive because: (i) computation of the test statistic only requires a quadratic form which involves the difference of the parameter estimates and an estimator of its asymptotic variance matrix, (ii) the test does not need assumptions on the distribution of the individual effects, (iii) individual effects can be correlated with the observed explanatory variables, (iv) it can be used regardless of the nature of the dependent variable, and (v) it can be

Table 7
Sample statistics ($n = 4094$, $T = 10$).

Variable	Description	Mean	SD	Min	Max
SRHS (categorical)	Self-rated health status, ranging from poor (1) to excellent (5)	3.47	1.04	1	5
SRHS (binary)	Dummy equal to 1 if SRHS >2	0.827	0.378	0	1
Age	Age of the respondent (years)	65.20	6.98	50	93
Female	Dummy for female	0.519	0.5	0	1
High school grad	Dummy for high school completed	0.336	0.472	0	1
Some college	Dummy for college dropout	0.203	0.402	0	1
College grad	Dummy for college completed	0.218	0.413	0	1
Non white	Dummy for asian, black or hispanic	0.149	0.357	0	1
BMI	Body mass index	27.41	4.76	12.8	51.9

Table 8
Test implementation for binary and ordered logit models.

	Binary logit		Ordered logit	
	M1	M2	M1	M2
FCML				
Age-65	−0.053***	−0.103	−0.054***	−0.110
	(0.0096)	(0.1125)	(0.0067)	(0.0771)
$(\text{Age-65})^2$	0.002**	0.002	0.003***	0.002**
	(0.0011)	(0.0013)	(0.0008)	(0.0009)
$(\text{Age-65})^2_+$	−0.006***	−0.005***	−0.005***	−0.005***
	(0.0018)	(0.0019)	(0.0012)	(0.0013)
BMI	−0.011	−0.011	−0.018**	−0.018**
	(0.0109)	(0.0110)	(0.0078)	(0.0079)
PCML				
Age-65	−0.052***	−0.217	−0.058***	−0.162
	(0.0125)	(0.1463)	(0.0085)	(0.0988)
$(\text{Age-65})^2$	0.005***	0.005**	0.004***	0.004***
	(0.0015)	(0.0018)	(0.0010)	(0.0012)
$(\text{Age-65})^2_+$	−0.008***	−0.010***	−0.007***	−0.009***
	(0.0024)	(0.0025)	(0.0016)	(0.0017)
BMI	−0.021	−0.025*	−0.008	−0.011
	(0.0141)	(0.0144)	(0.0101)	(0.0103)
Wave dummies	No	Yes	No	Yes
H_0 = time-invariant individual effects				
Test statistic	15.53	95.43	23.39	225.01
p-value	0.004	0.000	0.000	0.000

Standard errors in parenthesis.
* $p < 10\%$.
** $p < 5\%$.
*** $p < 1\%$.

easily implemented using existing software for fixed effects panel data models.

We provide an empirical illustration using the same model for SRHS as Heiss (2008) but exploiting a longer balanced panel from the HRS. We reject the null hypothesis of time-invariant unobserved heterogeneity for both binary and ordered logit versions of the model, confirming the results in Heiss (2008) but using a procedure that is both simpler and more robust. We conclude that a better model for this data may be based on the assumption that SRHS depends on unobservable "true" health which follows some time-series process with declining autocorrelations.

Appendix. Inconsistency of the FE and FD estimators

In Appendix A.1 we first consider the case of individual effects that are independent across sample units and follow either a stationary AR(1) process, as in Heiss (2008), or a pure random walk. Then, in Appendix A.2 we consider the case in which they are correlated across sample units, as in Bai (2009). In the latter case, we derive asymptotic results for both n and T diverging to infinity.

A.1. Cross-sectional independence

Suppose that the α_{it} obey a stationary AR(1) process parameterized as in (7). Under this assumption, the vector $\boldsymbol{\alpha_i} = (\alpha_{i1}, \ldots, \alpha_{iT})'$ has mean zero and variance matrix $\boldsymbol{\Sigma}$, whose generic element σ_{rs} is equal to $\rho^{|r-s|}$. Therefore, the vector $\tilde{\boldsymbol{\alpha}}_i = \boldsymbol{L}\boldsymbol{\alpha_i}$ has mean zero and variance matrix equal to $\boldsymbol{L}\boldsymbol{\Sigma}\boldsymbol{L}$, and

$$\tilde{\tau} = \operatorname{tr}\boldsymbol{L}\boldsymbol{\Sigma}\boldsymbol{L} = T - \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\rho^{|t-s|} = T - 1 - 2\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho^t,$$

where tr denotes the trace operator. If $\rho = 0$ then $\tilde{\tau} = T - 1$, whereas if $\rho = 1$ then $\tilde{\tau} = 0$. Notice that, if $-1 < \rho < 1$, then

$$\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho^t = \frac{\rho}{1-\rho}\left(1-\frac{1}{T}\frac{1-\rho^T}{1-\rho}\right).$$

Substituting the expression for $\tilde{\tau}$ in (5) shows that the inconsistency of $\hat{\beta}_1$ increases with T unless either $\phi = 0$ or $\rho = 1$, in which case it is equal to zero. On the other hand, $\operatorname{plim}\hat{\beta}_1 - \beta = \phi$ if $\rho = 0$.

The case of the FD estimator is simpler because $\Delta\alpha_{it} = \alpha_{it} - \alpha_{i,t-1}$ has mean zero and variance equal to $2(1-\rho)$, so $\Delta\tau = 2(T-1)(1-\rho)$, which is equal to $2(T-1)$ if $\rho = 0$ and is equal to zero if $\rho = 1$. Substituting in (6), we obtain

$$\operatorname{plim}\hat{\beta}_2 - \beta = \frac{\phi(1-\rho)}{\phi^2(1-\rho) + (1-\phi^2)} = \phi\,\frac{1-\rho}{1-\rho\phi^2}.$$

Table 9
AR(1) random-effects binary and ordered logit models.

	Binary logit		Ordered logit	
	M1	M2	M1	M2
Age-65	-0.058^{***}	0.015	-0.065^{***}	-0.001
	(0.0118)	(0.0215)	(0.0062)	(0.0138)
(Age-65)^2	0.003^{**}	0.003^{*}	0.004^{***}	0.003^{***}
	(0.0014)	(0.0017)	(0.0007)	(0.0008)
(Age-65)$^2_+$	-0.008^{***}	-0.009^{***}	-0.007^{***}	-0.007^{***}
	(0.0022)	(0.0024)	(0.0012)	(0.0013)
Female	0.254^{**}	0.326^{***}	0.076	0.143
	(0.1245)	(0.1266)	(0.0893)	(0.0901)
High school grad	2.079^{***}	2.118^{***}	1.457^{***}	1.490^{***}
	(0.1633)	(0.1648)	(0.1237)	(0.1242)
Some college	2.514^{***}	2.582^{***}	1.926^{***}	1.980^{***}
	(0.1876)	(0.1896)	(0.1369)	(0.1373)
College grad	3.788^{***}	3.836^{***}	2.731^{***}	2.774^{***}
	(0.2035)	(0.2055)	(0.1355)	(0.1356)
Non white	-1.269^{***}	-1.264^{***}	-1.100^{***}	-1.092^{***}
	(0.1678)	(0.1685)	(0.1270)	(0.1273)
BMI	-0.066^{***}	-0.063^{***}	-0.068^{***}	-0.066^{***}
	(0.0102)	(0.0103)	(0.0064)	(0.0065)
Constant	3.940^{***}	4.621^{***}		
	(0.3262)	(0.3643)		
Wave dummies	No	Yes	No	Yes
σ^2	3.522^{***}	3.544^{***}	2.854^{***}	2.864^{***}
	(0.1001)	(0.1018)	(0.0452)	(0.0453)
ρ	0.948^{***}	0.947^{***}	0.950^{***}	0.949^{***}
	(0.0037)	(0.0037)	(0.0022)	(0.0022)
Log-lik	$-12\,580.92$	$-12\,533.92$	$-44\,103.21$	$-43\,974.68$

Standard errors in parenthesis.
* $p < 10\%$.
** $p < 5\%$.
*** $p < 1\%$.

Now suppose that the α_{it} follow a pure random walk

$$\alpha_{it} = \begin{cases} \upsilon_{i1}, & t = 1, \\ \alpha_{it-1} + \upsilon_{it}, & t = 2, \dots, T, \end{cases}$$

where the υ_{it} are again i.i.d., independently of ϵ_{it} and z_{it}. The generic element of the matrix $\boldsymbol{\Sigma}$ is now equal to $\sigma_{rs} = \min(r, s)$, so $\tilde{\tau} = \mathrm{tr}\,\boldsymbol{L\Sigma L} = (T^2 - 1)/6$. Substituting this expression in (5), we find that the inconsistency of the FE estimator again increases with T unless $\phi = 0$, in which case it is equal to zero. As for the FD estimator, since now $\Delta\alpha_{it}$ has zero mean and unit variance, it follows that

$$\mathrm{plim}\hat{\beta}_2 - \beta = \frac{\phi(T-1)}{(T-1)(2-\phi^2)} = \frac{\phi}{2-\phi^2}.$$

A.2. Interactive fixed-effects

In the case considered by Bai (2009), $\alpha_{it} = \lambda_i f_t$ with f_t parameterized as in (10). Since $\alpha_{it} - \bar{\alpha}_{it} = \lambda_i(f_t - \bar{f})$ and $\alpha_{it} - \alpha_{i,t-1} = \lambda_i(f_t - f_{t-1})$, from the Law of Iterated Expectations we obtain the same limits in probability as in (5) and (6), except that they are now defined for $n \to \infty$ and $T \to \infty$. Because $\tilde{\tau}/T \to 1$ and $\Delta\tau/T \to 2(1-\rho)$ as $T \to \infty$, we have that

$$\mathrm{plim}\hat{\beta}_1 - \beta \to \frac{\phi}{\phi^2 + 1 - \phi^2} = \phi, \quad \text{as } T \to \infty,$$

while $\mathrm{plim}\hat{\beta}_2 - \beta$ is exactly the same as (9).

References

Ahn, S.C., Lee, Y.H., Schmidt, P., 2001. GMM estimation of linear panel data models with time-varying individual effects. J. Econometrics 101, 219–255.
Ahn, S.C., Lee, Y.H., Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. J. Econometrics 174, 1–14.
Baetschmann, G., Staub, K., Winkelmann, R., 2011. Consistent estimation of the fixed effects ordered logit model. IZA Discussion Papers 5443.
Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229–1279.
Bartolucci, F., Bacci, S., Pennoni, F., 2011. Mixture latent autoregressive models for longitudinal data. ArXiv e-prints, arXiv:1108.1498.
Bartolucci, F., Farcomeni, A., 2009. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. J. Amer. Statist. Assoc. 104, 816–831.
Bonhomme, S., Manresa, E., 2012. Grouped patterns of heterogeneity in panel data. CEMFI Working Papers 1208.
Cameron, A., Trivedi, P., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.
Chamberlain, G., 1980. Analysis of covariance with qualitative data. Rev. Econom. Stud. 47, 225–238.
Chamberlain, G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.
Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman and Hall, London.
Diggle, P.J., Heagerty, P., Liang, K.-Y., Zeger, S.L., 2002. Analysis of Longitudinal Data, seocnd ed. Oxford University Press, Oxford.
Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: theory. Econometrica 52, 681–700.
Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.
Hausman, J., Hall, B., Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52, 909–938.
Heiss, F., 2008. Sequential numerical integration in nonlinear state space models for microeconometric panel data. J. Appl. Econometrics 23, 373–389.
Holly, A., Monfort, A., 1986. Some useful equivalence properties of Hausman's test. Econom. Lett. 20, 39–43.
Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. Econometrica 56, 1371–1395.
Kneip, A., Sickles, R.C., Song, W., 2012. A new panel data treatment for heterogeneity in time trends. Econometric Theory 28, 590–628.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, second ed. Chapman and Hall, London.
Ruud, P.A., 1984. Tests of specification in econometrics. Econometric Rev. 3, 211–242.
Sartori, N., Severini, T.A., 2004. Conditional likelihood inference in generalized linear mixed models. Statist. Sinica 14, 349–360.
Stowasser, T., Heiss, F., McFadden, D., Winter, J., 2011. "Healthy, wealthy and wise?" revisited: an analysis of the causal pathways from socio-economic status to health. NBER Working Papers 17273.
Varin, C., 2008. On composite marginal likelihoods. Adv. Stat. Anal. 92, 1–28.
Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Statist. Sinica 21, 5–42.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, Massachusetts.
Xu, X., Reid, N., 2012. On the robustness of maximum composite likelihood estimate. J. Statist. Plann. Inference 141, 3047–3054.