Testing for state dependence in binary panel data with individual covariates

Bartolucci, Francesco and Nigro, Valentina and Pigini, Claudia

11 July 2013

Testing for state dependence in binary panel data with individual covariates

Francesco Bartolucci
University of Perugia (IT)
bart@stat.unipg.it

Valentina Nigro
Bank of Italy (IT)
valentina.nigro@bancaditalia.it

Claudia Pigini
University of Perugia(IT)
pigini@stat.unipg.it

July 11, 2013

Abstract

We propose a test for state dependence in binary panel data under the dynamic logit model with individual covariates. For this aim, we rely on a quadratic exponential model in which the association between the response variables is accounted for in a different way with respect to more standard formulations. The level of association is measured by a single parameter that may be estimated by a conditional maximum likelihood approach. Under the dynamic logit model, the conditional estimator of this parameter converges to zero when the hypothesis of absence of state dependence is true. This allows us to implement a Wald test for this hypothesis which may be very simply performed and attains the nominal significance level under any structure of the individual covariates. Through an extensive simulation study, we find that our test has good finite sample properties and it is more robust to the presence of (autocorrelated) covariates in the model specification in comparison with other existing testing procedures for state dependence. The test is illustrated by an application based on data coming from the Panel Study of Income Dynamics.

Keywords: conditional inference, dynamic logit model, quadratic exponential model, Wald test

1 Introduction

In the analysis of panel data, a question of main interest is if the choice (or the condition) of an individual in the current period may influence his/her future choice (or condition), either directly (via the so–called "true state dependence") or through the presence of unobserved time-invariant heterogeneity; see Feller (1943) and Heckman (1981). Different policy consequences may derive from disentangling the individual unobserved heterogeneity from the true state dependence, where idiosyncratic shocks may last for a long time. In the case of binary panel data, a very relevant model in which the effects are disentangled is the dynamic logit model (Hsiao, 2005). This model includes individual-specific intercepts and, further to time-constant and/or time-varying individual covariates, the lagged response variable. In particular, the regression coefficient for the lagged response is a measure of the true state dependence.

A drawback of the dynamic logit model, with respect to the static logit model that does not include the lagged response among the covariates, is that simple sufficient statistics do not exist for the individual-specific intercepts. Therefore, conditional likelihood inference becomes more complex and may be performed only under certain conditions on the distribution of the covariates (Chamberlain, 1993; Honoré and Kyriazidou, 2000). We recall that the main advantage of this approach, as other fixed-effects approaches, is that it does not require to formulate any assumption on the distribution of the individual intercepts and on the correlation between these effects and the covariates; assumptions of this type are instead required within the random-effects approach.

In order to test for (true) state dependence, Halliday (2007) developed an interesting approach which does not require any distributional assumption for the individual-specific intercepts. However, this approach is explicitly formulated for the case of two time periods (further to an initial observation) and it cannot be easily applied for panel data with more than two periods. Moreover, it does not explicitly allow for individual covariates. In particular, when covariates are present, they can be accounted for only by splitting the overall sample in strata corresponding to different configurations of these covariates, but this makes the procedure more complex and its results depending on arbitrary choices. At least to our knowledge, no other approaches having a complexity comparable to that of Halliday (2007) exist in the literature for testing for state dependence.

In the econometric literature, Bartolucci and Nigro (2010) introduced a dynamic model, belonging to the quadratic exponential family (Cox, 1972), which may be used to analyze binary panel data; its parameters have an interpretation very similar to the parameters of the dynamic logit model. Moreover, under the model of Bartolucci and Nigro (2010), simple sufficient statistics for the the individual-specific intercepts exist which are the sums of the response variables at individual level. A conditional maximum likelihood estimator may be then used to estimate the model parameters. Moreover, Bartolucci and Nigro (2012) showed how to use this technique also for estimating the parameters of the dynamic logit model. The resulting pseudo conditional maximum likelihood (PCML) estimator has good performance in comparison with the method proposed by Honoré and Kyriazidou (2000) and Carro (2007) and related estimators. In particular, with respect to the estimator of Honoré and Kyriazidou (2000), the PCML estimator has the advantage of better exploiting the information in the sample and of allowing for aggregate time-covariates (such as time-dummies).

In this paper, we propose a test for state dependence based on a modified version of the quadratic exponential model of Bartolucci and Nigro (2010), which relies on a different formulation of the structure regarding the conditional association between the response variables given the individual-specific intercepts for the unobserved heterogeneity and the covariates. We show that the proposed model may be still represented as a latent index model where the errors are logistically distributed and the systematic part is formulated in terms of future expectations. This model may be estimated in a simple way by a conditional likelihood approach based on the same sufficient statistics as the original quadratic exponential model. Moreover, we show that, when data are generated from the dynamic logit model, the estimator of the parameter measuring the association between the response variables converges to zero in absence of state dependence even in the presence of covariates. It is then natural to test for state dependence on the basis of the proposed quadratic exponential model by a Wald test statistic.

The test we propose is directly comparable with the one of Halliday (2007) in terms of simplicity of implementation. Differently from Halliday's, our test may be used in the presence of individual covariates and for panel settings of length greater than two. In addition, we show that, in the special case of two time periods and no individual covariates, the two procedures employ the same information in the data to test for state dependence.

We compared the proposed test with that of Halliday (2007) through a deep simulation study. It turns out that in absence of individual covariates and with two time occasions, the two testing approaches perform very similarly in terms of significance level and power, but in the other cases the former clearly outperforms the latter. In particular, we show that ignoring the covariates, as Halliday (2007)'s test does, may lead to unsatisfactory finite–sample properties. Furthermore, the proposed test, which is based on the modified quadratic exponential model, proves to be more powerful than a Wald test based on more standard formulations of this model and a test directly based on the PCML estimator.

With the aim of illustrating the proposed test, we consider an application about the relation between women's employment and fertility, which is based on a dataset deriving from the Panel Study of Income Dynamics (PSID). This analysis is related to those proposed by Hyslop (1999) and Bartolucci and Farcomeni (2009).

The paper is organized as follows. In Section 2 we describe the dynamic logit model and the alternative quadratic exponential model of Bartolucci and Nigro (2010); for the purpose of our comparison, in the same section we also illustrate Halliday (2007)'s testing approach. In Section 3 we introduce the proposed Wald test for state dependence based on a new formulation of the quadratic exponential model. The empirical size and the power of this test are studied by simulation in Section 4. Finally, in Section 5 we provide the empirical illustration based on the PSID dataset. In the last section we draw the main conclusions.

We make our `R` implementation of all the algorithms illustrated in this paper, and in particular of the algorithm to perform the proposed test for state dependence, available to the reader upon request.

2 Preliminaries

For a panel of n subjects observed at T time occasions, let y_{it} denote the binary response variable for subject i at occasion t and let $\boldsymbol{x}_{it}$ denote the corresponding vector of individual covariates. Also let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ denote the vector of all outcomes for subject i and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1} \ \cdots \ \boldsymbol{x}_{iT})$ denote the matrix of all covariates for this subject.

In the following, we briefly review the dynamic logit model for these data and then the quadratic exponential model as an alternative model that includes a state dependence parameter. We also review the test for state dependence proposed by Halliday (2007).

2.1 Dynamic logit model

The dynamic logit model assumes that, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, the binary response y_{it} has conditional distribution

$$p(y_{it}|\alpha_i, \boldsymbol{X}_i, y_{i0}, \ldots, y_{i,t-1}) = p(y_{it}|\alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}) = \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]}{1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)}, \qquad (1)$$

where $\boldsymbol{\beta}$ and γ are the parameters of interest and the individual-specific intercepts α_i are often considered as nuisance parameters; moreover, the initial observation y_{i0} is given. Therefore, the joint probability of $\boldsymbol{y}_i$ given α_i, $\boldsymbol{X}_i$, and y_{i0} has expression

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \prod_t p(y_{it}|\alpha_i, \boldsymbol{x}_{it}, y_{i,t-1}) = \frac{\exp(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i*}\gamma)}{\prod_t[1 + \exp(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma)]},$$

where $y_{i+} = \sum_t y_{it}$ and $y_{i*} = \sum_t y_{i,t-1}y_{it}$, with the product $\prod_t$ and the sum $\sum_t$ ranging over $t = 1, \ldots, T$.

It is important to stress that γ measures the effect of the true state dependence and then the hypothesis on which we focus is $H_0 : \gamma = 0$, meaning absence of this form of dependence. The parameter γ may be identified and consistently estimated if the α_i parameters are properly taken into account. In particular, Chamberlain (1985) showed that a conditional approach, in the case of no covariates, may identify the state dependence parameter by suitable sufficient conditions for the α_i parameters. Along these lines, Honoré and Kyriazidou (2000) extended the conditional estimator including exogenous covariates in the model. Under particular conditions on the support of the covariates, they showed that, for $T = 3$, $\boldsymbol{y}_i$ is conditionally independent of α_i given the initial and the final observations of the response variable and that $y_{i1} + y_{i2} = 1$. Their estimator has the advantage, as in any fixed-effects estimator, to let α_i be freely correlated with the covariates in $\boldsymbol{X}_i$ and it does not require to formulate any assumption on the distribution of these effects. One of its characteristics is that a weight is attached to each observation; this weight depends on the covariates through a kernel function which reduces the rate of convergence of the estimator, which is slower than $\sqrt{n}$. Also due to this condition, the sample size substantially shrinks lowering the overall efficiency of the estimator. Moreover, this approach does not allow for time-dummies or trend variables and may be applied only with $T > 2$, beyond the initial observation.

A different fixed-effects approach is based on bias corrected estimators; see Hahn and Newey (2004), Carro (2007), Fernandez-Val (2009), and Hahn and Kuersteiner (2011). They are only consistent as $T \to \infty$ but have a reduced order of bias and they remain asymptotically efficient. For this reason, these estimators are shown to have good performance also in quite short panels.

2.2 Quadratic exponential model

The quadratic exponential model directly defines the joint probability of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$ and y_{i0}, and also given an individual-specific effect here denoted by δ_i, as follows:

$$p(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi} + y_{i*}\psi)}{\sum_{\boldsymbol{z}} \exp(z_+\delta_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + z_{i*}\psi)}, \tag{2}$$

where the sum $\sum_{\boldsymbol{z}}$ ranges over all the possible binary response vectors $\boldsymbol{z} = (z_1, \ldots, z_T)'$, $z_+ = \sum_t z_t$, and $z_{i*} = y_{i0}z_1 + \sum_{t>1} z_{t-1}z_t$. We refer to this model as QE1. Here we use a different notation from that used for the dynamic logit model, where the vector of regression coefficients is denoted by $\boldsymbol{\phi}$ and the state dependence parameter by ψ. These parameters are collected in the vector $\boldsymbol{\theta} = (\boldsymbol{\phi}', \psi)'$.

The model is a special case of that proposed by Bartolucci and Nigro (2010) with the parameters $\boldsymbol{\phi}$ assumed to be equal for all time occasions[1]. The same results of their paper apply straight to model (2); therefore, the conditional probability of y_{it} may be represented as

$$p(y_{it}|\delta_i, \boldsymbol{X}_i, y_{i0}, \ldots, y_{i,t-1}) = \frac{\exp\{y_{it}[\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi} + y_{i,t-1}\psi + e_t(\delta_i, \boldsymbol{X}_i)]\}}{1 + \exp[\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi} + y_{i,t-1}\psi + e_t(\delta_i, \boldsymbol{X}_i)]},$$

where

$$e_t(\delta_i, \boldsymbol{X}_i) = \log \frac{1 + \exp[\delta_i + \boldsymbol{x}_{i,t+1}'\boldsymbol{\phi} + e_{t+1}(\delta_i, \boldsymbol{X}_i) + \psi]}{1 + \exp[\delta_i + \boldsymbol{x}_{i,t+1}'\boldsymbol{\phi} + e_{t+1}(\delta_i, \boldsymbol{X}_i)]} = \log \frac{p_{t+1}(y_{i,t+1} = 0|\delta_i, \boldsymbol{X}_i, y_{it} = 0)}{p_{t+1}(y_{i,t+1} = 0|\delta_i, \boldsymbol{X}_i, y_{it} = 1)},$$

for $t = 1, \ldots, T-1$, with $e_T(\delta_i, \boldsymbol{X}_i) = 0$. This correction term may be interpreted as the effect of the current choice y_{it} on the expected utility of the future choice $y_{i,t+1}$. Furthermore, the quadratic exponential model shares with the dynamic logit the same interpretation of the state dependence parameter as log-odds ratio between any pair of response variables $(y_{i,t-1}, y_{it})$; moreover, y_{it} is conditionally independent of any other response variable given $y_{i,t-1}$ and $y_{i,t+1}$. Actually, when $\psi = 0$ this model coincides with

[1] Bartolucci and Nigro (2010) used a different parametrization for $t = T$ in order to approximate the probability in the last time period, where future covariates cannot be observed.

the static logit model, and this is an important point for the approach here proposed.

The main advantage of model QE1 defined above is the availability of simple sufficient statistics for the unobserved heterogeneity parameters. In particular, the sufficient statistic for each parameter α_i is y_{i+}. Therefore, a $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\phi}}', \hat{\psi})'$ may be derived maximizing a likelihood based on the conditional probability

$$p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}) = \frac{\exp\left(\sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi} + y_{i*}\psi\right)}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp\left(\sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + z_{i*}\psi\right)}, \tag{3}$$

where the sum $\sum_{\boldsymbol{z}:z_+=y_{i+}}$ is extended to all response configurations $\boldsymbol{z}$ with sum equal to y_{i+}. The estimator may be computed through a Newton-Raphson algorithm. Moreover, it allows for time-dummies and can be used even with $T = 2$.

Bartolucci and Nigro (2012) also showed that, up to a correction term, a quadratic exponential model of the type above may approximate the dynamic logit. On the basis of this result they derived a PCML estimator which is very competitive in terms of efficiency compared with the other estimators proposed in the econometric literature. In particular, this estimator typically results more efficient than the estimator of Honoré and Kyriazidou (2000) and it does not impose any conditions on the support of the covariates.

2.3 Available test for state dependence

Halliday (2007) proposed a test for state dependence allowing for the presence of aggregate time variables in a dynamic logit model of type (1). The proposed approach, which follows the lines of the conditional approach of Chamberlain (1985), is based on the construction of conditional probability inequalities that depend only on the sign of the state dependence parameter γ.

In the case of $T = 2$, Halliday (2007) considered the events $A_{i1} = \{y_{i0} = 1, y_{i1} = 1, y_{i2} = 0\}$ and $B_{i1} = \{y_{i0} = 0, y_{i1} = 1, y_{i2} = 0\}$ and he proved that

$$p(A_{i1}|\boldsymbol{X}_i, y_{i0} = 1, y_{i+} = 1) \geq p(B_{i1}|\boldsymbol{X}_i, y_{i0} = 0, y_{i+} = 1) \quad \text{for} \quad \gamma \geq 0$$

and

$$p(A_{i1}|\boldsymbol{X}_i, y_{i0}=1, y_{i+}=1) \leq p(B_{i1}|\boldsymbol{X}_i, y_{i0}=0, y_{i+}=1) \quad \text{for} \quad \gamma \leq 0,$$

under assumption (1). When $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$ are constant across subjects, and therefore there are only time-dummies common to all sample units, it is possible to consistently estimate $p_A = p(A_{i1}|\boldsymbol{X}_i, y_{i0}=1, y_{i+}=1)$ and $p_B = p(B_{i1}|\boldsymbol{X}_i, y_{i0}=0, y_{i+}=1)$ as follows

$$\hat{p}_A = \frac{\sum_i 1\{A_{i1}|\boldsymbol{X}_i\}}{\sum_i 1\{y_{i+}=1, y_{i0}=1|\boldsymbol{X}_i\}} = \frac{n_{110}}{m_1} \quad \text{and} \quad \hat{p}_B = \frac{\sum_i 1\{B_{i1}|\boldsymbol{X}_i\}}{\sum_i 1\{y_{i+}=1, y_{i0}=0|\boldsymbol{X}_i\}} = \frac{n_{010}}{m_0},$$

where $1\{\cdot\}$ is the indicator function, $n_{y_0y_1y_2}$ is the frequency of sample units with response configuration (y_0, y_1, y_2), and $m_{y_0} = n_{y_001} + n_{y_010}$. The test statistic for $H_0 : \gamma = 0$ is then defined as

$$S = \sqrt{n}\frac{\hat{p}_A - \hat{p}_B}{\hat{\sigma}(\hat{p}_A - \hat{p}_B)} \tag{4}$$

where $\hat{\sigma}(\hat{p}_A - \hat{p}_B)$ is the estimated standard deviation of the numerator. As the sample size grows to infinity, the test statistic converges to a standard normal distribution only under H_0; otherwise it diverges to $+\infty$ or to $-\infty$, as n grows to infinity, according to whether $\gamma > 0$ or $\gamma < 0$. It is worth noting that the test statistic exploits all the possible configurations of the response variable, conditionally on $y_{i+} = 1$; in fact, after some simple algebra, the numerator of (4) may be written as

$$\hat{p}_A - \hat{p}_B = \frac{n_{001}n_{110} - n_{101}n_{010}}{m_1 m_0}. \tag{5}$$

The method of Halliday (2007) identifies the sign of the state dependence parameter without estimating γ and avoiding distributional assumptions on the unobserved heterogeneity parameters. Nevertheless, this result cannot be easily generalized for $T > 2$. A possible solution may be using a multiple testing technique (Hochberg and Tamhane, 1987), where tests for all possible triples $(y_{i,t-1}, y_{it}, y_{i,t+1})$, $t = 1, \ldots, T-1$, are combined together. Furthermore, to take into account individual covariates that vary across individuals and/or time occasions, the test must to be performed for different configurations of the covariates. Consequently, the results may depend on how the covariate configurations are taken and may give a final ambiguous solution.

3 Proposed test for state dependence

In the following, we first illustrate a modified version of the quadratic exponential model QE1 outlined in Section 2.2 and then we show how to test for state dependence on the basis of the estimates of the parameters of this model.

3.1 Modified quadratic exponential model

To construct a test for state dependence we propose to modify the quadratic exponential model based on expression (2) in the following way:

$$\tilde{p}(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{y}_{i*}\psi)}{\sum_{\boldsymbol{z}} \exp(z_{i+}\delta_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{z}_{i*}\psi)}, \tag{6}$$

where $\tilde{y}_{i*} = \sum_t 1\{y_{it} = y_{i,t-1}\}$ and $\tilde{z}_{i*} = 1\{z_1 = y_{i0}\} + \sum_{t>1} 1\{z_t = z_{t-1}\}$, and where we use $\tilde{p}(\cdot|\cdot)$ instead of $p(\cdot|\cdot)$ for the probability function in order to avoid confusion, given that the two models use the same parameters; in particular, the parameters of interest are still collected in the vector $\boldsymbol{\theta}$. This model is referred to as QE2. The main difference between models QE1 and QE2 is in how the association between the response variables is modeled. Here this structure is based on the statistic $\tilde{y}_{i*}$ that, differently from y_{i*}, is equal to the number of consecutive pairs of outcomes which are equal each other, regardless if they are equal to 0 or 1.

Regarding the interpretation of model QE2 based on assumption (6), it useful to consider how this expression becomes after recursive marginalizations of the response variables in backward order. In particular, for $t = 1, \ldots, T-1$, we have that

$$\tilde{p}(y_{i1}, \ldots, y_{it}|\boldsymbol{X}_i, y_{i0}) = \frac{\exp(\sum_{h=1}^t y_{ih}\delta_i + \sum_{h=1}^t y_{ih}\boldsymbol{x}_{ih}'\boldsymbol{\phi} + \sum_{h=1}^t 1\{y_{ih} = y_{i,h-1}\}\psi)\tilde{g}_t(y_{it}, \delta_i, \boldsymbol{X}_i)}{\sum_{\boldsymbol{z}} \exp(z_{i+}\delta_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\phi} + \tilde{z}_{i*}\psi)},$$

where

$$\tilde{g}_t(y_{it},\delta_i,\boldsymbol{X}_i)=\tilde{g}_{t+1}(0,\delta_i,\boldsymbol{X}_i)\exp[(1-y_{it})\psi]+\tilde{g}_{t+1}(1,\delta_i,\boldsymbol{X}_i)\exp(\delta_i+\boldsymbol{x}'_{i,t+1}\boldsymbol{\phi}+y_{it}\psi),$$

with $\tilde{g}_T(y_{iT},\delta_i,\boldsymbol{X}_i)=1$. Consequently, for $t=1,\dots,T$, we have that

$$\log\frac{\tilde{p}(y_{i1},\dots,y_{i,t-1},y_{it}=1|\boldsymbol{X}_i,y_{i0})}{\tilde{p}(y_{i1},\dots,y_{i,t-1},y_{it}=0|\boldsymbol{X}_i,y_{i0})}=\delta_i+\boldsymbol{x}'_{it}\boldsymbol{\phi}+(2y_{i,t-1}-1)\psi+\tilde{e}_t(\delta_i,\boldsymbol{X}_i),$$

where

$$\begin{aligned}\tilde{e}_t(\delta_i,\boldsymbol{X}_i) &= \log\frac{\tilde{g}_t(1,\delta_i,\boldsymbol{X}_i)}{\tilde{g}_t(0,\delta_i,\boldsymbol{X}_i)}=\log\frac{\tilde{g}_{t+1}(0,\delta_i,\boldsymbol{X}_i)+\tilde{g}_{t+1}(1,\delta_i,\boldsymbol{X}_i)\exp(\delta_i+\boldsymbol{x}'_{i,t+1}\boldsymbol{\phi}+\psi)}{\tilde{g}_{t+1}(0,\delta_i,\boldsymbol{X}_i)\exp(\psi)+\tilde{g}_{t+1}(1,\delta_i,\boldsymbol{X}_i)\exp(\delta_i+\boldsymbol{x}'_{i,t+1}\boldsymbol{\phi})}\\ &= \log\frac{1+\exp[\delta_i+\boldsymbol{x}'_{i,t+1}\boldsymbol{\phi}+\psi+e_{t+1}(\delta_i,\boldsymbol{X}_i)]}{\exp(\psi)+\exp[\delta_i+\boldsymbol{x}'_{i,t+1}\boldsymbol{\phi}+\tilde{e}_{t+1}(\delta_i,\boldsymbol{X}_i)]},\end{aligned}\tag{7}$$

for $t=1,\dots,T-1$, with $\tilde{e}_T(\delta_i,\boldsymbol{X}_i)=0$. For $t=T$, this implies that

$$\tilde{p}(y_{it}|\delta_i,\boldsymbol{X}_i,y_{i,t-1})=\frac{\exp\{y_t[\delta_i+\boldsymbol{x}'_{it}\boldsymbol{\phi}+(2y_{i,t-1}-1)\psi]}{1+\exp[\delta_i+\boldsymbol{x}'_{it}\boldsymbol{\phi}+(2y_{i,t-1}-1)\psi]};$$

this expression may be seen as a reparametrization of the probability expression holding under the dynamic logit model (1). For $t=1,\dots,T-1$, instead, we have

$$\tilde{p}(y_{it}|\delta_i,\boldsymbol{X}_i,y_{i,t-1})=\frac{\exp\{y_t[\delta_i+\boldsymbol{x}'_{it}\boldsymbol{\phi}+(2y_{i,t-1}-1)\psi-\psi+\tilde{e}_t(\delta_i,\boldsymbol{X}_i)]}{1+\exp[\delta_i+\boldsymbol{x}'_{it}\boldsymbol{\phi}+(2y_{i,t-1}-1)\psi-\psi+\tilde{e}_t(\delta_i,\boldsymbol{X}_i)]}.$$

Regarding the interpretation of the last expression, first of all consider that for $\psi=0$ definition (7) implies that $\tilde{e}_t(\delta_i,\boldsymbol{X}_i)=0$ and then we have again a reparametrization of the dynamic logit model. Moreover, it may be easily proved that

$$\tilde{e}_t(\delta_i,\boldsymbol{X}_i)=\log\frac{\tilde{p}(y_{i,t+1}=0|\delta_i,\boldsymbol{X}_i,y_{it}=0)}{\tilde{p}(y_{i,t+1}=0|\delta_i,\boldsymbol{X}_i,y_{it}=1)}.$$

This correction term depends on the data only through $\boldsymbol{x}_{i,t+1},\dots,\boldsymbol{x}_{iT}$ and has a similar interpretation in terms of the probability of future choices as model QE1 and the quadratic exponential model of Bartolucci and Nigro (2010).

Moreover, model QE2 reproduces the same conditional independence relations of the dynamic logit model between the response variable y_{it} and $y_{i0},\dots,y_{i,t-2},y_{i,t+2},\dots,y_{iT}$, given α_i, $\boldsymbol{X}_i$, $y_{i,t-1}$, and $y_{i,t+1}$ $(t=2,\dots,T-1)$.

Finally, we have

$$\log\frac{\tilde{p}(y_{it}=1|\delta_i,\boldsymbol{X}_i,y_{i,t-1}=1)}{\tilde{p}(y_{it}=0|\delta_i,\boldsymbol{X}_i,y_{i,t-1}=1)}\frac{\tilde{p}(y_{it}=0|\delta_i,\boldsymbol{X}_i,y_{i,t-1}=0)}{\tilde{p}(y_{it}=1|\delta_i,\boldsymbol{X}_i,y_{i,t-1}=0)}=2\psi,\quad i=1,\dots,n,\ t=1,\dots,T,$$

meaning that the log-odds ratio between every consecutive pair of response variables has the same sign of ψ and it is equal to 0 if there is no state dependence.

3.2 Model estimation and testing for state dependence

As for model QE1, the sums of the response variables at the individual level, y_{i+}, are sufficient statistics for the individual-specific intercepts δ_i. Conditioning on the sum of the response variables, we obtain for model QE2 the following conditional probability function:

$$\tilde{p}(\boldsymbol{y}_i|\boldsymbol{X}_i,y_{i0},y_{i+})=\frac{\exp(\sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\phi}+\tilde{y}_{i*}\psi)}{\sum_{\boldsymbol{z}:z_+=y_{i+}}\exp(\sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\phi}+\tilde{z}_{i*}\psi)},\tag{8}$$

which directly compares with (3). For making inference on the state dependence parameter ψ, the new quadratic exponential model allows us to use a larger set of information with respect to the initial model QE1. This issue will be discussed in more detail at the end of this session.

On the basis of expression (8), we obtain the conditional log-likelihood

$$\tilde{\ell}(\boldsymbol{\theta})=\sum_i 1\{0<y_{i+}<T\}\tilde{\ell}_i(\boldsymbol{\theta}),\tag{9}$$

where

$$\begin{aligned}\tilde{\ell}_i(\boldsymbol{\theta}) &= \log\tilde{p}(\boldsymbol{y}_i|\boldsymbol{X}_i,y_{i0},y_{i+})\\ &= \sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\phi}+\tilde{y}_{i*}\psi-\log\sum_{\boldsymbol{z}:z_+=y_{i+}}\exp\left(\sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\phi}+\tilde{z}_{i*}\psi\right)\end{aligned}\tag{10}$$

is the individual contribution to the conditional log-likelihood. Note that the response configurations with y_{i+} equal to 0 or T do not contribute to the likelihood.

Function $\tilde{\ell}(\boldsymbol{\theta})$ may by maximized by a Newton–Raphson algorithm in a similar way as for model QE1, using the score vector and the information matrix reported below; see also Bartolucci and Nigro (2010). In this regard, it is convenient to write

$$\tilde{\ell}_i(\boldsymbol{\theta}) = \tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)'\boldsymbol{\theta} - \log \sum_{\boldsymbol{z}:z_+=y_{i+}} \exp[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z})'\boldsymbol{\theta}],$$

with

$$\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) = \left(\sum_t y_{it}\boldsymbol{x}_{it}', \tilde{y}_{i*}\right)',$$

so that, using the standard theory about the regular exponential family, we have the following expressions for the score for $\tilde{\ell}(\boldsymbol{\theta})$:

$$\begin{aligned} \tilde{\boldsymbol{s}}(\boldsymbol{\theta}) &= \boldsymbol{\nabla}_{\boldsymbol{\theta}}\, \tilde{\ell}(\boldsymbol{\theta}) = \sum_i \tilde{\boldsymbol{s}}_i(\boldsymbol{\theta}), \\ \tilde{\boldsymbol{s}}_i(\boldsymbol{\theta}) &= \tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i) - \tilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}]. \end{aligned}$$

Regarding the observed information matrix we have

$$\tilde{\boldsymbol{J}}(\boldsymbol{\theta}) = -\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\, \tilde{\ell}(\boldsymbol{\theta}) = \sum_i \tilde{V}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}]. \qquad (11)$$

In these expressions, $\tilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}]$ denotes the conditional expected value of $\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)$ given $\boldsymbol{X}_i$ and y_{i+} under model QE2, whereas the corresponding conditional variance is denoted by $\tilde{V}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}]$. These are given by

$$\begin{aligned} \tilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}] &= \sum_{\boldsymbol{z}:z_+=y_{i+}} \tilde{p}(\boldsymbol{z}|\boldsymbol{X}_i, y_{i0}, y_{i+})\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z}) \\ \tilde{\mathrm{V}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}] &= \sum_{\boldsymbol{z}:z_+=y_{i+}} \tilde{p}(\boldsymbol{z}|\boldsymbol{X}_i, y_{i0}, y_{i+})\tilde{\boldsymbol{d}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z})\tilde{\boldsymbol{d}}(\boldsymbol{X}, y_{i0}, \boldsymbol{z})', \end{aligned}$$

with $\tilde{\boldsymbol{d}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z}) = \tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{z}) - \tilde{\mathrm{E}}_{\boldsymbol{\theta}}[\tilde{\boldsymbol{u}}(\boldsymbol{X}_i, y_{i0}, \boldsymbol{y}_i)|\boldsymbol{X}_i, y_{i0}, y_{i+}]$. Note that $\tilde{\boldsymbol{J}}(\boldsymbol{\theta})$ is always non-negative definite since it corresponds to the sum of a series of variance-covariance matrices and therefore $\tilde{\ell}(\boldsymbol{\theta})$ is always concave. Moreover, a necessary condition for the Information matrix $\tilde{\boldsymbol{J}}(\boldsymbol{\theta})$ to be non-singular, and then for $\tilde{\ell}(\boldsymbol{\theta})$ to be strictly concave, is that time-constant covariates are ruled out, as it happens in any fixed-effect approach.

The conditional maximum likelihood estimator of $\boldsymbol{\theta}$ based on the maximization of (9) is denoted by $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\phi}}', \tilde{\psi})'$.

Once the parameters of the proposed quadratic exponential model are estimated, it is straightforward to construct a standard Wald statistic for testing that $\psi = 0$, as follows:

$$W = \frac{\tilde{\psi}}{\mathrm{se}(\tilde{\psi})}, \tag{12}$$

where se$(\cdot)$ is the standard error which is estimated by the sandwich formula reported in Appendix. It is worth noting that when $\gamma = 0$, the dynamic logit model corresponds to the proposed quadratic exponential model QE2 when $\psi = 0$. This implies the following main result.

Proposition 1 *Under the dynamic logit model and if the null hypothesis $H_0 : \gamma = 0$ holds, the test statistic W defined in (12) has asymptotic standard normal distribution as $n \to \infty$.*

Moreover, if data are generated from the dynamic logit model but $\gamma \neq 0$, then the value of W is expected to diverge to $+\infty$ or $-\infty$ according to whether the true value of γ is positive or negative. This is because, as we show at the end of Section 3.1, the sign of ψ is the same of the log-odds ratio between pairs of consecutive response variables and the latter is equal to γ under the dynamic logit model. Therefore, within the proposed approach we reject H_0 against the unidirectional alternative $H_1 : \gamma > 0$ at the significance level α if the observed value of W is greater than z_α, where z_α is the 100α-th upper percentile of the standard normal distribution. Similarly, we reject H_0 against $H_1 : \gamma < 0$ if the observed value of W is smaller than $-z_\alpha$ and we reject H_0 against the bidirectional alternative $H_1 : \gamma \neq 0$ if the observed value of $|W|$ is greater than $z_{\alpha/2}$.

In the following, we illustrate the case with $T = 2$ and only time dummies considered by Halliday (2007) and illustrated in Section 2.3. In this case, the conditional log-likelihood of model QE2, as defined in (9), allows us to identify two parameters, that is, ϕ corresponding to the difference between the effect to the two time-dummies and ψ

for the state dependence, so that $\boldsymbol{\theta} = (\phi, \psi)'$. Moreover, after some simple algebra we have

$$\begin{aligned}\tilde{\ell}(\boldsymbol{\theta}) &= \sum_i 1\Big\{y_{i+} = 1\}\{y_{i2}\phi + (1\{y_{i1} = y_{i0}\} + 1\{y_{i2} = y_{i1}\})\psi \\ &- \log \sum_{\boldsymbol{z}:z_+ = y_{i+}} \exp[z_2\phi + (1\{z_1 = y_{i0}\} + 1\{z_2 = z_1\}\psi)\Big\},\end{aligned}$$

that, in terms of sample frequencies, may be expressed as

$$\tilde{\ell}(\boldsymbol{\theta}) = (n_{001} + n_{101})\phi + (n_{001} + n_{110})\psi - m_0 \log k(0) - m_1 \log k(1),$$

where $k(y_{i0}) = \exp[\phi + (1 - y_{i0})\psi] + \exp(y_{i0}\psi)$. Consequently, the score function is

$$\begin{aligned}\tilde{\boldsymbol{s}}(\boldsymbol{\theta}) &= \sum_i 1\{y_{i+} = 1\} \begin{pmatrix} y_{i2} - \frac{\exp[\phi + (1-y_{i0})\psi]}{k(y_{i0})} \\ 1\{y_{i1} = y_{i0}\} - \frac{1\{y_{i1}=y_{i0}\}\exp[(1-y_{i0})\phi + 1\{y_{i1}=y_{i0}\}\psi]}{k(y_{i0})} \end{pmatrix} \\ &= \begin{pmatrix} n_{001} + n_{101} - \frac{m_0 \exp(\phi+\psi)}{k(0)} - \frac{m_1 \exp(\phi)}{k(1)} \\ n_{001} + n_{110} - \frac{m_0 \exp(\phi+\psi)}{k(0)} - \frac{m_1 \exp(\psi)}{k(1)} \end{pmatrix}.\end{aligned}$$

In order to solve the system of two equations $\tilde{\boldsymbol{s}}(\boldsymbol{\theta}) = \boldsymbol{0}$, we initially subtract the first equation from the second and, after some algebra, we obtain

$$\exp(\psi) = \frac{n_{110}}{n_{101}} \exp(\phi). \tag{13}$$

We then substitute this result in the first equation obtaining

$$\hat{\phi} = \frac{1}{2} \log \frac{n_{001} n_{101}}{n_{010} n_{110}}.$$

Finally, by substituting this solution in (13), we have

$$\hat{\psi} = \frac{1}{2} \log \frac{n_{001} n_{110}}{n_{010} n_{101}}.$$

This result shows that our test statistic is based on the same response variable configurations of Halliday's test statistic in expression (5) and then it exploits the same information.

Moreover, the two test statistics always exhibit the same sign since

$$\text{sign}(\tilde{\psi}) = \text{sign}[\log(n_{001}n_{110}) - \log(n_{010}n_{101})] = \text{sign}(n_{001}n_{110} - n_{010}n_{101}) = \text{sign}(\hat{p}_A - \hat{p}_B),$$

where $\hat{p}_A - \hat{p}_B$ in the numerator of Halliday's test statistic S defined in (4). This confirms that our estimator $\tilde{\psi}$ identifies the sign of the state dependence parameter γ under the dynamic logit model. A consequence is that the proposed test statistic W and the test statistic S have the same asymptotic distribution with mean centered in 0 under the dynamic logit model when $H_0 : \gamma = 0$ holds; both test statistics diverge to $+\infty$ or $-\infty$ under the dynamic logit model with $\gamma \neq 0$ (the first case when the true value of γ is positive and the second when it is negative). This is in agreement with the similar performance of the tests based on the two statistics, S and W, in terms of actual size and power that we note in the simulation study (see Section 4) when $T = 2$ and in absence of individual covariate.

A relevant issue is if the same Wald test as above may be based on the initial quadratic exponential model QE1, using a statistic of type $\hat{\psi}/se(\hat{\psi})$, as also this model is equal to the static logit model when $\psi = 0$. Our conjecture is that this test is less powerful than the test based on the test statistic W defined above, since the latter is based on a version of the quadratic exponential model, the estimator of which better exploits the information about the association between the response variables. In order to illustrate this point, we consider the simple case in which there are two time occasions, no covariates, and no time-dummies. In this case it is possible to prove that the conditional likelihood estimator of ψ under model QE1 is equal to

$$\hat{\psi} = \log \frac{n_{110}}{n_{101}}$$

in terms of sample frequencies, whereas, under model QE2, the estimator of this parameter is equal to

$$\tilde{\psi} = \frac{n_{001} + n_{110}}{n_{010} + n_{101}}.$$

Both estimators converge in probability to the true value of γ under the dynamic logit model if $H_0 : \gamma = 0$ holds. However, the first estimator exploits a reduced amount of information with respect to the second, as it ignores the response configurations $n_{y_0 y_1 y_2}$ with $y_0 = 0$; this also happens in more complex situations. Consequently the test based on estimator $\hat{\psi}$ (QE1) attains a reduced power than the one based on the estimator $\tilde{\psi}$ (QE2) when the true value of γ is different from 0. As will be shown in Section 4.2, this different behavior is also confirmed by the simulation results. A related point is how the proposed test compares with a Wald test based on one of the fixed-effects estimators for the dynamic logit model, as the PCML estimator proposed by Bartolucci and Nigro (2012). Since this estimator is based on model QE1, as an approximating model, we expect a similar difference in terms of power with respect to the proposed test for state dependence.

4 Simulation study

In order to study the finite–sample properties of the Wald test for state dependence proposed in Section 3, we performed a comprehensive Monte Carlo experiment based on a simulation design similar to the one adopted by Honoré and Kyriazidou (2000).

4.1 Design

In particular, we generated samples from a dynamic logit model where the conditional mean specification includes individual-specific intercepts, one covariate, and the lag of the response variable, as follows:

$$y_{it} = 1\left\{\alpha_i + x_{it}\beta + y_{i,t-1}\gamma + \varepsilon_{it} \geq 0\right\},$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, with initial condition

$$y_{i0} = 1\left\{\alpha_i + x_{i0}\beta + \varepsilon_{i0} \geq 0\right\}, \; i = 1, \ldots, n.$$

The error terms ε_{it} are independent, and have zero–mean logistic distribution with variance $\pi^2/3$. For $T \geq 2$, the individual intercepts α_i are defined as $\alpha_i = \frac{1}{3}\sum_{t=0}^{2} x_{it}$, where the covariate x_{it} is generated as

$$
\begin{aligned}
x_{i0} &\sim N\left(0, \pi^2/3\right), \\
x_{it} &= x_{i,t-1}\rho + u_{it}, \\
u_{it} &\sim N\left(0, \left(1-\rho^2\right)\pi^2/3\right),
\end{aligned}
$$

so that x_{it} and ε_{it} have the same stationary variance. In this way, the generating model admits a correlation between the covariates and the individual-specific intercepts and also it allows for an autocorrelation of the covariate for the same unit according to an AR(1) dependence structure. In particular, the covariate is autocorrelated if the parameter ρ is different from 0, whereas if ρ is equal to 0 we have a simulation design with independent covariate values at different occasions.

Based on the above generating model, we ran experiments for values of γ on a grid between -1.0 and 1.0 with step 0.1. The values of sample size we considered are 500 and 1000; we also considered panels of length $T = 2, 5$, $\beta = 0, 1$, and $\rho = 0.5$. The number of Monte Carlo replications was fixed at $1,000$.

We performed the proposed Wald test for state dependence based on model QE2 comparing its behavior with that of the test of Halliday (2007). An important feature of the latter is that it does not allow to take into account individual covariates. A possible solution is performing the test separately for subgroups of individuals (see Section 2.3). The problem is relevant when the covariates are autocorrelated, as it is reasonable to expect in standard economic applications. A procedure that ignores the presence of these explanatory variables may confound state dependence with the persistence that comes from the correlation of $y_{i,t-1}$ with x_{it}, as both depend on x_{it-1}. We, therefore, expect Halliday's test to exhibit rather wrong size properties in these circumstances.

Another issue is that, from Halliday (2007), it is not obvious how to test for state dependence when $T > 2$. In our simulation, we considered all the possible triples $(y_{i,t-1}, y_{it}, y_{i,t+1})$, $t = 1, \ldots, T-1$, computed Halliday's test for each of these triples and then decided when to reject $H_0 : \gamma = 0$ by a multiple testing technique (see Hochberg and Tamhane, 1987). In particular, H_0 is rejected if at least one of the p–values that are obtained from each of the $T-1$ triples of consecutive observations is smaller than the Bonferroni corrected nominal size. Such a correction ensures that the family-wise error rate is controlled for. For instance, if we test the null hypothesis for a nominal size of 0.05, the corrected nominal size is $1 - \sqrt[T-1]{0.95}$ for each single test.

4.2 Simulation results

Figure 1 depicts the power curves resulting from the simulation study illustrated in the previous section for the proposed Wald test based on statistic W defined in (12); and that proposed by Halliday (2007) and based on statistic S defined in (4). For a better comparison with the Halliday's test (see Section 2.3), we estimated model QE2 on which our test is based, including the covariate x_{it} and without this covariate. All rejection rates are displayed for a nominal size $\alpha = 0.05$ and considering the null hypothesis $H_0 : \gamma = 0$ against the bidirectional alternative hypothesis $H_1 : \gamma \neq 0$. For the Wald test statistic, the curve labelled by "test_cov" refers to the situation in which x_{it} is included in the model specification; the curve is labelled by "test_nocov" when this covariate is ignored. For certain relevant values of γ, Table 1 displays the rejection rate of this bidirectional test.

[Figure 1 about here.]

[Table 1 about here.]

For each approach, we also considered both lower and upper tailed tests (Figures 2 and 3) which are referred to $H_1 : \gamma < 0$ and $H_1 : \gamma > 0$, respectively.

[Figure 2 about here.]

[Figure 3 about here.]

The top panels of Figures 1–3 show that the proposed test has size equal to the nominal level α when $\beta = 0$ and $T = 2$ and with both sample sizes, whereas a sample size of at least 1000 is needed to exhibit satisfactory power properties. In these scenarios, Halliday's test statistic presents a behavior very similar to the proposed test. With $T = 5$ and $\beta = 0$, the rejection rate for the proposed test sensibly increases (see the second–row panels of Figures 1–3) reaching almost 100% for $|\gamma| = 0.6$ (with $n = 500$) and $|\gamma| = 0.4$ (with $n = 1000$). On the contrary, the generalization of Halliday's test statistic to cases with $T > 2$ leads to a remarkable power loss: with $n = 500$ and $|\gamma| = 0.5$ the rejection rate is about one half of that of the proposed test (see Table 1).

The third and fourth row–panels of Figures 1–3 provide an illustration of the simulation results with $\beta = 1$. While the proposed "test_cov" maintains its size properties, Halliday's test over-rejects the null hypothesis of absence of state dependence when this hypothesis is true. Moreover, the test size bias grows with the sample size: when $T = 2$, for example, it rises from 23% with $n = 500$ to 44% with $n = 1000$ (see Table 1). As expected, ignoring the presence of the covariate x_{it} in testing for state dependence leads to mistakenly detect a significant persistence in the dependent variable. This result is also confirmed by the rejection rates of "test_nocov" that exhibit the same behavior as Halliday's. Regarding the power, it decreases for all the three tests when $\beta = 1$ rather than $\beta = 0$. Nevertheless, our test shows a better performance in the case $\gamma < 0$. On the other hand, for $\gamma > 0$ and $T = 2$ this test has less power than Halliday's, which, however, confounds the positive autocorrelation in the covariate with a positive state dependence.

In conclusion, the above simulation results confirm our conjecture that in absence of individual covariates and when $T = 2$, the proposed Wald test for state dependence performs similarly, in terms of size and power, to the test proposed by Halliday (2007). Furthermore, in all other situations, our test is superior to the other one, mainly due to the fact that its nominal size level is attained with any $T \geq 2$ and in the presence of individual covariates.

A final point concerns performing a Wald test for state dependence which is based on the initial quadratic exponential model QE1 defined in (2). We recall that the main difference with the model here adopted for the test (QE2) is in the way the association between the response variables is accounted for. In this regard, Table 2 shows the Monte Carlo results for the test based on model QE1 and in particular on the estimator of ψ coming from the maximization of conditional likelihood (3).

[Table 2 about here.]

As expected, the power properties of the test based on the initial QE1 model are less satisfactory compared to those of the proposed test. This is due to an information loss: model QE1 only considers the information concerning pairs of consecutive observations such that $(y_{i,t-1} = 1, y_{it} = 1)$. Nevertheless, this test shows a better size with respect to Halliday's, in particular for $\beta = 1$.

5 Empirical application

In this section, we illustrate an application of the proposed testing procedure based on a dataset derived from the Panel Study of Income Dynamics[2]. Our example closely resembles the empirical analyses in Hyslop (1999) and in Bartolucci and Farcomeni (2009) that focus on the effect of fertility on women's employment and on the magnitude of the state dependence effect in both variables.

For the purposes of the present article, we restrict the analysis to testing for state dependence for women's employment and fertility. The dataset concerns $n = 1446$ married women between 18 and 46 years of age followed for $T = 5$ time occasions, from 1987 to 1992 (the first year of observation is taken as initial condition). The covariates in the model specification are: number of children in the family between 1 and 2 years of age

[2]The database is available at `http://psidonline.isr.umich.edu`

"child 1-2" and, similarly, "child 3-5", "child 6-14", "child 14-", "income" of the husband in dollars, time-dummies, "lagged employment", and "lagged fertility".

We computed the proposed Wald test statistic for $H_0 : \gamma = 0$, which is based on the modified the quadratic exponential model QE2 defined in (8), and the Halliday's test statistic defined in (4) for each triple $(y_{i,t-1}, y_{it}, y_{i,t+1})$, $t = 1, \ldots, T-1$, applying the Bonferroni corrected nominal size (see Section 4).

In the case the null hypothesis of no state dependence is rejected, it is necessary to estimate the model parameters in a dynamic framework. As illustrated in Section 2.2, suitable approaches are based on the conditional estimator of Honoré and Kyriazidou (2000), the PCML estimator (Bartolucci and Nigro, 2012), or biased-corrected fixed-effects estimators (Carro, 2007). If the null hypothesis is not rejected, we simply estimate a static logit model.

For the dataset here considered, Table 3 reports the test statistics computed by the different approaches using the complete sample of $n = 1446$ women. The proposed test strongly rejects the null hypothesis of absence of state dependence for both response variables employment and fertility. The signs of Halliday's test statistics also indicate positive state dependence for employment and negative for fertility. The values of these statistics are such that the null hypothesis is rejected: in both cases there is at least one p–value lower than the Bonferroni corrected nominal size.

[Table 3 about here.]

Since the null hypothesis of absence of state dependence is rejected, we estimated the dynamic logit model by the PCML estimator. The estimation results, reported in Table 4, confirm a strong positive state dependence for employment with an estimated coefficient $\hat{\gamma}$ close to 1.5 and a negative state dependence for fertility with $\hat{\gamma}$ equal to -0.9. However, in the second case, the value of the Wald test statistic based on this estimate is closer to zero than the value of the proposed Wald test statistic. The difference is that the PCML estimator is based on an approximating model having a structure similar to model QE1 and then it leads to a less powerful test with respect to the proposed approach based on model QE2.

[Table 4 about here.]

As discussed in Section 2.3, in order for Halliday's procedure to take into account observed heterogeneity in terms of individual covariates, the test must be performed separately for different configurations of such covariates. As an illustration, we performed this test for two different groups of women created on the basis of the covariate "education": we grouped observations of women with at most 12 years of schooling and those of women with more than 12 years of schooling (12 is the education median value). Table 5 shows the test statistics for the two groups: in both cases the positive state dependence in employment is detected by all the test statistics, as they reject H_0.

[Table 5 about here.]

On the contrary, Halliday's test does not seem to detect the negative state dependence in fertility for less educated women; this is an important difference with respect to our approach. For this subsample, the PCML estimate of the state dependence parameter is negative (equal to -0.14; see Table 6) but not significantly different form 0. Overall, in this case the higher power of the proposed testing approach emerges over both the Halliday's approach and the approach based on the PCML estimator.

Finally, for women with more than 12 years of schooling, there is agreement between all the tests which reject the null hypothesis of absence of state dependence in fertility. For this subsample, the PCML estimate of γ is equal to -1.3 (see Table 7).

[Table 6 about here.]

[Table 7 about here.]

6 Conclusions

In this paper, we propose a test for state dependence under the dynamic logit model with individual covariates. The test is based on a modified version of the quadratic exponential model proposed in Bartolucci and Nigro (2010) in order to exploit more information about the association between the response variables. We show that this model correctly identifies the presence of state dependence regardless of whether individual covariates are present or not.

Our test directly compares with the one proposed by Halliday (2007), which, however, cannot be easily applied in a panel with more than two periods (further to the initial observation) and does not allow for individual covariates. In the special case of two time periods and no covariates, the proposed test employs the same information on the response variables as Halliday's.

We studied the finite–sample properties of the Wald test for state dependence proposed in this paper by means of a comprehensive Monte Carlo experiment in which we also compare it with the test proposed by Halliday (2007). Simulation results show that the proposed test attains the nominal size even with not large samples (500 sample units), while it exhibits satisfactory power properties with large sample sizes. As expected, ignoring the presence of time-varying covariates in testing for state dependence leads to mistakenly detect a significant persistence in the response variable: the proposed test maintains its size properties, whereas Halliday's test over-rejects the true null hypothesis of absence of state dependence. Moreover, when state dependence is negative and the covariate positively affects the response variable, Halliday's test shows a remarkable power loss.

This result is confirmed by our empirical study based on a dataset derived from the Panel Study of Income Dynamics: when using either the whole sample or different subsamples, the proposed test always rejects the null hypothesis of absence of state dependence in fertility for parameter estimates of about -1, whereas Halliday's test may fail to detect the state dependence.

Overall, the main advantages of the proposed test are the simplicity of use and its flexibility. In fact, it can be very simply performed and does not require to formulate any parametric assumption on the distribution of the individual-specific intercepts (or on the correlation between these intercepts and the covariates) as random-effects approaches instead require. Moreover, it may be used even with only two time occasions (further to an initial observations) and with individual covariates of any nature (time-constant or time-varying), including time-dummies. It is also worth noting that, as in typical fixed-effects approaches, how the time-constant covariates affect the response variables needs not to be specified, as their effect is absorbed into to the individual-specific intercepts.

Finally, it is worth noting that the proposed test, being based on a modified quadratic exponential model, is more powerful than a Wald test based on more traditional quadratic exponential models or on the pseudo conditional likelihood estimator of Bartolucci and Nigro (2012). We noticed this aspect in the simulation study and in the empirical application.

Acknowledgments

Francesco Bartolucci and Claudia Pigini acknowledge the financial support from the grant RBFR12SHVV of the Italian Government (FIRB project "Mixture and latent variable models for causal inference and analysis of socio-economic data").

Appendix: the standard error of conditional maximum likelihood estimator

In order to derive an expression for the standard error used in the Wald test statistic, we rely on the sandwich estimator of White (1982). From the log-likelihood equation defined in (10), the variance covariance matrix of $\tilde{\boldsymbol{\theta}}$ is

$$\tilde{\boldsymbol{V}}(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{J}}(\tilde{\boldsymbol{\theta}})^{-1}\tilde{\boldsymbol{H}}(\tilde{\boldsymbol{\theta}})[\tilde{\boldsymbol{J}}(\tilde{\boldsymbol{\theta}})^{-1}]',$$

where

$$\tilde{\boldsymbol{H}}(\boldsymbol{\theta}) = \sum_i 1\{0 < y_{i+} < T\}\tilde{\boldsymbol{s}}_i(\boldsymbol{\theta})\tilde{\boldsymbol{s}}_i(\boldsymbol{\theta})'$$

and $\tilde{\boldsymbol{J}}(\boldsymbol{\theta})$ is the information matrix defined in (11). Once the matrix $\tilde{\boldsymbol{V}}(\tilde{\boldsymbol{\theta}})$ has been computed as above, the standard error for $\tilde{\psi}$ may be obtained in the usual way from the main diagonal of this matrix.

References

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.

Bartolucci, F. and Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. *Econometrica*, 78:719–733.

Bartolucci, F. and Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data. *Journal of Econometrics*, 170:102–116.

Carro, J. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, 140:503–528.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In Heckman, J. J. and Singer, B., editors, *Longitudinal analysis of labor market data*. Cambridge University Press: Cambridge.

Chamberlain, G. (1993). Feedback in panel data models. Technical report, Department of Economics, Harvard University.

Cox, D. (1972). The analysis of multivariate binary data. *Applied statistics*, 21:113–120.

Feller, W. (1943). On a general class of "contagious" distributions. *Annals of Mathematical Statistics*, 14:389–400.

Fernandez-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit model. *Journal of Econometrics*, 150:71–85.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27:1152–1191.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Halliday, T. H. (2007). Testing for state dependence with time-variant transition probabilities. *Econometric Reviews*, 26:685–703.

Heckman, J. (1981). Heterogeneity and state dependence. In Rosen, S., editor, *Studies in labor markets*, pages 91–140. Chicago University Press: Chicago.

Hochberg, Y. and Tamhane, A. (1987). *Multiple comparison procedures*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.

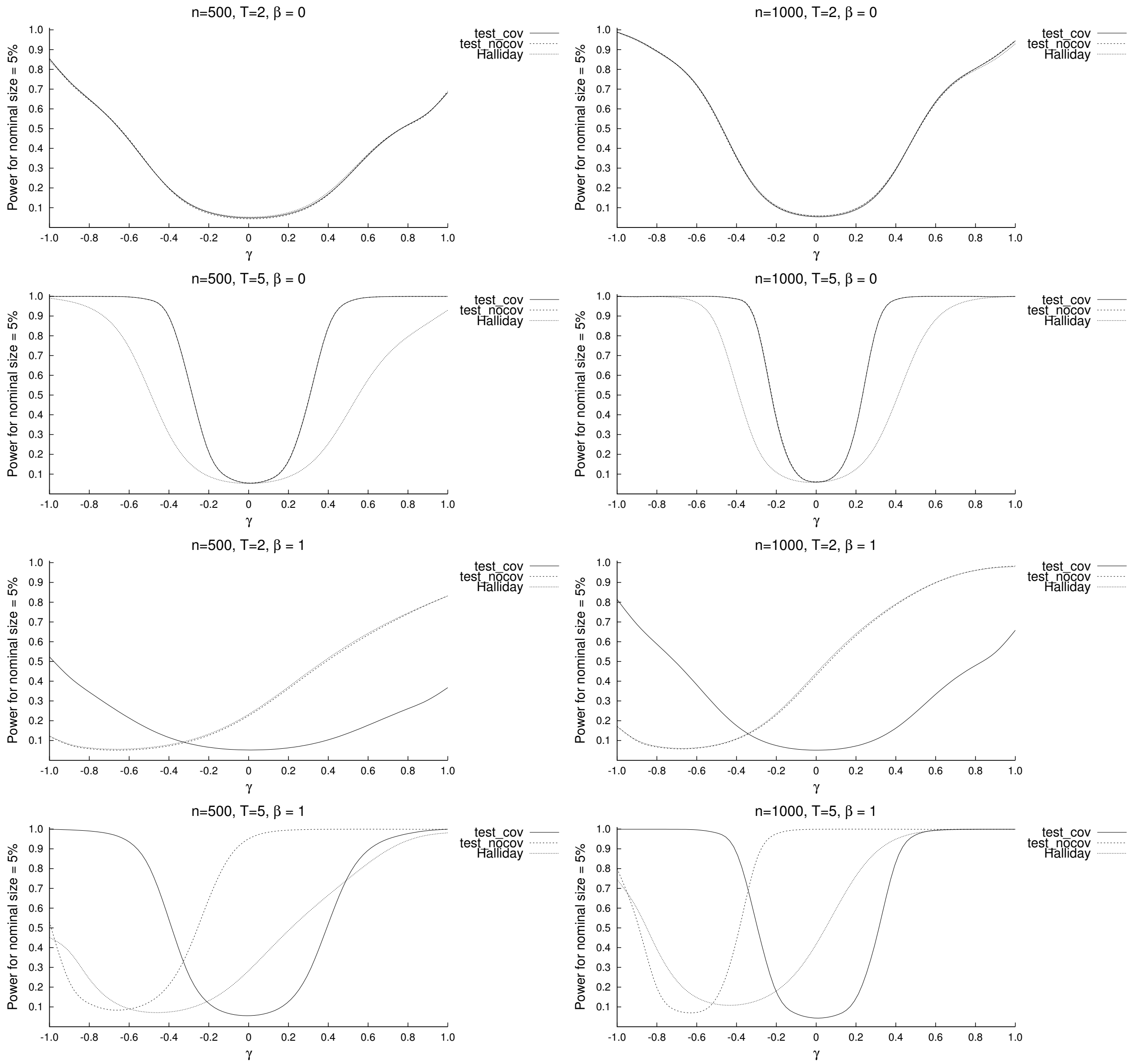
Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68:839–874.

Hsiao, C. (2005). *Analysis of Panel Data*. Cambridge University Press, New York, 2nd edition.

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

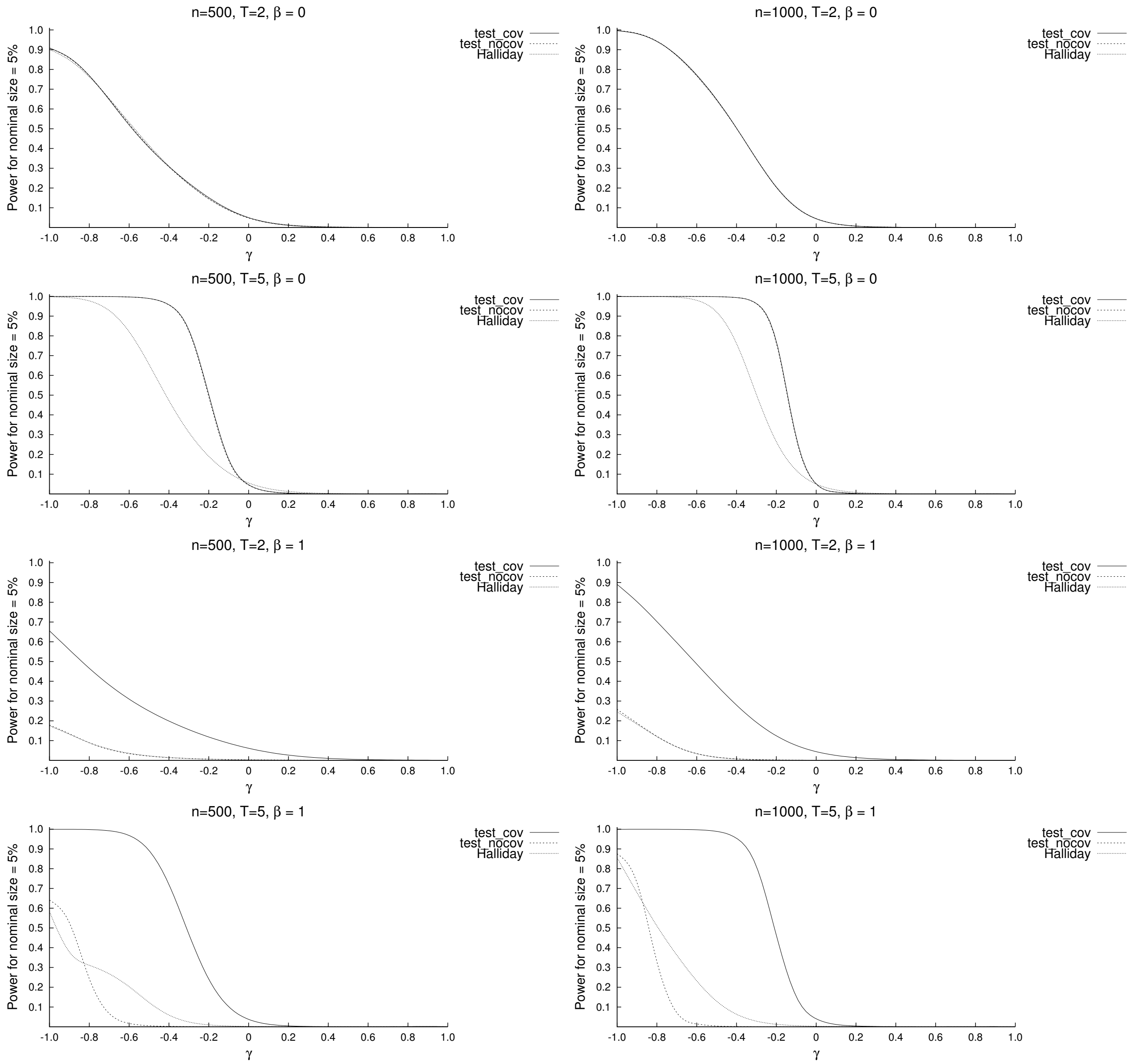
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–26.

Figure 1: Power plots for Wald (QE2) and Halliday's tests: bidirectional ($H_1 : \gamma \neq 0$)



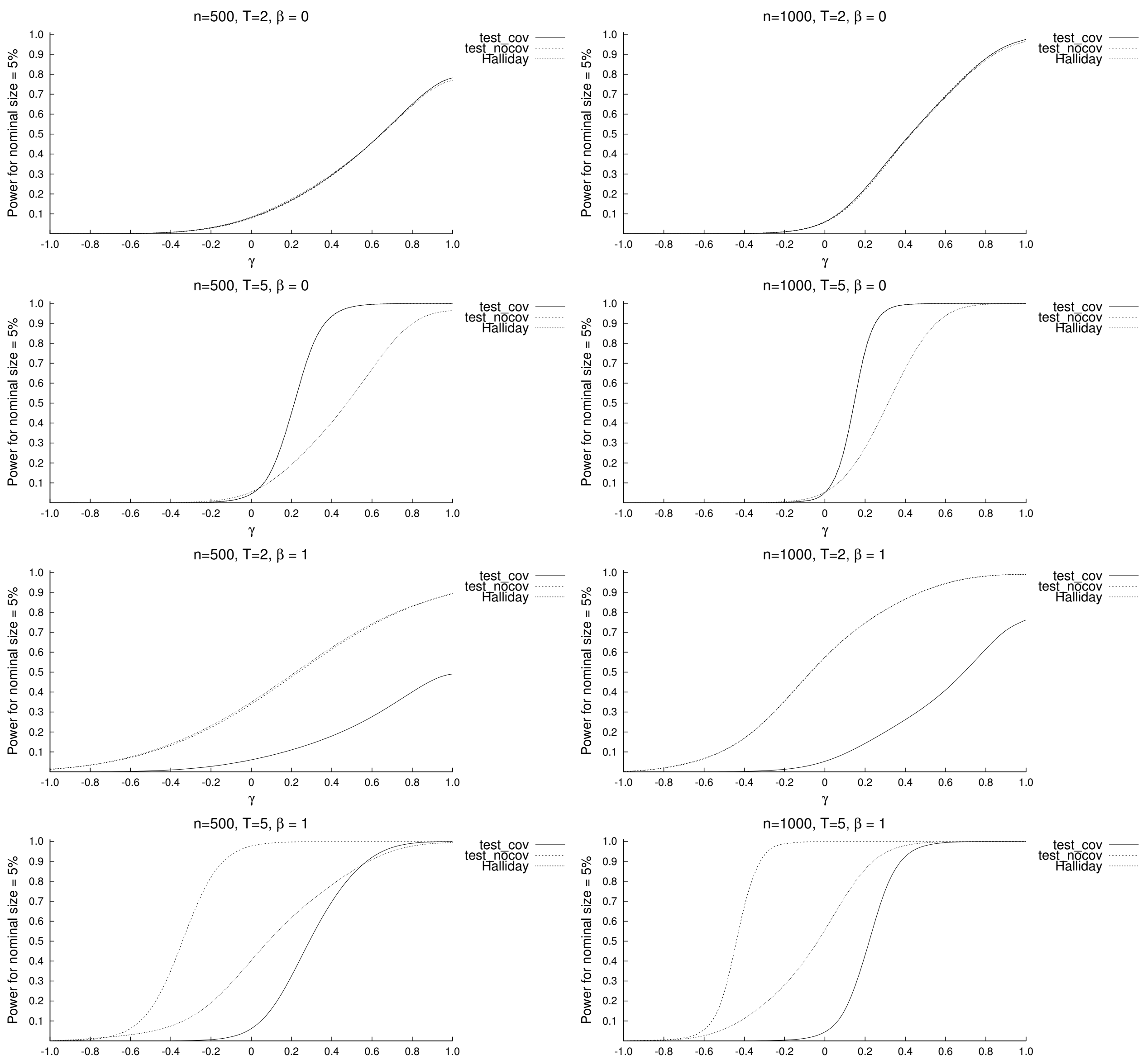
("test_cov" refers to the case in which the covariate x_{it} is included in the QE2 model; "test_nocov" is referred to the case in which the covariate is not included; power curves are smoothed by means of sixth–order polynomials)

Figure 2: Power plots for Wald (QE2) and Halliday's tests: lower tailed ($H_1 : \gamma < 0$)



("test_cov" refers to the case in which the covariate x_{it} is included in the QE2 model; "test_nocov" is referred to the case in which the covariate is not included; power curves are smoothed by means of sixth–order polynomials)

Figure 3: Power plots for Wald (QE2) and Halliday's tests: upper tailed ($H_1 : \gamma > 0$)



("test_cov" refers to the case in which the covariate x_{it} is included in the QE2 model; "test_nocov" is referred to the case in which the covariate is not included; power curves are smoothed by means of sixth–order polynomials)

Table 1: Simulation results for Wald (QE2) and Halliday's test statistics: bidirectional

	γ	$n=500$ test_cov	$n=500$ test_nocov	$n=500$ Halliday	$n=1000$ test_cov	$n=1000$ test_nocov	$n=1000$ Halliday
$\beta=0$	-1.0	0.854	0.855	0.846	0.989	0.988	0.987
$T=2$	-0.5	0.313	0.314	0.312	0.557	0.550	0.556
	0.0	0.048	0.044	0.052	0.054	0.058	0.059
	0.5	0.261	0.262	0.270	0.474	0.471	0.474
	1.0	0.684	0.690	0.682	0.943	0.946	0.930
$\beta=0$	-1.0	1.000	1.000	0.990	1.000	1.000	1.000
$T=5$	-0.5	0.983	0.982	0.524	0.998	0.998	0.849
	0.0	0.054	0.055	0.053	0.060	0.061	0.058
	0.5	0.971	0.972	0.413	0.998	0.998	0.719
	1.0	1.000	1.000	0.929	1.000	1.000	1.000
$\beta=1$	-1.0	0.523	0.122	0.122	0.815	0.172	0.171
$T=2$	-0.5	0.158	0.059	0.064	0.270	0.076	0.078
	0.0	0.052	0.226	0.233	0.051	0.430	0.441
	0.5	0.137	0.573	0.583	0.243	0.848	0.850
	1.0	0.368	0.832	0.832	0.658	0.983	0.980
$\beta=1$	-1.0	1.000	0.530	0.452	1.000	0.800	0.745
$T=5$	-0.5	0.810	0.123	0.073	0.980	0.146	0.113
	0.0	0.056	0.951	0.283	0.043	1.000	0.417
	0.5	0.761	1.000	0.748	0.971	1.000	0.980
	1.0	1.000	1.000	0.981	1.000	1.000	1.000

("test_cov" refers to the case in which the covariate x_{it} is included in the QE2 model; "test_nocov" is referred to the case in which the covariate is not included)

Table 2: Simulation results for Wald test (QE1) test statistic: bidirectional

	γ	$n = 500$ test_cov	$n = 500$ test_nocov	$n = 1000$ test_cov	$n = 1000$ test_nocov
$\beta = 0$	-1.0	0.709	0.710	0.935	0.933
$T = 2$	-0.5	0.195	0.191	0.351	0.354
	0.0	0.045	0.050	0.050	0.056
	0.5	0.130	0.130	0.220	0.222
	1.0	0.268	0.268	0.483	0.480
$\beta = 1$	-1.0	0.347	0.102	0.600	0.149
$T = 2$	-0.5	0.110	0.058	0.169	0.062
	0.0	0.039	0.148	0.045	0.246
	0.5	0.091	0.311	0.136	0.525
	1.0	0.158	0.448	0.308	0.772

("test_cov" refers to the case in which the covariate x_{it} is included in the QE1 model; "test_nocov" is referred to the case in which the covariate is not included)

Table 3: Tests for state dependence ($H_1 : \gamma \neq 0$): proposed Wald test (QE2) and Halliday's test statistics for the overall PSID dataset

	Employment stat.	Employment p-value	Fertility stat.	Fertility p-value
Proposed Wald test				
W	13.58	0.00	-6.80	0.00
Halliday's test				
S_1 (1st triple)	5.75	0.00	-4.74	0.00
S_2 (2nd triple)	4.80	0.00	-4.97	0.00
S_3 (3rd triple)	4.02	0.00	-1.09	0.27
S_4 (4th triple)	4.27	0.00	-5.10	0.00
Sample size	1446		1446	

(QE2 model is estimated with covariates; Bonferroni corrected nominal size: 0.010206)

Table 4: Estimation results based on the PCML approach (Bartolucci and Nigro, 2012): overall PSID dataset

	Employment				Fertility			
	coeff.	s.e.	Wald-stat.	p-value	coeff.	s.e.	Wald-stat.	p-value
Child 1–2	-0.675	0.13	-5.10	0.00	-0.719	0.15	-4.72	0.00
Child 3–5	-0.312	0.12	-2.52	0.01	-1.085	0.21	-5.05	0.00
Child 6–13	-0.032	0.12	-0.25	0.40	-1.055	0.26	-4.08	0.00
Child 14–	-0.010	0.14	-0.07	0.47	-0.800	0.43	-1.86	0.03
Income/1000	-0.007	0.00	-1.68	0.05	-0.000	0.00	-0.13	0.45
1989	0.089	0.14	-1.12	0.13	0.402	0.15	4.64	0.00
1990	0.317	0.13	0.65	0.26	0.445	0.19	2.64	0.00
1991	0.089	0.13	2.49	0.01	0.397	0.24	2.31	0.01
1992	0.001	0.13	0.67	0.25	0.448	0.29	1.66	0.05
Lag fertility	-0.185	0.17	0.01	0.50	-0.906	0.21	1.56	0.06
Lag employment	1.550	0.11	13.93	0.00	0.801	0.17	-4.35	0.00

Table 5: Tests for state dependence ($H_1 : \gamma \neq 0$): proposed Wald test (QE2) and Halliday's test statistics for the PSID dataset

	Years of schooling ≤ 12				Years of schooling > 12			
	Employment		Fertility		Employment		Fertility	
	stat.	p-value	stat.	p-value	stat.	p-value	stat.	p-value
Proposed Wald test								
W	10.18	0.00	-2.70	0.01	9.04	0.00	-6.25	0.00
Halliday's test								
S_1 (1st triple)	3.04	0.00	-1.60	0.11	5.27	0.00	-5.79	0.00
S_2 (2nd triple)	3.49	0.00	-0.81	0.41	3.30	0.00	-6.28	0.00
S_3 (3rd triple)	2.83	0.00	-0.79	0.43	2.75	0.01	-0.78	0.00
S_4 (4th triple)	4.40	0.00	-1.18	0.24	1.68	0.09	-5.45	0.00
Sample size	773				673			

(QE2 model is estimated with covariates; Bonferroni corrected nominal size: 0.010206)

Table 6: Estimation results based on the PCML approach (Bartolucci and Nigro, 2012): PSID dataset for the subsample with years of schooling ≤ 12

	Employment				Fertility			
	coeff.	s.e.	Wald-stat.	p-value	coeff.	s.e.	Wald-stat.	p-value
Child 1–2	-0.419	0.20	-2.12	0.02	-0.726	0.24	-2.96	0.00
Child 3–5	0.055	0.17	0.32	0.33	-0.750	0.33	-2.26	0.02
Child 6–13	0.127	0.16	0.77	0.22	-0.524	0.38	-1.36	0.17
Child 14–	0.067	0.17	0.38	0.35	-0.757	0.69	-1.10	0.27
Income/1000	-0.018	0.01	-2.47	0.01	-0.004	0.01	-0.27	0.78
1989	0.153	0.19	0.13	0.45	0.328	0.23	3.99	0.00
1990	0.193	0.17	0.82	0.20	0.189	0.29	1.41	0.16
1991	0.341	0.18	1.13	0.13	0.004	0.36	0.66	0.51
1992	-0.102	0.18	1.92	0.03	-0.516	0.46	0.01	0.99
Lag fertility	0.031	0.25	-0.57	0.28	-0.145	0.32	-1.13	0.26
Lag employment	1.529	0.15	10.10	0.00	1.090	0.27	-0.45	0.65

Table 7: Estimation results based on the PCML approach (Bartolucci and Nigro, 2012): PSID dataset for the subsample with years of schooling > 12

	Employment				Fertility			
	coeff.	s.e.	Wald-stat.	p-value	coeff.	s.e.	Wald-stat.	p-value
Child 1–2	-0.932	0.19	-0.86	0.19	-0.763	0.20	-3.89	0.00
Child 3–5	-0.690	0.19	-0.64	0.19	-1.350	0.29	-4.68	0.00
Child 6–13	-0.220	0.21	-4.95	0.00	-1.438	0.36	-3.99	0.00
Child 14–	-0.154	0.25	-3.59	0.00	-0.811	0.61	-1.34	0.09
Income/1000	-0.001	0.00	-1.05	0.15	-0.000	0.00	-0.02	0.49
1989	0.038	0.29	-0.61	0.27	0.529	0.21	2.99	0.00
1990	0.505	0.20	-0.24	0.41	0.670	0.26	2.57	0.01
1991	-0.139	0.21	-1.43	0.08	0.694	0.33	2.53	0.01
1992	0.208	0.21	0.18	0.43	1.071	0.39	2.10	0.02
Lag fertility	-0.325	0.23	2.57	0.01	-1.275	0.28	2.71	0.00
Lag employment	1.591	0.17	-0.68	0.25	0.684	0.23	-4.57	0.00