EIEF WORKING PAPER SERIES

EIEF

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 12/13

May 2013

Testing for Time-Invariant Unobserved Heterogeneity in Generalized Linear Models for Panel Data

by

Francesco Bartolucci
(University of Perugia)
Federico Belotti
(University of Rome "Tor Vergata")
Franco Peracchi
(University of Rome "Tor Vergata" and EIEF)

Testing for time-invariant unobserved heterogeneity in generalized linear models for panel data*

Francesco Bartolucci
University of Perugia

Federico Belotti
University of Rome Tor Vergata

Franco Peracchi
University of Rome Tor Vergata and EIEF

May 14, 2013

Abstract

Recent literature on panel data has emphasized the importance of accounting for time-varying unobserved heterogeneity, which may stem either from time-varying omitted variables or macro-level shocks that affect each individual unit differently. In this paper, we propose a computationally convenient test for the null hypothesis of time-invariant individual effects. The proposed test is an application of Hausman (1978) specification test procedure and can be applied to generalized linear models for panel data, a wide class of models that includes the Gaussian linear model and a variety of nonlinear models typically employed for discrete or categorical outcomes. The basic idea is to compare fixed effects estimators defined as the maximand of full and pairwise conditional likelihood functions. Thus, the proposed approach requires no assumptions on the distribution of the individual effects and, most importantly, it does not require them to be independent of the covariates 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 quite well, with small size distortions and good power properties. A health economics example based on data from the Health and Retirement Study is used to illustrate the proposed test.

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

JEL: C12, C33, C35.

* Preliminary draft. We thank Bill Greene and Pravin Trivedi for many useful comments and 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. Corresponding author: Franco Peracchi, Department of Economics and Finance, University of Rome Tor Vergata, via Columbia 2, 00133 Rome, Italy. E-mail: `franco.peracchi@uniroma2.it`.

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 individual (i.e., unit-specific) effects. For a detailed treatment see Hsiao (2005), Wooldridge (2010), and Arellano and Bonhomme (2011). 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 time-varying. This is especially true in the case of long panels.

A few studies have recently tried to relax the assumption of time-invariant individual effects by modeling unobserved heterogeneity as a unit-specific time-series process. In the case of nonlinear panel data models, one strategy is to include time-varying random effects, treated as continuous or discrete and assumed to be independent of the covariates. For example, Heiss (2008) proposes a limited dependent variable model that relies on a sequence of time-varying effects which are assumed to follow a first-order autoregressive process whith 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 unit-specific effects which are assumed to follow a time-homogeneous Markov chain for every sample unit.

These approaches have both pros and cons. Although the specification in Heiss (2008) is parsimonious (it uses only one additional parameter with respect to a standard random-effects model) and more easily justifiable (continuous random effects are perhaps more natural to conceive 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. On the other hand, 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 propose a computationally convenient 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 proposed 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.

The basic idea of the test is to compare alternative estimators obtained by maximizing full and pairwise conditional likelihood functions. Since our test is a pure specification test[1] based on the comparison of two alternative estimators of the same parameter vector, we shall 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 argue 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. Clearly, when both estimators have the same asymptotic biases, as in the case of a panel with only two waves, the test has no power.[2]

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 covariates in the model. Second, it can be easily implemented using standard statistical software, as the computation of the test statistic only requires a quadratic form which involves the difference of the parameter estimates and a consistent estimator of its asymptotic variance matrix.[3] Third, it does not require assumption on how time-invariant regressors enter the model, as the conditional likelihood function does not depend on them.

[1] A pure specification test is a testing procedure in which no structure is placed on the alternative hypothesis; see Cox and Hinkley (1974) and Ruud (1984) for a detailed discussion.

[2] See Holly (1982) and Newey (1985) for a detailed discussion of the conditions under which a specification test is inconsistent.

[3] The proposed specification test has been implemented in a series of R and Stata functions which are available from the corresponding author upon request.

The remainder of this paper is organized as follows. Section 2 presents the statistical framework and the proposed test. Section 3 investigates the small sample properties of the proposed test through a set of Monte Carlo experiments. Our results show that the test performs quite well, with small size distortions and good power properties. Section 4 provides an empirical illustration based on data from the Health and Retirement Study. Finally, Section 5 offers some conclusions.

2 The proposed approach

We assume that the data consist of a balanced panel where n units, randomly drawn from a given population, are observed for T periods each. For each sample unit $i = 1, \dots, n$, we denote by $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})$ the vector of observed outcomes and by $\boldsymbol{X}_i$ the $T \times k$ matrix of observed covariates, with tth row equal to $\boldsymbol{x}_{it}$.

2.1 The statistical framework

Under the null hypothesis of time-invariant unobserved heterogeneity, our model is a standard GLM (McCullagh and Nelder, 1989), that is, the conditional distribution of y_{it} given $\boldsymbol{X}_i$ and the individual effect α_i is assumed to belong to the linear exponential family with density function of the form

$$f(y_{it}|\boldsymbol{X}_i, \alpha_i) = f(y_{it}|\boldsymbol{x}_{it}, \alpha_i) = \exp\left[\frac{y_{it}\eta_{it} - b(\eta_{it})}{\gamma} + c(y_{it}, \gamma)\right] \tag{1}$$

where η_{it} is a parameter that varies both across units and over time depending on the observed covariates 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 $\mu_{it} = h(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})$, where h is an invertible function called the inverse link function and $\boldsymbol{\beta}$ is the $k \times 1$ vector of parameters of interest.

If the individual effect α_i is time-invariant and the covariates are strictly exogenous conditional on α_i then, under certain conditions, inference about $\boldsymbol{\beta}$ may be based on the conditional density of the data given a sufficient statistic for the time-invariant individual effect, such as $y_{i+} = \sum_{t=1}^{T} y_{it}$; see Chamberlain (1980), Diggle et al. (2002), and Sartori and Severini (2004).

This conditional density depends only on $\boldsymbol{\beta}$, not on α_i. The resulting maximum likelihood estimator of $\boldsymbol{\beta}$, called the full conditional maximum likelihood (FCML) estimator and denoted by $\hat{\boldsymbol{\beta}}_1$, maximizes the full conditional log-likelihood function

$$L_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} \ln f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}), \tag{2}$$

where $f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+})$ is the multivariate density of the observed sequence of outcomes $\boldsymbol{y}_i$ for the ith unit conditional on the covariates in $\boldsymbol{X}_i$ and the sufficient statistic y_{i+} for α_i. To ensure the existence of the conditional likelihood (2), we restrict the inverse link function to be canonical, namely such that $h = b'$, in which case $\eta_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}$. For example, h is the log-odds transformation for the binomial regression model, the log transformation for the Poisson regression model, and the identity function for the Gaussian linear model. Notice that the existence of the conditional likelihood depends on the structure of the model and is not guaranteed in general.

In order to construct a Hausman-like specification test, we need an alternative estimator of $\boldsymbol{\beta}$ that is also consistent if the unit-specific effects are time-invariant but has different convergence properties if they are time-varying. One such estimator, called the pairwise conditional maximum likelihood (PCML) estimator and denoted here by $\hat{\boldsymbol{\beta}}_2$, may be obtained by maximizing the pairwise conditional log-likelihood function

$$L_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{t=2}^{T} \ln f(y_{i,t-1}, y_{it}|\boldsymbol{x}_{i,t-1}, \boldsymbol{x}_{it}, y_{i,t-1} + y_{it}), \tag{3}$$

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

It is easily shown that, under the null hypothesis of time-invariant individual effects, $\hat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}_2 \xrightarrow{p} \boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ denotes the true value of $\boldsymbol{\beta}$. Thus, both estimators are consistent under the null. Further, under the null,[4]

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

[4] The asymptotic distribution of the conditional maximum likelihood estimator has been obtained elsewhere under various sets of assumptions; see Andersen (1970), McFadden (1974), Huque and Katti (1976) and Pfanzagl (1993) among others. For a formal proof of consistency and asymptotic normality of the pairwise conditional likelihood estimator, see Appendix A.

where the asymptotic variance-covariance matrix $\boldsymbol{W}_0$ has the following structure

$$\boldsymbol{W}_0 = \begin{bmatrix} \boldsymbol{W}_{11} & \boldsymbol{W}_{12} \\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22} \end{bmatrix}.$$

Notice that the difference $\boldsymbol{W}_{22} - \boldsymbol{W}_{11}$ is a non-negative definite matrix because $\hat{\boldsymbol{\beta}}_2$ is asymptotically inefficient relative to $\hat{\boldsymbol{\beta}}_1$ since it discards some of the available information.

A consistent estimator of the asymptotic variance-covariance matrix $\boldsymbol{W}_0$ is

$$\widehat{\boldsymbol{W}}_0 = \begin{bmatrix} \boldsymbol{H}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{H}_2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{H}_2 \end{bmatrix}^{-1},$$

with

$$\boldsymbol{H}_p = \frac{1}{n} \frac{\partial^2 L_p(\hat{\boldsymbol{\beta}}_p)}{\partial \boldsymbol{\beta} \, \partial \boldsymbol{\beta}'}, \qquad p = 1, 2,$$

and

$$\boldsymbol{S}_{pq} = \frac{1}{n} \frac{\partial L_p(\hat{\boldsymbol{\beta}}_p)}{\partial \boldsymbol{\beta}} \frac{\partial L_q(\hat{\boldsymbol{\beta}}_q)}{\partial \boldsymbol{\beta}'}, \qquad p, q = 1, 2.$$

Thus, the finite-sample variances of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ may be estimated by $n^{-1}\widehat{\boldsymbol{W}}_{11} = n^{-1}\boldsymbol{H}_1^{-1}\boldsymbol{S}_{11}\boldsymbol{H}_1^{-1}$ and $n^{-1}\widehat{\boldsymbol{W}}_{22} = n^{-1}\boldsymbol{H}_2^{-1}\boldsymbol{S}_{22}\boldsymbol{H}_2^{-1}$ respectively, and their finite-sample covariance by $n^{-1}\widehat{\boldsymbol{W}}_{12} = n^{-1}\boldsymbol{H}_1^{-1}\boldsymbol{S}_{12}\boldsymbol{H}_2^{-1}$. Notice that, to account for the fact that outcomes are generally correlated over time for the same sample unit, the estimates $\boldsymbol{S}_{pq}$ must be cluster-robust.[5]

Under the alternative hypothesis of time-varying individual effects, neither $\boldsymbol{\beta}_1$ nor $\boldsymbol{\beta}_2$ are consistent for $\boldsymbol{\beta}$ in general. Further, being based on different functions of the data when $T > 2$, they 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.

2.2 The proposed test

The results in the previous section suggest a test that rejects the null hypothesis of time-invariant unobserved heterogeneity for large value of the statistic

$$\hat{\xi} = n \, \hat{\boldsymbol{\delta}}' \widehat{\boldsymbol{V}}_0^{-1} \hat{\boldsymbol{\delta}}, \tag{4}$$

where $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2$ and $\widehat{\boldsymbol{V}}_0 = \boldsymbol{D}\widehat{\boldsymbol{W}}_0\boldsymbol{D}'$, with $\boldsymbol{D} = (\boldsymbol{I}_k, -\boldsymbol{I}_k)$ and $\boldsymbol{I}_k$ the identity matrix of size k, is a consistent estimator of the asymptotic variance matrix $\boldsymbol{V}_0$ of $\hat{\boldsymbol{\delta}}$. Notice that $\widehat{\boldsymbol{V}}_0$ is

[5] See Moulton (1986) for a detailed discussion of cluster-robust variance-covariance matrices.

guaranteed to be positive definite, and that the resulting Hausman-like test is valid even when the data are heteroskedastic or serially correlated (Cameron and Trivedi, 2005).

If the asymptotic variance matrix $\boldsymbol{V}_0$ is nonsingular, then the test statistic $\hat{\xi}$ exists with probability approaching one for large values of n, and its asymptotic null distribution is χ^2 with number of degrees of freedom equal to k. Based on this result, we can test the null hypothesis in the usual way and compute an asymptotic p-value measuring the evidence provided by the sample against this hypothesis. It might be the case that the true asymptotic covariance matrix is singular. This may happen when there exists a linear relationship between the estimator contrasts which holds under the null hypothesis for all parameter values (Ruud, 1984). In this case, if one replaces the inverse of $\widehat{\boldsymbol{V}_0}$ in (4) with a generalized inverse, then the asymptotic null distribution of $\hat{\xi}$ is still χ^2 with number of degrees of freedom equal to the rank of $\boldsymbol{V}_0$, but this rank is now less than k.[6]

As pointed out by Verbeek and Nijman (1992), if both estimators converge to the same point in the parameter space under the alternative, then the test completely loses its power. This is clearly the case when $T = 2$. When $T > 2$, the two estimators are based on different functions of the data, so $\hat{\boldsymbol{\delta}}$ will generally have a nonzero probability limit under the alternative (Xu and Reid, 2012), which means that our test will have power against a broad class 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.

2.3 Examples

In this section we provide more detail for some commonly used panel data GLMs in which the dispersion parameter is known (or may be treated as known), namely the logit, ordered logit and Poisson regression models, and the Gaussian linear model with known dispersion parameter. We refer to Wooldridge (2010) and Hausman and Griliches (1984) for a detailed discussion of conditional maximum likelihood estimation of other GLMs, such as the exponential and gamma models for continuos nonnegative outcomes and the negative binomial (type I) model.

[6] Generalized inverses are not unique, but Holly and Monfort (1986) show that all generalized inverses give the same test statistic.

2.3.1 Logit model

In this case, possible values of the sufficient statistic for α_i are $y_{i+} = 1, 2, \dots, T-1$, and the elements of the conditional log-likelihood $L_1(\boldsymbol{\beta})$ are

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}) = \frac{\prod_{t=1}^{T} \exp(y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\sum_{\boldsymbol{z}_i \in \mathcal{Z}_{i+}} \prod_{t=1}^{T} \exp(z_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta})}, \tag{5}$$

where $\mathcal{Z}_{i+}$ is the set of all T-dimensional vectors whose elements are equal to zero or one and add up to y_{i+}. The elements of the pairwise conditional log-likelihood $L_2(\boldsymbol{\beta})$ are instead

$$f(y_{i,t-1}, y_{it}|\boldsymbol{x}_{i,t-1}, \boldsymbol{x}_{it}, y_{i,t-1} + y_{it} = 1) = \frac{\exp(y_{i,t-1}\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta} + y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\exp(z_{i,t-1}\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta} + z_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta})}, \tag{6}$$

where $z_{i,t-1}$ and z_{it} are equal to zero or one and add up to one.

2.3.2 Ordered logit model

In this case, following Baetschmann et al. (2011), we consider the estimator that can be obtained by using all possible dichotomizations $y_{it}^{(j)}$ of the ordinal outcome y_{it} for each unit in the sample, where

$$y_{it}^{(j)} = 1\{y_{it} > j-1\}, \qquad j = 1, \dots, J-1,$$

with $J > 2$. Assuming that the unknown parameter vector is the same for all $y_{it}^{(j)}$, 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 $f(\boldsymbol{y}_i^{(j)}|\boldsymbol{X}_i, \alpha_i, y_{i+}^{(j)})$ is specified as in (5) for each possible dichotomization of the ordered outcome. 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 $f(y_{i,t-1}^{(j)}, y_{it}^{(j)}|\boldsymbol{X}_i, y_{i,t-1}^{(j)} + y_{it}^{(j)} = 1)$ is specified as in (6).

2.3.3 Poisson model

In this case, the elements of the conditional log-likelihood $L_1(\boldsymbol{\beta})$ are

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}) = \frac{(\sum_{t=1}^{T} y_{it})!}{\prod_{t=1}^{T} y_{it}!} \prod_{t=1}^{T} \left[\frac{\exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\sum_{t=1}^{T} \exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})} \right]^{y_{it}}. \tag{7}$$

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

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

2.3.4 Gaussian linear model

In this case, the elements of the conditional log-likelihood $L_1(\boldsymbol{\beta})$ are

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+}) = \frac{(2\pi\gamma)^{-T/2}}{(2\pi\gamma/T)^{-1/2}}\exp\left[-\frac{1}{2\gamma}\sum_{t=1}^T(\tilde{y}_{it}-\tilde{\boldsymbol{x}}_{it}'\boldsymbol{\beta})^2\right],$$

where γ is the dispersion parameter, $\tilde{y}_{it} = y_{it}-\bar{y}_i$ and $\tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i$, with $\bar{y}_i = T^{-1}\sum_{t=1}^T y_{it}$ and $\bar{\boldsymbol{x}}_i = T^{-1}\sum_{t=1}^T \boldsymbol{x}_{it}$. The resulting FCML estimator does not depend on the dispersion parameter and can be shown to be equivalent to the within-group estimator

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

where $\tilde{\boldsymbol{X}}_i = \boldsymbol{L}_T\boldsymbol{X}_i$, $\tilde{\boldsymbol{y}}_i = \boldsymbol{L}_T\boldsymbol{y}_i$, $\boldsymbol{L}_T = \boldsymbol{I}_T - T^{-1}\boldsymbol{J}_T$ is a $T\times T$ symmetric idempotent matrix that transforms a vector (matrix) into deviations from the average of its elements, $\boldsymbol{I}_T$ is the identity matrix, and $\boldsymbol{J}_T$ is a matrix whose elements are all equal to one. A cluster-robust estimate of the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_1$ is

$$\hat{V}(\hat{\boldsymbol{\beta}}_1) = \left[\sum_{i=1}^n \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right]^{-1}\left[\sum_{i=1}^n \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{u}}_i\tilde{\boldsymbol{u}}_i'\tilde{\boldsymbol{X}}_i\right]\left[\sum_{i=1}^n \tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right]^{-1}, \tag{8}$$

where $\tilde{\boldsymbol{u}}_i = \tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i'\hat{\boldsymbol{\beta}}_1$.

The elements of the pairwise conditional log-likelihood $L_2(\boldsymbol{\beta})$ are instead

$$f(y_{i,t-1}, y_{it}|\boldsymbol{x}_{i,t-1}, \boldsymbol{x}_{it}, y_{i,t-1}+y_{it}) = \frac{(2\pi\gamma)^{-1}}{(\pi\gamma)^{-1/2}}\exp\left[-\frac{1}{4\gamma}(\Delta y_{it}-\Delta\boldsymbol{x}_{it}'\boldsymbol{\beta})^2\right],$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ and $\Delta\boldsymbol{x}_{it} = \boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1}$. The resulting PCML estimator again does not depend on the dispersion parameter and is equivalent to the first-difference estimator

$$\hat{\boldsymbol{\beta}}_2 = \left[\sum_{i=1}^n \check{\boldsymbol{X}}_i'\check{\boldsymbol{X}}_i\right]^{-1}\sum_{i=1}^n \check{\boldsymbol{X}}_i'\check{\boldsymbol{y}}_i,$$

where $\check{\boldsymbol{X}}_i = \boldsymbol{P}\boldsymbol{X}_i$, $\check{\boldsymbol{y}}_i = \boldsymbol{P}\boldsymbol{y}_i$ and $\boldsymbol{P}$ is $(T-1)\times T$ matrix that transforms the vector (or matrix) of all pairs of adjacent observations for each unit i into first differences. For example, when $T = 4$, $\boldsymbol{P}$ is the following 3×4 matrix

$$\boldsymbol{P} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}. \tag{9}$$

A cluster-robust estimate of the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_2$ has the same form as (8) with $\tilde{\boldsymbol{X}}_i$ replaced by $\check{\boldsymbol{X}}_i$ and $\tilde{\boldsymbol{u}}_i$ replaced by $\check{\boldsymbol{u}}_i = \check{\boldsymbol{y}}_i - \check{\boldsymbol{X}}_i'\hat{\boldsymbol{\beta}}_2$.

3 Monte Carlo evidence

We now present some Monte Carlo evidence about the size and power properties of the proposed test. We analyze four commonly used GLMs, namely (i) the logit model, (ii) the ordered logit model, (iii) the Poisson model, and (iv) the Gaussian linear model.

3.1 Setup

For logit and ordered logit models, the outcome of interest is generated as $y_{it} = \sum_{j=1}^{J-1} 1\{y_{it}^* > \tau_j\}$, where $1\{A\}$ is the indicator of the event A, the τ_j are thresholds, $J \geq 2$ is the number of outcome categories, and y_{it}^* is a latent continuous variable which is assumed to obey the following regression model

$$y_{it}^* = \alpha_{it} + x_{it}\beta + \varepsilon_{it}, \qquad i = 1, \ldots, n;\ t = 1, \ldots, T, \tag{10}$$

where the ε_{it} are independently and identically distributed (iid) errors with a standard logistic distribution. We use $\tau_1 = 0$ for the binary logit case ($J = 2$) and $\tau_j = -2, -0.75, 0.75, 2$ for the ordinal logit case with $J = 4$ categories. For the Poisson model, we define the mean as $\lambda_{it} = \exp(\alpha_{it} + x_{it}\beta)$ while, for the Gaussian linear model, we specify the following DGP

$$y_{it} = \alpha_{it} + x_{it}\beta + \varepsilon_{it}, \qquad i = 1, \ldots, n;\ t = 1, \ldots, T,$$

where the ε_{it} are iid errors with a standard Gaussian distribution.

For all DGPs, x_{it} is a scalar regressor and α_{it} is an individual effect that follows a stationary Gaussian first-order autoregressive process with zero mean and unit variance. Specifically, we assume that

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

where the v_{it} are iid standard Gaussian random variables. When $\rho = 1$, the unit-specific effects are time-invariant, as $\alpha_{i1} = \cdots = \alpha_{iT}$ with probability one. When $\rho = 0$, the α_{it} collapse to a Gaussian white noise with no persistence. Intermediate cases are obtained by letting ρ take different value between 0 and 1. In our Monte Carlo experiments, we consider a set of eleven equally spaced ρ values ranging from 0 to 1.

To allow for dependence between the unit-specific effects and the regressor, we generate the latter as follows

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

where the z_{it} are iid standard Gaussian random variables. Thus, x_{it} has zero mean and unit variance, and its correlation with α_{it} is equal to ϕ.

Since the FCML and the PCML estimators are both inconsistent under model misspecification, even when $\phi = 0$, we consider the following design. In the baseline scenario, we assume the absence of correlation between the individual effects and the regressor ($\phi = 0$). We also set $\beta = 1$, implying a low regression R^2 (≈ 0.19) for the latent model (10).[7] We also consider two departures from the baseline:

(i) $\phi = 0$ and $\beta = 2$, that is, no correlation between x_{it} and α_{it} and a high latent regression R^2 (≈ 0.48);

(ii) $\phi = 0.5$ and $\beta = 1$, that is, positive correlation between x_{it} and α_{it} and a low latent regression R^2 (≈ 0.19).

For the Poisson and Gaussian models (Sections 2.3.3 and 2.3.4), the FCML and the PCML estimators are always consistent, even when the unobserved heterogeneity is time-varying, provided that there is no correlation between the unit-specific effects and the regressor. Since our test has no power in this case, we do not report results for $\phi = 0$. Thus, our baseline scenario has a low degree of correlation ($\phi = 0.1$) and the only departure from the baseline is obtained by setting $\phi = 0.5$. In both cases, we set $\beta = 1$.

For each value of ρ and each scenario, we investigate the behavior of the test for two different sample sizes ($n = 1,000$ and 4,000) and three different panel lengths ($T = 5$, 7 and 10), for a total of $11 \times 2 \times 3 \times 3 = 198$ experiments in the case of logit and ordered logit models, and

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

$11 \times 2 \times 3 \times 2 = 132$ experiments for Poisson and Gaussian models. The sample sizes are selected with an eye to the empirical illustration in Section 4. Size and power of the test are computed using $1,000$ replications of each experiment.

3.2 Results

Tables 1 and 2 present the empirical size of our test for all the GLMs considered, along with the Monte Carlo mean and standard deviation of the test statistic under all scenarios considered.[8]

In line with the asymptotic properties of the two estimators under the null hypothesis ($\rho = 1$), the size distortion of the test is always very small and not statistically different from zero. However, the test presents a slight tendency to over-reject.

As for power, results are presented in tabular form in Tables 3–6 for selected values of ρ, and are summarized graphically in Figures 1–4. In the baseline scenario ($\phi = 0$ and $\beta = 1$ for logit and ordered logit models, $\phi = 0.1$ and $\beta = 1$ for Poisson and Gaussian linear models), the empirical power of the test is very low for short panels ($T = 3$) but increases rapidly as the longitudinal dimension increases. The relationship between the empirical power and the autocorrelation coefficient is inversely U-shaped, with evidence of skewness for higher values of ρ. As expected, in the extreme case of no persistence ($\rho = 0$), α_{it} collapses to a Gaussian white noise and the test has no power at all.

An intuitive explanation for this behaviour is linked to the robustness of the pairwise conditional estimator to the violation of the time-invariant unobserved heterogeneity assumption. Indeed, as noted by Varin (2008), the maximand of a slightly misspecified pairwise log-likelihood may be closer to the true parameter value than that from a roughly specified high-dimensional log-likelihood. In our case, provided that $T > 2$, the pairwise conditional log-likelihood will be slightly misspecified when ρ is close to one (i.e., high persistence) and always "less misspecified" than the full conditional log-likelihood as T increases. Apart from marginal differences, this evidence is consistent across scenarios and model specifications, and is stronger when the sample size n is large.

Looking at the first variant scenario for logit and ordered logit models, that is when it is

[8] Notice that we have 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. Since we got comparable (qualitatively and quantitatively) evidence, we deliberately do not report simulation results for this case. This alternative output is available from the authors upon request.

imposed that a large portion of the total variability is explained by the true model for the latent outcome, there seems to be evidence of a slight increase in the empirical power, especially for the ordered logit case.

Finally, a much higher power is observed for the more realistic case when the regressor and the unit-specific effects are moderately correlated ($\phi = 0.5$). This evidence is stronger for the Poisson and the Gaussian linear models in which, except for the case of very short panels and low autoregressive coefficient, the empirical power of the test seems to be always greater than seventy percent.

4 Empirical illustration

In this Section, we illustrate our testing procedure through an empirical application to self-rated health status (SRHS) of the elderly American population based on data from the Health and Retirement Study (HRS), a longitudinal survey that interviews every two years a representative sample of over 26,000 Americans aged 50 and older.

4.1 The data

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 and containing all waves from 1992 to 2010 ($T = 10$). We restrict attention to the subset of $n = 4,094$ individuals who responded to all waves (this gives a total of 40,940 observations). SRHS is measured on a 5-point ordered scale (poor, fair, good, very good, excellent). We also transform the original variable into a binary indicator that is equal to one if SRHS is good or better, and is equal to zero otherwise. Our covariates include a set of socio-demographic characteristics (gender, age, education and ethnicity), the number of doctor visits and the body mass index (BMI). Definitions and summary statistics of all variables considered are presented in Table 7.[9]

Following Heiss (2008), we estimate logit and ordered logit models for SRHS in wave 10. In addition to our set of covariates we also include lagged values of SRHS (Table 8). This simple exercise gives an idea of the SRHS correlation pattern over a longer period of time and highlights two interesting points: (i) the coefficients on most lags are strongly statistically different from

[9] Notice that, due to a failure in the AR(1) ordered logit model convergence, we are forced to drop outliers in the distribution of BMI and doctor visits. We use the method of percentiles. Since it does not seem to be outliers in the left tail, we drop out only values > 99.9 percentile, losing 37 individuals (370 observations).

zero, and (ii) they get smaller in size as the lag length increases. This suggests a model where SRHS depends on unobserved "true" health status, which follows some time-series process with declining autocorrelation.[10]

4.2 Results

We consider two model specifications. The first (Model M1) includes as regressors a constant term, age (specified as age splines), BMI and the number of GP visits. The second (Model M2) adds to Model M1 a set of wave dummies.

Tables 9 reports the four sets of estimates which are used to construct the test statistic for the logit and the ordered logit models respectively. The top panel shows the full conditional maximum likelihood estimates for the two model specifications, whereas the bottom panel reports the corresponding pairwise estimates. The key point here is that, regardless of the model type and specification, we strongly reject the null hypothesis of time-invariant unobserved heterogeneity confirming the results in Heiss (2008) even for this longer release of the HRS panel.

Given these results, we estimate the latent AR(1) random-effects logit and ordered logit models proposed by Heiss (2008), to which we refer the reader interested in more details.[11] Tables 10 shows the estimates under the two model specifications in which we also include the same time-invariant socio-demographic covariates as in Heiss (2008) application. It is worth emphasizing that omitting these covariates in performing our test (Tables 9) does not affect its power since, being time-invariant, they are eliminated from the conditional likelihood function. The estimated ρ is close to 0.95, basically the same value found by Heiss (2008), and appears to be strongly statistically significant regardless of the model type and specification. Hence, 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 autocorrelation.

[10] See Heiss (2008) for a detailed discussion. An alternative to the approach in Heiss (2008) would be a random-effects model with state dependence. However, state dependence in SRHS is not very convincing from a theoretical point of view. In fact, lagged outcomes can causally affect current outcome in a model of female labor force participation (Hyslop, 1999), this causality is less plausible in the case of SRHS, as it implies that the simple perception of own health affects future true health status.

[11] The estimation is performed using the `arldv` Stata package produced by Florian Heiss. The likelihood function of this model does not have a closed-form solution, so estimation involves numerical integration. We use 50 integration points.

5 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 fixed effects estimators defined as maximand of full and pairwise conditional likelihood functions.

The finite-sample properties of the proposed test are investigated via a set of Monte Carlo experiments. Our results suggest that the test generally performs well, showing small size distortions and good power properties especially when $n > 1,000$ and $T > 5$ (common sample sizes 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 require assumptions on the distribution of the individual effects, (iii) it allows the individual effects to 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 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. The null hypothesis of time-invariant unobserved heterogeneity is rejected for both logit and ordered logit specifications of the model, thus confirming the results in Heiss (2008) while 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 decaying autocorrelation.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society Series B*, 32:283–301.

Arellano, M. and Bonhomme, S. (2011). Nonlinear panel data analysis. *Annual Review of Economics*, 3:395–424.

Baetschmann, G., Staub, K., and Winkelmann, R. (2011). Consistent estimation of the fixed effects ordered logit model. *IZA Discussion Papers*, 5443.

Bartolucci, F., Bacci, S., and Pennoni, F. (2011). Mixture latent autoregressive models for longitudinal data. *ArXiv e-prints*, 1108.1498.

Bartolucci, F. and Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent markov heterogeneity structure. *Journal of the American Statistical Association*, 104(486):816–831.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225–38.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press, Oxford, 2nd edition.

Hausman, J.and Hall, B. and Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52(4):909–938.

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

Heiss, F. (2008). Sequential numerical integration in nonlinear state space models for microeconometric panel data. *Journal of Applied Econometrics*, 23(3):373–389.

Holly, A. (1982). A remark on Hausman's specification test. *Econometrica*, 50(3):749–759.

Holly, A. and Monfort, A. (1986). Some useful equivalence properties of hausman's test. *Economics Letters*, 20(1):39–43.

Hsiao, C. (2005). *Analysis of Panel Data*. Cambridge University Press, New York.

Huque, F. and Katti, S. K. (1976). A note on maximum conditional likelihood estimators. *The Indian Journal of Statistics, series B*, 38(1):1–13.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67(6):1255–1294.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, 2nd edition.

McFadden, D. (1974). *Frontiers in Econometrics*, chapter Conditional logit analysis of qualitative choice behaviour. Academic Press, New York.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics*, 32(3):385–397.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, 53(5):1047–1070.

Pfanzagl, J. (1993). On the consistency of conditional maximum likelihood estimators. *Annals of the Institute of Statistical Mathematics*, 45(4):703–719.

Ruud, P. A. (1984). Tests of specification in econometrics. *Econometric Reviews*, 3(2):211–242.

Sartori, N. and Severini, T. A. (2004). Conditional likelihood inference in generalized linear mixed models. *Statistica Sinica*, 14(2):349–360.

Stowasser, T., Heiss, F., McFadden, D., and 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. *Advances in Statistical Analysis*, 92:1–28.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.

Verbeek, M. and Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 33(3):681–703.

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

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Xu, X. and Reid, N. (2012). On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference*, 141(9):3047–3054.

Table 1: Size analysis for logit and ordered logit models.

T	Logit						Ordered logit					
	n=1000			n=4000			n=1000			n=4000		
	mean	sd	power	mean	sd	power	mean	sd	power	mean	sd	power
	$\boldsymbol{\phi = 0}$ $\boldsymbol{\beta = 1}$											
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
	$\boldsymbol{\phi = 0}$ $\boldsymbol{\beta = 2}$											
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
	$\boldsymbol{\phi = 0.5}$ $\boldsymbol{\beta = 1}$											
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 linear models.

T	Poisson						Gaussian					
	n=1000			n=4000			n=1000			n=4000		
	mean	sd	power	mean	sd	power	mean	sd	power	mean	sd	power
	$\boldsymbol{\phi = 0.1}$ $\boldsymbol{\beta = 1}$											
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
	$\boldsymbol{\phi = 0.5}$ $\boldsymbol{\beta = 1}$											
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 logit model.

T	$\boldsymbol{\phi = 0}$ $\boldsymbol{\beta = 1}$						$\boldsymbol{\phi = 0}$ $\boldsymbol{\beta = 2}$						$\boldsymbol{\phi = 0.5}$ $\boldsymbol{\beta = 1}$					
	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			
	$\rho = 0.2$																	
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
	$\rho = 0.4$																	
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
	$\rho = 0.6$																	
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
	$\rho = 0.8$																	
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

Table 4: Power analysis for the ordered logit model.

T	$\boldsymbol{\phi = 0 \quad \beta = 1}$						$\boldsymbol{\phi = 0 \quad \beta = 2}$						$\boldsymbol{\phi = 0.5 \quad \beta = 1}$					
	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			
$\rho = 0.2$																		
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
$\rho = 0.4$																		
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
$\rho = 0.6$																		
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
$\rho = 0.8$																		
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	$\boldsymbol{\phi = 0.1} \quad \boldsymbol{\beta = 1}$						$\boldsymbol{\phi = 0.5} \quad \boldsymbol{\beta = 1}$					
	n=1000			n=4000			n=1000			n=4000		
	mean	sd	power	mean	sd	power	mean	sd	power	mean	sd	power
$\rho = 0.2$												
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
$\rho = 0.4$												
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
$\rho = 0.6$												
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
$\rho = 0.8$												
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

Table 6: Power analysis for the Gaussian linear model.

T	$\boldsymbol{\phi = 0.1 \quad \beta = 1}$						$\boldsymbol{\phi = 0.5 \quad \beta = 1}$					
	n=1000			n=4000			n=1000			n=4000		
	mean	sd	power	mean	sd	power	mean	sd	power	mean	sd	power
$\rho = 0.2$												
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
$\rho = 0.4$												
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
$\rho = 0.6$												
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
$\rho = 0.8$												
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

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

Variable	Description	Mean	Std. Dev.	Min.	Max.
SRHS	(1 poor; 2 fair; 3 good; 4 very good; 5 excellent)	3.471	1.038	1	5
SRHS (binary)	Dummy equal to 1 if SRHS > 2	0.827	0.378	0	1
Age	Age of the respondent (year)	65.198	6.983	50	93
Female	Dummy for female	0.519	0.5	0	1
High school	Dummy for high school (`raeduc` $= 3$)	0.336	0.472	0	1
Some college	Dummy for college degree (`raeduc` $= 4$)	0.203	0.402	0	1
College degree+	Dummy for higher education (`raeduc` $= 5$)	0.218	0.413	0	1
Non white	Dummy for hispanic and black	0.149	0.357	0	1
BMI	Body mass index	27.414	4.762	12.8	51.9
GP visits	Number of GP visits (last two years)	7.474	10.455	0	300

Table 8: Logit and ordered logit models for SRHS in wave 10 on past SRHS.

	logit		ologit	
Age	-0.043	***	-0.029	***
Female	0.045		0.043	
High school	0.192		0.018	
Some college	-0.007		-0.029	
College degree+	0.303	**	0.089	
Non white	-0.186		-0.021	
SRHS wave 9	1.632	***	0.870	***
SRHS wave 8	0.946	***	0.608	***
SRHS wave 7	0.709	***	0.305	***
SRHS wave 6	0.415	***	0.259	***
SRHS wave 5	0.325	**	0.122	**
SRHS wave 4	0.197		0.061	
SRHS wave 3	0.182		0.017	
SRHS wave 2	0.338	**	0.104	**
SRHS wave 1	0.108		0.022	
Constant	-2.366	***	-	
cut-off1	-		1.492	**
cut-off2	-		3.860	***
cut-off3	-		6.369	***
cut-off4	-		9.399	***
Obs	4094		4094	
Log-lik	-1447.37		-4324.98	

Significance levels: * $p < 10\%$; ** $p < 5\%$, *** $p < 1\%$

Table 9: Test implementation for logit and ordered logit models.

	Logit				**Ordered logit**			
	S1		**S2**		**S1**		**S2**	
Standard								
Age splines: 50+	-0.086	***	-0.086		-0.108	***	-0.113	
Age splines: 60+	0.040	**	0.041	*	0.047	***	0.039	***
Age splines: 70+	-0.070	***	-0.079	***	-0.044	***	-0.050	***
Age splines: 80+	-0.001		0.000		0.016		0.016	
BMI	-0.010		-0.010		-0.028	***	-0.028	***
GP visits	-0.038	***	-0.038	***	-0.041	***	-0.041	***
Pairwise								
Age splines: 50+	-0.108	***	-0.255	*	-0.118	***	-0.164	*
Age splines: 60+	0.061	**	0.059	*	0.044	***	0.022	
Age splines: 70+	-0.074	***	-0.107	***	-0.028	*	-0.057	***
Age splines: 80+	0.082		0.067		0.059		0.033	
BMI	-0.024	*	-0.028	*	-0.021	**	-0.024	**
GP visits	-0.029	***	-0.029	***	-0.031	***	-0.031	***
Wave dummies	No		Yes		No		Yes	
H_0 = time-invariant individual effects								
Test statistic	16.240		67.770		35.440		165.120	
P-value	0.012		0.000		0.000		0.000	

Significance levels: * $p < 10\%$; ** $p < 5\%$, *** $p < 1\%$

Table 10: AR(1) random-effects logit and ordered logit models (Heiss, 2008).

	Logit				Ordered logit			
	S1		**S2**		**S1**		**S2**	
Age splines: 50+	-0.096	***	-0.044	*	-0.117	***	-0.053	***
Age splines: 60+	0.047	**	0.057	**	0.054	***	0.044	***
Age splines: 70+	-0.082	***	-0.109	***	-0.043	***	-0.063	***
Age splines: 80+	-0.007		-0.020		0.004		-0.007	
BMI	-0.061	***	-0.059	***	-0.067	***	-0.065	***
GP visits	-0.052	***	-0.051	***	-0.044	***	-0.044	***
Female	0.302	**	0.360	***	0.126		0.179	**
High school	2.019	***	2.053	***	1.444	***	1.473	***
Some college	2.487	***	2.544	***	1.938	***	1.985	***
College degree+	3.747	***	3.790	***	2.752	***	2.791	***
Non white	-1.256	***	-1.253	***	-1.109	***	-1.105	***
Constant	5.302	***	4.929	***				
Wave dummies	No		Yes		No		Yes	
σ	3.314	***	3.337	***	2.743	***	2.758	***
ρ	0.9504	***	0.9496	***	0.9515	***	0.9510	***
Log-lik	-12334.71		-12291.41		-43653.07		-43533.11	

Significance levels: * $p < 10\%$; ** $p < 5\%$, *** $p < 1\%$

Figure 1: Power curves for the logit model.

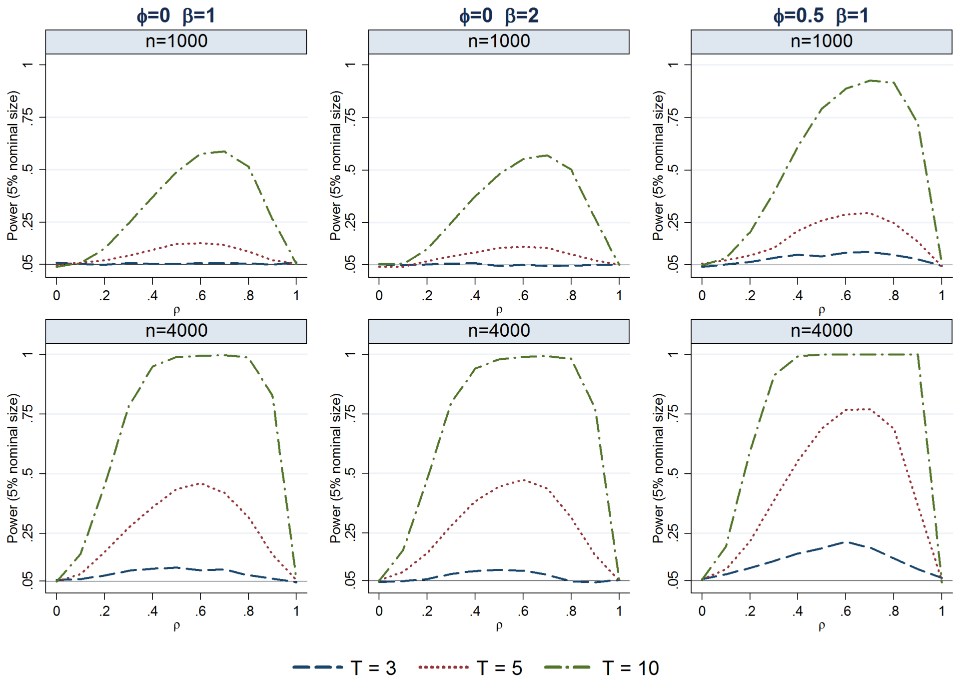


Figure 2: Power curves for the ordered logit model.

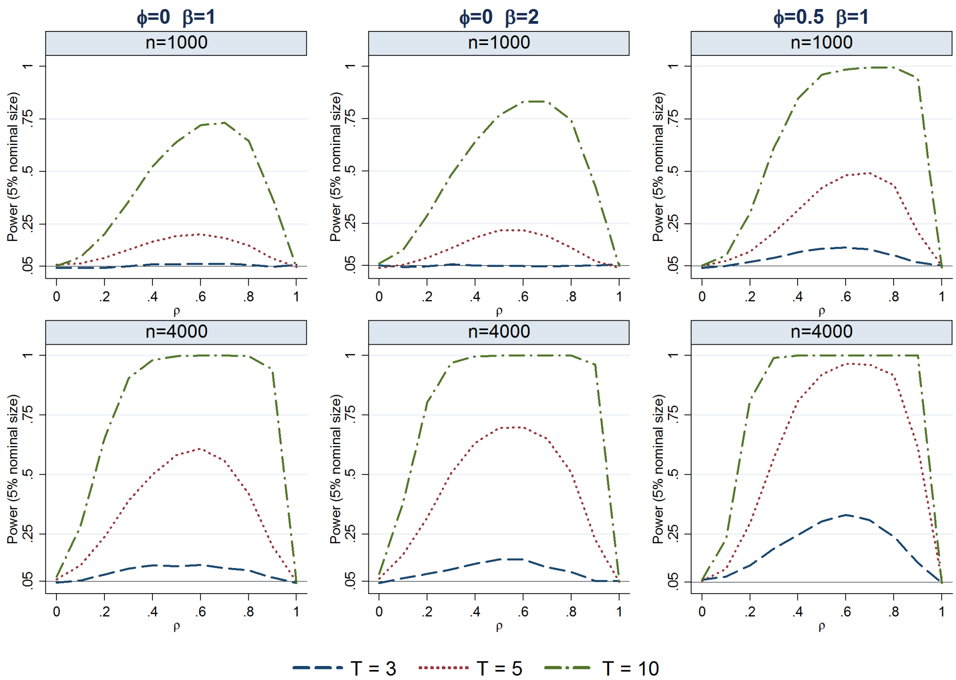


Figure 3: Power curves for the Poisson model.

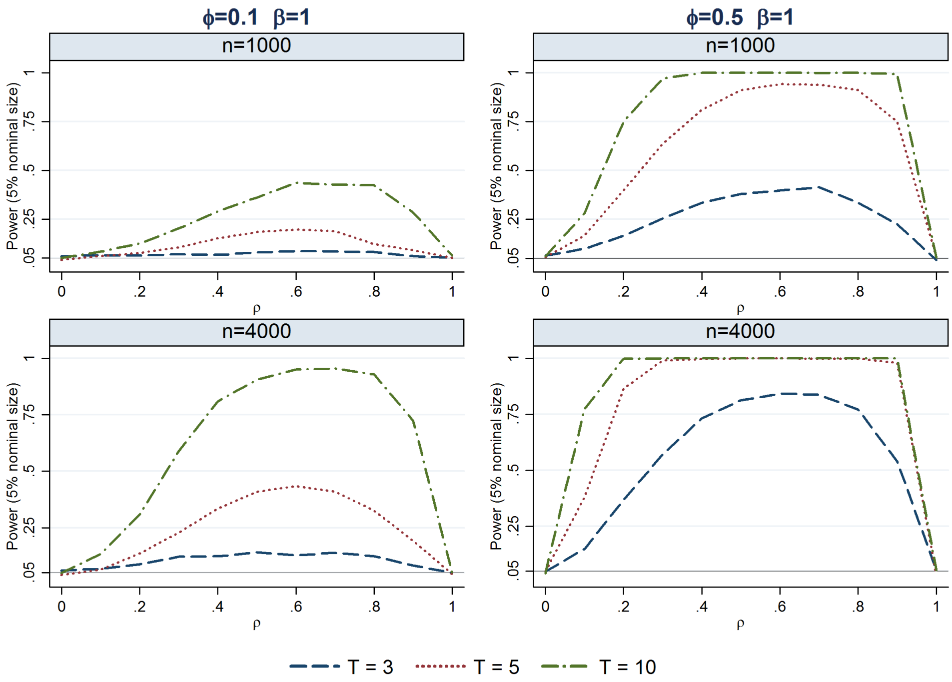
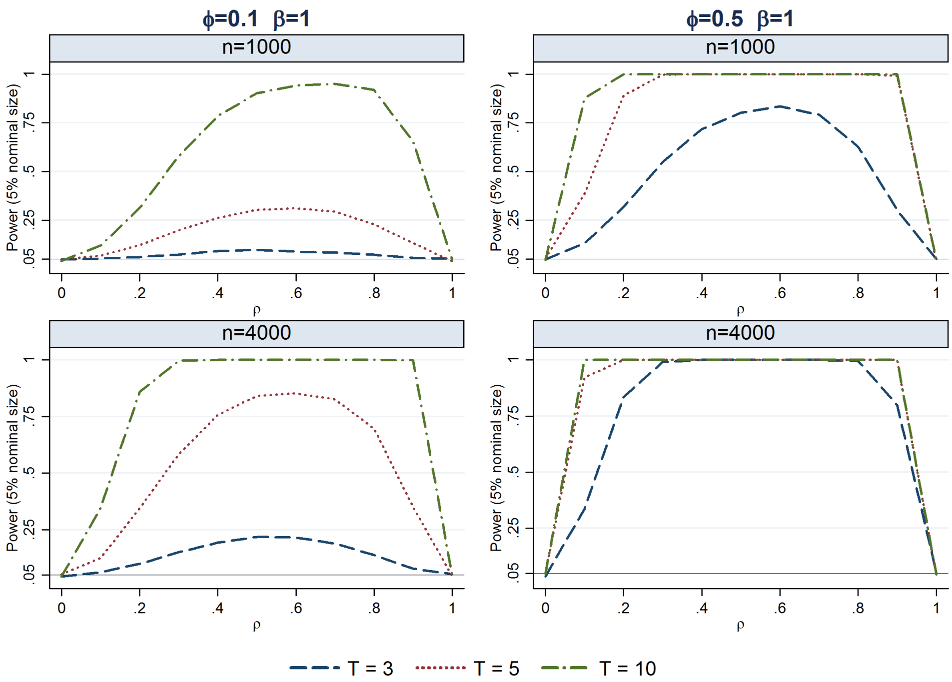


Figure 4: Power curves for the Gaussian linear model.



A Asymptotic properties of the pairwise conditional maximum likelihood estimator

In this appendix we provide formal arguments to establish the asymptotic properties (with $n \to \infty$ and fixed T) of the pairwise conditional maximum likelihood (PCML) estimator.

A.1 Asymptotics under H_0

The pairwise conditional log-likelihood function,

$$L_2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{t=2}^{T} \ln p_{\boldsymbol{\theta}}(y_{i,t-1}, y_{it} | y_{i,t-1} + y_{it}), \tag{A.1}$$

is a specific type of composite log-likelihood function formed by the sum of log-conditional probabilities where $p_{\boldsymbol{\theta}}(y_{i,t-1}, y_{it} | y_{i,t-1} + y_{it})$ is defined as in equation (5) but for each single pair of consecutive observations in the sample. Note that the parameter vector $\boldsymbol{\theta}$ has been added as a subscript to $p(.|.)$ in order to indicate that this probability depends on $\boldsymbol{\theta}$. As pointed out by Lindsay (1988), each component of $L_2(\boldsymbol{\theta})$ is a standard log-likelihood object. Since the asymptotic properties of the conditional maximum likelihood estimator have been proved elsewhere under different sets of assumptions (Andersen, 1970; McFadden, 1974; Huque and Katti, 1976; Pfanzagl, 1993), we discuss the asymptotic properties of the PCML estimator assuming consistency and asymptotic normality of the standard conditional maximum likelihood estimator defined for each single pair of consecutive observations in the sample.

Let us denote with $\boldsymbol{\theta}_0$ the true parameter vector and with $\hat{\boldsymbol{\theta}}$ a generic point in the parameter space Θ. Being the PCML estimator an M-estimator, the following theorem gives sufficient conditions that guarantee its consistency.

Theorem 1. *If (i) $\boldsymbol{\theta}_0$ is an interior point of a compact set Θ; (ii) the sequence of random functions $\{\hat{Q}_n\}$ converges in probability uniformly on Θ to a continuous function Q_0; (iii) Q_0 attains a unique maximum on Θ at $\boldsymbol{\theta}_0$; then $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ exists and is unique with probability approaching one as $n \to \infty$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} \xrightarrow{p} \boldsymbol{\theta}_0$.*

Proof: If we denote the CML estimator's criterion function for the subsample defined by a single pair of consecutive observations t and $t-1$ as

$$\hat{Q}_n^t(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \ln p_{\boldsymbol{\theta}}(y_{i,t-1}, y_{it} | y_{i,t-1} + y_{it}),$$

then the criterion function of the PCML estimator can be expressed as

$$\hat{Q}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\hat{Q}_n^t(\boldsymbol{\theta}).$$

When $p(.|.)$ is logistic, it can be show by standard asymptotic results that each $\hat{Q}_n^t(\boldsymbol{\theta})$ converges in probability uniformly to a continuous function $Q_0^t(\boldsymbol{\theta})$ defined as

$$Q_0^t(\boldsymbol{\theta}) = \mathbb{E}_0[\ln p_{\boldsymbol{\theta}}(y_{i,t-1}, y_{it}|y_{i,t-1} + y_{it})],$$

where $\mathbb{E}_0[.]$ denotes the expected value under the true model. Then,

$$\begin{aligned}
Q_0(\boldsymbol{\theta}) &= \mathbb{E}_0\left[\sum_{t=2}^{T}\ln p_{\boldsymbol{\theta}}(y_{i,t-1}, y_{it}|y_{i,t-1} + y_{it})\right] \\
&= \sum_{t=2}^{T}\mathbb{E}_0[\ln p_{\boldsymbol{\theta}}(y_{i,t-1}, y_{it}|y_{i,t-1} + y_{it})] \\
&= \sum_{t=2}^{T}Q_0^t(\boldsymbol{\theta}).
\end{aligned}$$

Therefore, by linearity of the expected value and the fact that each term in $\hat{Q}_n(\boldsymbol{\theta})$ converges in probability uniformly on Θ to its population counterpart, $\hat{Q}_n(\boldsymbol{\theta})$ converges in probability uniformly on Θ to $Q_0(\boldsymbol{\theta})$.

The third condition requires $\boldsymbol{\theta}_0$ to be the unique maximizer of $Q_0(\boldsymbol{\theta})$, so it suffices to show that $\boldsymbol{\theta}_0$ is the unique maximizer of $Q_0^t(\boldsymbol{\theta})$ for all $t = 2, \dots, T$. Since in a conditional maximum likelihood framework $\boldsymbol{\theta}_0$ is identified (and this is a sufficient condition for a unique maximum), under the same regularity conditions $\boldsymbol{\theta}_0$ uniquely maximizes each $Q_0^t(\boldsymbol{\theta})$, which complete the proof. $\square$

Theorem 2. *If the assumptions of the Theorem 1 are satisfied and each subsample CML estimator* $\sqrt{n}(\hat{\boldsymbol{\theta}}^{CMLE} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, J_0^{-1})$, *then* $\sqrt{n}(\hat{\boldsymbol{\theta}}^{PCMLE} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, A_0^{-1}B_0A_0^{-1})$.

Proof: The proof of this result is based on the mean value theorem assuming that all the $T-1$ subsample CML estimators are asymptotically normally distributed. Let us denote the $\hat{\boldsymbol{\theta}}^{PCMLE}$ estimator with $\tilde{\boldsymbol{\theta}}$. Since $\tilde{\boldsymbol{\theta}}$ is an extremum estimator and the conditional log-likelihood is twice differentiable, then

$$\sum_{i=1}^{n}\sum_{t=2}^{T}\nabla\ln p_{\tilde{\boldsymbol{\theta}}}(y_{i,t-1}, y_{it}|y_{i,t-1} + y_{it}) = 0,$$

where ∇ denotes the vector of first derivatives with respect to $\boldsymbol{\theta}$. By the mean value theorem,

$$\sum_{i=1}^{n}\sum_{t=2}^{T}\nabla \ln p_{\boldsymbol{\theta}_0}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it}) + \left[\sum_{i=1}^{n}\sum_{t=2}^{T}\nabla^2 \ln p_{\bar{\boldsymbol{\theta}}}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it})\right](\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = 0,$$

where $\bar{\boldsymbol{\theta}}$ is a point on the line joining $\tilde{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}_0$ and ∇^2 denotes the Hessian of the objective function. We can rearrange the previous equation as

$$\begin{aligned}\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = -&\left[\sum_{t=2}^{T}\frac{1}{n}\sum_{i=1}^{n}\nabla^2 \ln p_{\bar{\boldsymbol{\theta}}}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it})\right]^{-1}\times \\ &\times\sum_{t=2}^{T}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla \ln p_{\boldsymbol{\theta}_0}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it})\right].\end{aligned}$$

Since $\sqrt{n}(\hat{\boldsymbol{\theta}}^{CMLE}-\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, J_0^{-1})$ as $n\to\infty$, we have that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla \ln p_{\boldsymbol{\theta}_0}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, J_0)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\nabla^2 \ln p_{\bar{\boldsymbol{\theta}}}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it}) \xrightarrow{p} J_0^{-1}.$$

It then follows from Slutzky theorem and the closeness property of linear combination of Gaussian random variables that $\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ is asymptotically distributed as a Gaussian random variable with mean zero and asymptotic variance $A_0^{-1}B_0A_0^{-1}$, where

$$A_0 = -\sum_{t=2}^{T}\mathbb{E}_0\left[\nabla^2 \ln p_{\boldsymbol{\theta}_0}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it})\right],$$

and

$$B_0 = \mathbb{E}_0\left[\sum_{t=2}^{T}\nabla \ln p_{\boldsymbol{\theta}_0}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it})\right]\left[\sum_{t=2}^{T}\nabla \ln p_{\boldsymbol{\theta}_0}(y_{i,t-1}, y_{it}|y_{i,t-1}+y_{it})\right]'.$$

□