



# Testing the binomial fixed effects logit model, with an application to female labour supply

Rainer Winkelmann<sup>1</sup> · Lin Xu<sup>1</sup>

Received: 22 October 2019 / Accepted: 10 February 2021  
© The Author(s) 2021

## Abstract

Regression models for proportions are frequently encountered in applied work. The conditional expectation function is bounded between 0 and 1 and therefore must be nonlinear, requiring nonstandard panel data extensions. One possible approach is the binomial panel logit model with fixed effects (Machado in *J Econom* 119:73–98, 2004). We propose a new and simple implementation of this conditional maximum likelihood estimator for standard software. We investigate the properties of the estimator under misspecification and derive a new test for overdispersion. Estimator and test are applied in a study of contracted working volumes, measured as proportion of full-time work, for women in Switzerland.

**Keywords** Proportions data · Unobserved heterogeneity · Conditional maximum likelihood · Overdispersion

**JEL Classification** C23 · J21

## 1 Introduction

After half a century of research on econometric models for limited dependent variables (Maddala 1983; Wooldridge 2002), it remains the case that only a small portion of it deals with proportions data, and even a smaller one with panel models for such proportions. Machado (2004) proposes the binomial fixed effects logit model, Papke

---

We are grateful to Damian Kozbur, Johannes Kunz, Kevin Staub and Tiemen Woutersen for helpful discussions, and to two anonymous referees for very useful comments. The data used in the application have been provided by the Swiss Household Panel. We gratefully acknowledge financial support by the Swiss National Science Foundation.

---

✉ Rainer Winkelmann  
rainer.winkelmann@econ.uzh.ch

Lin Xu  
lin.xu@econ.uzh.ch

<sup>1</sup> Department of Economics, University of Zurich, Zurich, Switzerland

and Wooldridge (2008) a correlated random effects probit quasi-likelihood estimator, and Ramalho et al. (2016) a class of exponential GMM estimators.

And yet, proportions and related types of data are regularly encountered in applied econometric work. They sometimes correspond to the fraction of “successes” in a sequence of Bernoulli trials. Examples are the proportion of successful patent applications (Machado 2004) and the proportion of days absent from work (Barmby et al. 2001). Similarly, variety scores (e.g. the number of applicable items in a general health questionnaire), bounded count data, as well as ratings, can be re-scaled to the (0, 1)-interval. For example, the degree of (customer or life) satisfaction in surveys often has a lower bound of zero (meaning “completely dissatisfied”) and some upper bound (e.g. 10, meaning “completely satisfied”), which can then be re-coded as 100%. All these variables share the key features of being discrete and bounded, and the binomial model with a logit function for the expected proportion provides a natural starting point for modelling.

For the fixed effects setting, Machado (2004) shows that the incidental parameters problem can be overcome by a conditional maximum likelihood (CML) estimator, much like it is the case for the binary response logit model (Chamberlain 1980). She also provides Monte Carlo evidence indicating that the dummy variables (DV) approach is subject to an upward bias that is decreasing both in the length of the panel,  $T$ , and in the number of Bernoulli trials,  $K$ . For  $T > 5$  and  $K > 5$ , CML and DV approaches yield quite comparable results with minor bias (Machado 2004).

This paper advances the earlier literature in three directions: First, we show how the binomial logit fixed effects estimator can be implemented in any off-the-shelf statistical software that includes a conditional logit routine, using the idea of cloning, or data expansion. Second, we study the properties of the CML and DV estimators for the case where the binomial distributional assumption fails. The leading example is that of overdispersion, originating from random unobserved heterogeneity or dependence among the Bernoulli trials. The CML estimator is not a pseudo-ML estimator in the sense of Gourieroux et al. (1984), and it does not possess formal robustness properties. We therefore investigate the extent of bias in a series of simulation experiments. Third and finally, we derive and implement a new test for the binomial assumption, i.e. a test for the hypothesis of no dispersion, as existing tests (e.g. Dean 1992) cannot be applied because the fixed effects are not estimated by the CML estimator.

To illustrate the proposed methods, we conduct a study of the determinants of women’s work behaviour in Switzerland. The outcome variable is the contracted work-time percentage, where 0 means no work and 1 means full-time work. Data are extracted from the Swiss Household Panel for the years 2012–2016. The binomial logit estimates indicate that having children is associated with substantially reduced work-time percentage, *ceteris paribus*. Perhaps more surprisingly, having a partner makes the effect more pronounced, whereas speaking French reduces it.

## 2 Model and estimation

A proper panel model for proportions  $y_{it} \in [0, 1]$  must overcome two challenges. First, the model should observe the restricted support of the outcome, as well as being able

to handle data clustering at the end points. For instance, the log-odds transformation  $\log[y_{it}/(1 - y_{it})]$  is not defined for  $y_{it} = 0$  or  $y_{it} = 1$ . Another method facing the same limitation is beta regression, which is flexible for fitting continuous proportional data but cannot give predictions at the boundaries with positive probability. Second, direct control for unobserved time-invariant individual heterogeneity (that may or may not be correlated with the regressors), using a dummy for each cross-sectional unit is subject to the incidental parameters problem, leading to inconsistent estimation of structural parameters when the length of panel  $T$  is fixed.

Machado (2004) addresses these two issues by considering a binomial logit model with fixed effects. The application she had in mind was using information on the number of patent applications and patents granted at the firm level to estimate the probability of obtaining a patent (i.e. proportion of patents granted). She derived a consistent conditional maximum likelihood estimator based on the following assumptions:

**Assumption 1** Let  $Y_{it} = Ky_{it}$ , where  $K$  is a known integer and

$$y_{it} \in \left\{0, \frac{1}{K}, \frac{2}{K}, \dots, 1\right\}$$

such that

$$Y_{it}|p_{it} \sim \text{binomial}(K, p_{it}), \quad i = 1, \dots, N; \quad t = 1, \dots, T \quad (1)$$

Here,  $K$  is the number of “trials”,  $Y_{it} = Ky_{it}$  is the “number of successes”, and  $y_{it}$  is the proportion, or fraction of successes for observation unit  $i$  in period  $t$ .

**Assumption 2** Let the expected proportion depend on covariates  $x_{it}$ , and an individual-specific effect  $\alpha_i$  as follows:

$$E(y_{it}|x_{it}, \alpha_i) = p_{it} = \frac{\exp(x'_{it}\beta + \alpha_i)}{1 + \exp(x'_{it}\beta + \alpha_i)} \equiv \Lambda_{it} \quad (2)$$

$x_{it}$  and  $\alpha_i$  can be correlated.

**Assumption 3** Observations are independent between individuals and, conditional on group effects  $\alpha_i$ , serially uncorrelated.

The objective of the analysis is estimation of  $\beta$ . Under Assumptions 1–3, the joint binomial density for  $Y_{i1}, Y_{i2}, \dots, Y_{iT}$  conditional on  $\sum_t Y_{it}$  is given by (see Machado 2004)

$$\begin{aligned} f\left(Y_{i1}, Y_{i2}, \dots, Y_{iT} \mid \sum_t Y_{it}\right) &= \frac{\prod_t \binom{K}{Y_{it}} p_{it}^{Y_{it}} (1 - p_{it})^{K - Y_{it}}}{\sum_{q \in Q_i} \prod_t \binom{K}{q_t} p_{it}^{q_t} (1 - p_{it})^{K - q_t}} \\ &= \frac{\exp(\sum_t Y_{it} x'_{it} \beta) \prod_t \binom{K}{Y_{it}}}{\sum_{q \in Q_i} \exp(\sum_t q_t x'_{it} \beta) \prod_t \binom{K}{q_t}} \quad (3) \end{aligned}$$

where  $Q_i = \{(q_1, q_2, \dots, q_T) | q_t \in \{0, 1, 2, \dots, K\}, \sum_t q_t = \sum_t Y_{it}\}$ . The conditional binomial approach eliminates the fixed effects  $\alpha_i$  which appear in the numerator and denominator with same power. Observations for which  $\sum_t Y_{it} = 0$  or  $\sum_t Y_{it} = KT$  have a conditional probability of 1 and do not contribute to estimation of  $\beta$ . For proportion data, such outcomes tend to be much less prevalent than they are for binary outcomes.

In principle, the Machado (2004) approach solves an important problem in the analysis of panel data for proportions. In contrast to Papke and Wooldridge (2008), it is “semi-parametric”, as there is no need to specify the relationship between the individual effect and the regressors, and also no need to add an assumption on the distribution of the individual effects. And yet, subsequent applications have been few, perhaps, because the estimator has a couple of limitations. First, the estimator is not readily available in standard econometric software packages. We therefore develop a simple modification that makes it easily implementable in standard software. And second, the binomial assumption may be violated, and the properties of the estimator under misspecification are unknown so far. We provide such a misspecification analysis in Sect. 2.3, and also derive a test for the binomial assumption in Sect. 3. In addition, it is important to point out that the binomial fixed effects estimator can be applied in a broader range of situations than hitherto considered, i.e. beyond those relating to the number (or proportion) of successes in a sequence of  $K$  independent Bernoulli trials. Even in the absence of such a process, the model can be a good starting point for fractions and shares, as we illustrate in an application to work-time percentages.

## 2.1 An alternative implementation

To understand, how the binomial logit fixed effects estimator can be implemented using any off-the-shelf statistical software with a conditional logit routine, note that the binomial distribution arises as the sum of  $K$  independent Bernoulli trials. Therefore, two estimators are equivalent: one based on a binomial log-likelihood function and the other based on a Bernoulli log-likelihood for an expanded dataset.

For the expanded dataset, one simply generates a sequence of  $K$  copies for each  $i$ , keeping the regressors unchanged, where the proportion  $y_{it}$  is replaced by a sequence of 0/1 indicator variables  $d_{ijt}$  in arbitrary order such that

$$\sum_{j=1}^K d_{ijt} = Ky_{it} \quad (4)$$

It follows that  $d_{ijt}$  and  $y_{it}$  have the same CEF:

$$E(y_{it}|x_{it}) = E\left(\frac{\sum_{j=1}^K d_{ijt}}{K} \middle| x_{it}\right) = E(d_{ijt}|x_{it}). \quad (5)$$

The logit (Bernoulli) log-likelihood function of the expanded dataset is given by

$$\begin{aligned} \log L &= \sum_t^T \sum_i^N \left[ \sum_j^K d_{ijt} \log(\Lambda_{it}) + (1 - d_{ijt}) \log(1 - \Lambda_{it}) \right] \\ &= \sum_t^T \sum_i^N Y_{it} \log(\Lambda_{it}) + (K - Y_{it}) \log(1 - \Lambda_{it}). \end{aligned} \tag{6}$$

This log-likelihood function is equal to the binomial log-likelihood as well as to the Bernoulli quasi-log-likelihood (Papke and Wooldridge 1996, replacing  $Y_{it}$  by  $y_{it}$  and  $(K - Y_{it})$  by  $(1 - y_{it})$ ), up to an additive constant, and the three ML estimators are therefore identical.

Similarly, the conditional density function for individual  $i$  at time  $t$  can be written as:

$$f\left(\{d_{ijt}\} \mid \sum_t \sum_j d_{ijt}\right) = \frac{\prod_t \prod_j p_{it}^{d_{ijt}} (1 - p_{it})^{1-d_{ijt}}}{\sum_{s \in S_i} \prod_t \prod_j p_{it}^{s_{jt}} (1 - p_{it})^{1-s_{jt}}} = \frac{\exp\left(\sum_t \sum_j d_{ijt} x'_{it} \beta\right)}{\sum_{s \in S_i} \exp\left(\sum_t \sum_j s_{jt} x'_{it} \beta\right)} \tag{7}$$

where  $S_i = \{(s_{11}, s_{21}, \dots, s_{K1}, s_{12}, \dots, s_{KT}) \mid s_{jt} \in \{0, 1\}, \sum_t \sum_j s_{jt} = \sum_t \sum_j d_{ijt}\}$ .

Compared with Eq. (3), the number of  $s$  such that  $\{s \mid \sum_j s_{ijt} = q_{it}\}$  is  $\binom{K}{q_{it}}$  for given  $q_{it}$ . Equation (7) is therefore basically the same as Eq. (3), except for the term  $\prod_t \binom{K}{Y_{it}}$  in the numerator of (3). But this term does not depend on any parameter and thus does not affect the first-order condition for the maximum of the log-likelihood function. Specifically, the conditional Bernoulli log-likelihood function is given by:

$$\log L = \sum_i \left[ \sum_t \sum_j d_{ijt} x'_{it} \beta - \log \left( \sum_{s \in S_i} \exp \left( \sum_t \sum_j s_{jt} x'_{it} \beta \right) \right) \right] \tag{8}$$

with the first derivative

$$\frac{\partial \log L}{\partial \beta} = \sum_i \left[ \sum_t K y_{it} x'_{it} - \frac{\sum_{s \in S_i} \exp \left( \sum_t \sum_j s_{jt} x'_{it} \beta \right) \sum_t \sum_j s_{jt} x'_{it}}{\sum_{s \in S_i} \exp \left( \sum_t \sum_j s_{jt} x'_{it} \beta \right)} \right] \tag{9}$$

which is the same as that of the conditional binomial model and therefore will yield the same consistent estimator of  $\beta$ , after elimination of the fixed effects. We from now on refer to this estimator as the Blogit (for binomial logit) conditional maximum likelihood estimator, or in short, Blogit CML, in contrast to the inconsistent binomial estimator with dummy variables included for each cross-sectional unit, Blogit DV.

Of course, there can be situations where the expansion approach becomes practically infeasible: as  $K$  gets large, for instance, because proportions are measured at the

granularity of percentage points, the size of the set  $S_i$  of the conditional Bernoulli log-likelihood expressions is increased from  $T$  to  $100 \times T$ , at which point one may run into computational constraints.

## 2.2 Overdispersion

Departures from the binomial proportions model can take a number of forms. The first one is a violation of the independence assumption for the underlying Bernoulli trials. Positive dependence, or contagion, among the sequence of Bernoulli trials causes overdispersion, a conditional variance exceeding the binomial variance  $K p_{it}(1 - p_{it})$ . Another violation stems from “random unobserved heterogeneity”. This is in addition to the time-invariant unobserved heterogeneity  $\alpha_i$ . Random unobserved heterogeneity is time- and individual-specific, as well as unrelated to  $x_{it}$ . Specifically, it means that  $p_{it}$  is no longer a constant but rather a random variable, say  $\tilde{p}_{it}$ . Marginalizing over  $\tilde{p}_{it}$  then leads to a mixture model that is characterized by overdispersion as well. Depending on the distribution of  $\tilde{p}_{it}$ , proportions can, for example, have a u-shaped probability function even conditional on  $\alpha_i$  and  $x_{it}$ , i.e. probability mass stacked at the endpoints of 0 and 1, which is never the case for a binomial distribution that has either a single, or two adjacent modes.

A prominent example for a continuous mixture is the beta-binomial model, where the conditional probability is

$$\tilde{p}_{it} \sim \text{beta}(u_{it}, v_{it}), \quad (10)$$

and

$$u_{it} = \phi \Lambda(x'_{it}\beta + \alpha_i), \quad v_{it} = \phi(1 - \Lambda(x'_{it}\beta + \alpha_i)).$$

where  $\phi > 0$  is a parameter that determines the degree of overdispersion. It is straightforward to show that a beta-binomial distribution with this parameterization has expectation  $K \Lambda_{it}$  and variance

$$\text{Var}(Y_{it}|K, \Lambda_{it}, \phi) = K \Lambda_{it}(1 - \Lambda_{it}) \left(1 + \frac{K - 1}{\phi + 1}\right) \quad (11)$$

Thus, the variance of the beta-binomial model is proportional to that of the binomial model. Overdispersion increases in  $K$ , the number of trials, and it decreases in the parameter  $\phi$ . The binomial variance is obtained for  $K = 1$ , or in the limit, for  $\phi \rightarrow \infty$ , which also means that  $\text{Var}(\tilde{p}_{it}) \rightarrow 0$ .

In general, fixed effects conditional maximum likelihood estimators are not consistent if the underlying model is misspecified. The reason is that the first-order condition is not a moment condition for the mean, but rather a function of the conditional probabilities. However, it might still be the case that the CML estimator works satisfactorily as long as the degree of overdispersion, in other words, the departure from the binomial assumption, is not too large. We will explore this type of robustness in a series of simulation experiments. We thereby extend results by Machado (2004), who considered

the severity of the incidental parameters problem and the small sample properties of the CML estimator under the maintained assumption of a correctly specified binomial model. In our simulations, this assumption is dropped.

### 2.3 Simulation study

The simulation experiments employ two different data generating processes: one where the binomial assumption is satisfied, and the other, based on the beta-binomial model, where overdispersion is present. Unobserved time-invariant individual heterogeneity is positively correlated with the regressor in both cases. The degree of overdispersion is varied from 10 to 200%.

Both set-ups use the same logit conditional expectation function with a single regressor

$$E(y_{it}|x_{it}, \alpha_i) = \Lambda(\beta_0 + \beta_1 x_{it} + \alpha_i) = \frac{\exp(\beta_0 + \beta_1 x_{it} + \alpha_i)}{1 + \exp(\beta_0 + \beta_1 x_{it} + \alpha_i)}, \quad (12)$$

where  $\beta_0 = 0$ ,  $\beta_1 = 2$  and the size of the cross section is either  $N = 100$  or  $N = 500$ . The time dimension increases from  $T = 2$ ,  $T = 5$  to  $T = 10$ .

The regressor  $x_{it}$  is drawn from a uniform distribution with support  $[-1, 1]$  and has therefore a mean of 0 and a variance of 1/3. Draws are independent both across individuals and over time. We make a correlated random effects assumption:

$$\alpha_i = \sqrt{T} \bar{x}_i + \varepsilon_i, \quad (13)$$

where  $\varepsilon_i \sim N(0, 1)$ . It follows that the correlation between  $\alpha_i$  and  $\bar{x}_i$  is 0.5, a substantial amount.

Once the mean is given, the dependent variable is obtained by generating pseudo-random numbers from either a binomial or a beta-binomial distribution. Specifically, we first draw integer random numbers from a (beta) binomial distribution with parameters  $K$  and  $\Lambda(x_{it}\beta_1 + \alpha_i)$  and then divide the result by the number of categories  $K$ , e.g.

$$y_{it} = \frac{K y_{it}}{K}, K y_{it} \sim \text{binomial}(K, p_{it}), p_{it} = \Lambda(\beta_0 + \beta_1 x_{it} + \alpha_i) \quad (14)$$

$K$  is exogenously set to 2, 5 or 10. For  $K = 2$ ,  $K \times y_{it}$  can be 0, 1, or 2, with corresponding fractions of  $y_{it} = 0, 0.5$ , or 1, respectively; if  $K = 10$ ,  $y_{it}$  takes on one-digit decimals: 0, 0.1, 0.2, ..., 1.

Ignoring the presence of the individual-specific component and estimating the marginal, pooled model instead has two effects:

- $\beta_1$  is upward biased due to the positive correlation between  $x_{it}$  and  $\alpha_i$ .
- $\beta_1$  is downward biased due to omitted heterogeneity. In the probit model, there is a closed form expression for this bias (Wooldridge 2002). In the logit model, it needs to be computed numerically, but the direction is the same.

Which one of the two biases is larger is an empirical matter. The DV estimator, on the other hand, suffers from the standard incidental parameters bias that is upward (Abrevaya 1997).

Table 1 shows the simulation results based on 1000 replications, for a sample size of  $N = 100$ . The mean and standard deviation of estimated coefficients across replications are reported. Three estimators were used: Blogit CML, Blogit DV, and pooled logit, respectively. Similar to Machado (2004), we find that the Blogit CML model estimates the true structural slope parameter very well even for small samples. There is a 2% upward bias for  $T = K = 2$  that vanishes quickly as either  $T$  or  $K$  increases. The sampling variability decreases not only in  $T$  but also in  $K$ , albeit at a less than  $\sqrt{K}$  rate. The Blogit DV estimator has a larger bias and a larger standard error, and hence a higher mean squared error, in all settings. The bias becomes small as  $T$  and  $K$  increase. For instance, for  $T = 10$  and  $K = 10$ , the mean Blogit DV estimate is 2.025, whereas the mean Blogit CML estimate is 2.000. On the other hand, the pooled logit estimator has no tendency to converge to the true parameter  $\beta_1 = 2$ , over- or underestimating it depending on  $K$  and  $T$ . In the lower panel of Table 1, simulations are repeated for a larger sample, with  $N = 500$  instead of  $N = 100$ . The qualitative conclusions remain unchanged.

## 2.4 Beta-binomial DGP

Simulations from the beta-binomial model add a further step: instead of directly obtaining binomial responses with (conditional on  $x_{it}$  and  $\alpha_i$ ) success probability  $p_{it} = \Lambda(\beta_0 + \beta_1 x_{it} + \alpha_i)$ ,  $\tilde{p}_{it}$  is now drawn from a beta distribution with mean  $p_{it}$ :

$$\tilde{p}_{it} \sim \text{beta}(\phi \Lambda(\beta_0 + \beta_1 x_{it} + \alpha_i), \phi(1 - \Lambda(\beta_0 + \beta_1 x_{it} + \alpha_i))) \quad (15)$$

From (11), we know that the multiplicative variance inflation factor depends both on  $K$  and  $\phi$ . To keep the degree of overdispersion constant for  $K = 2, 5, 10$ , we adjust  $\phi$  accordingly. For example, for 10% overdispersion and  $K = 2$ , we have  $1 + (K - 1)/(\phi + 1) = 1.1$ , so  $\phi = 9$ .

As a practical limitation, common beta random number generators set lower bounds (above the theoretical ones of 0) for the two parameters. In Stata, for example, these are given by 0.05 and 0.15, respectively. From (15) we see that attempts to draw from the beta using arguments violating these bounds are more likely to arise when the mean is close to zero or one, or when  $\phi$  is small (and therefore the degree of overdispersion is large). Since such occurrences only depend on exogenous factors, dropping these cases does not invalidate the estimation procedure. However, it affects the effective sample size and thus leads to higher standard errors than would otherwise be the case.

Figures 1 and 2 plot the relative biases of Blogit CML and Blogit DV against the degree of overdispersion, for  $N = 100$  and  $N = 500$ , respectively. Overdispersion varies from 10% to 200%. (The full results on the means and standard deviations of the estimators for each DGP are given in Tables 6 and 7 in ‘‘Appendix.’’)

Three key patterns emerge. First, overdispersion leads to an upward bias of both the Blogit CML and the Blogit DV estimators. The bias increases in the amount of



**Table 1** Simulation results under the binomial distribution

	T = 2			T = 5			T = 10		
	Blomit CML	Blomit DV	Pooled logit	Blomit CML	Blomit DV	Pooled logit	Blomit CML	Blomit DV	Pooled logit
N = 100									
K = 2	2.049 (0.419)	2.880 (0.621)	2.242 (0.255)	2.003 (0.178)	2.280 (0.211)	1.986 (0.145)	2.006 (0.118)	2.134 (0.128)	1.877 (0.101)
K = 5	2.012 (0.233)	2.279 (0.272)	2.243 (0.190)	2.000 (0.111)	2.101 (0.118)	1.990 (0.104)	2.002 (0.073)	2.051 (0.075)	1.872 (0.076)
K = 10	2.005 (0.157)	2.129 (0.169)	2.241 (0.153)	2.001 (0.078)	2.050 (0.081)	1.989 (0.091)	2.000 (0.052)	2.025 (0.052)	1.871 (0.063)
N = 500									
K = 2	2.013 (0.170)	2.826 (0.254)	2.234 (0.112)	2.002 (0.082)	2.278 (0.097)	1.987 (0.068)	2.002 (0.053)	2.130 (0.057)	1.870 (0.046)
K = 5	2.003 (0.102)	2.268 (0.119)	2.232 (0.078)	2.001 (0.051)	2.102 (0.054)	1.986 (0.050)	2.000 (0.033)	2.049 (0.034)	1.867 (0.033)
K = 10	1.997 (0.071)	2.121 (0.076)	2.235 (0.067)	1.999 (0.034)	2.049 (0.035)	1.984 (0.042)	2.000 (0.023)	2.024 (0.023)	1.867 (0.029)

Results for 1000 Monte Carlo replications; standard deviations in parentheses. The number of observation in each period is 100 in the upper part of the table and 500 in the second half.  $x_{it} \sim U[-1, 1]$ ,  $\beta_1 = 2$  and  $\alpha_i = \sqrt{T} \bar{x}_i + N(0, 1)$ . Outcome  $y_{it}$  follows the binomial distribution with  $p_{it} = \frac{\exp(\alpha_i \beta_1 + \alpha_i)}{1 + \exp(\alpha_i \beta_1 + \alpha_i)}$ . *Blomit CML* denotes the binomial logit conditional maximum likelihood estimator; *Blomit DV* is the binomial logit fixed effects estimator including dummy variables for each individual; and *Pooled Logit* is the pooled logit estimator that ignores the presence of individual effects

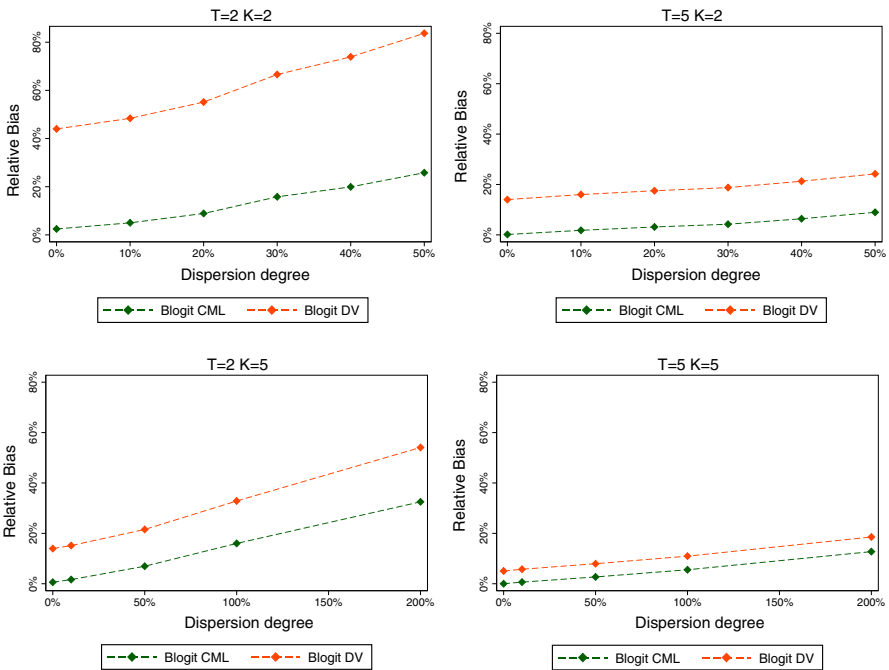


Fig. 1 Relative bias by dispersion degree ( $N = 100$ )

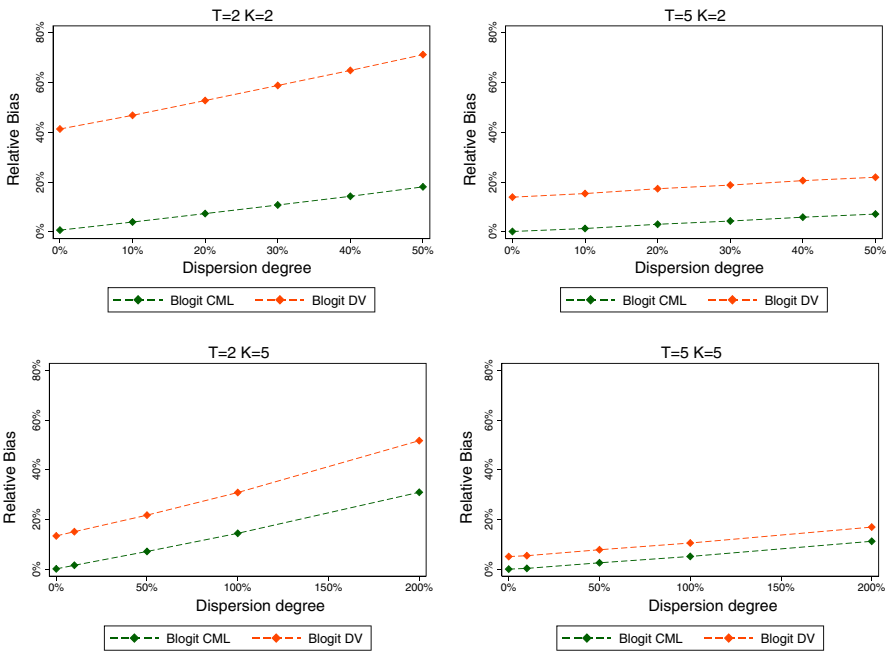


Fig. 2 Relative bias by dispersion degree ( $N = 500$ )

overdispersion. Second, the Blogit CML estimator always dominates the Blogit DV estimator, both in terms of bias and standard error. The same pattern was already found for the binomial case, and it persists in the presence of overdispersion. Third, for a given degree of overdispersion, the bias is decreasing in  $T$  as well as in  $K$ . However, increasing  $K$  alone not necessarily leads to a reduction in estimation bias, because it increases the amount of overdispersion, ceteris paribus. Again, results are qualitatively similar for  $N = 500$  (see Fig. 2).

The overall conclusion is that the Blogit CML estimator maintains a rather good performance even if the binomial model is misspecified, as long as the degree of overdispersion is modest, or else, as long as  $T$  is large. To take the two extreme cases, for  $N = 100$  and  $K = T = 2$ , the mean estimate with 10% overdispersion is 2.1, a 5% upward bias. For  $K = T = 10$ , the mean estimate with 100% overdispersion is 2.049, a 2.45% upward bias.

### 3 A test for overdispersion

Existing binomial tests for  $y_{it}$ , e.g. Dean's (1992) score test or regression-based tests regressing squared residuals  $y_{it} - \hat{\Lambda}_{it}$  on  $\hat{\Lambda}_{it}(1 - \hat{\Lambda}_{it})$ , require estimates  $\hat{\Lambda}_{it}$  in order to obtain conditional variances  $\widehat{\text{Var}}(y_{it}|x_{it})$ . However, the Blogit CML approach does not give us  $\hat{\alpha}_i$ , so this is not feasible. To ascertain the validity of the Blogit CML model assumption, i.e. that  $K y_{it}$  is binomial distributed conditional on  $\alpha_i$  and  $x_{it}$ , we propose an alternative approach that uses  $\hat{\beta}$  but does not require estimates of  $\alpha_i$ , based on taking differences.

To start, consider a binary random variable  $M_{it}$  defined by a draw from a Bernoulli distribution with mean  $y_{it}$ ,  $M_{it} \sim \text{Bernoulli}(y_{it})$ . Clearly, the conditional mean is  $E(M_{it}|y_{it}) = y_{it}$ , while the unconditional mean is  $E(M_{it}) = \Lambda_{it}$ . The conditional variance is  $\text{Var}(M_{it}|y_{it}) = y_{it}(1 - y_{it})$ , while the unconditional variance is  $\text{Var}(M_{it}) = E[y_{it}(1 - y_{it})] + \text{Var}(y_{it}) = \Lambda_{it}(1 - \Lambda_{it})$ .

The basic idea of the test is to compare the variances of the differences  $Y_{it} - Y_{is}$  and that of the difference  $M_{it} - M_{is}$ , for pairs of observations where the underlying probabilities  $p_{it} = \Lambda_{it}$  are the same (or similar) for the two periods. For notational simplicity, let  $t = 1$  and  $s = 2$ . In such cases, outcomes  $Y_{i1}, Y_{i2}$  can be regarded under  $H_0$  as random draws from i.i.d. binomial distributions and the variance of  $Y_{i1} - Y_{i2}$  should be equal to the sum of binomial variances, under assumptions A1 and A3. On the other hand, the Bernoulli draws from the same distributions have standard variances. If there is over- or under-dispersion, the variance of  $Y_{i1} - Y_{i2}$  will be larger or smaller than the variance calculated from Bernoulli draws.

Specifically, consider the variable

$$z_i = \frac{(Y_{i1} - Y_{i2})^2 - K(M_{i1} - M_{i2})^2}{K(K - 1)}. \tag{16}$$

Conditional on  $y_{i1}, y_{i2}$ ,

$$\begin{aligned} E[(M_{i1} - M_{i2})^2|y_{i1}, y_{i2}] &= y_{i1}(1 - y_{i1}) + y_{i2}(1 - y_{i2}) + (y_{i1} - y_{i2})^2 \\ &= y_{i1} + y_{i2} - 2y_{i1}y_{i2}. \end{aligned}$$

Therefore, under A1, A2 and A3, the expectation of  $z_i$  is given by

$$E(z_i) = \frac{1}{K(K-1)} \left[ \text{Var}(Y_{i1}) + \text{Var}(Y_{i2}) + (EY_{i1} - EY_{i2})^2 - K(\Lambda_{i1} + \Lambda_{i2} - 2\Lambda_{i1}\Lambda_{i2}) \right].$$

Under the binomial assumption,  $\text{Var}(Y_{it}) = K\Lambda_{it}(1 - \Lambda_{it})$ , and it follows that

$$\begin{aligned} E(z_i) &= \frac{1}{K(K-1)} \left[ K\Lambda_{i1}(1 - \Lambda_{i1}) + K\Lambda_{i2}(1 - \Lambda_{i2}) + K^2(\Lambda_{i1} - \Lambda_{i2})^2 \right. \\ &\quad \left. - K[\Lambda_{i1}(1 - \Lambda_{i1}) + \Lambda_{i2}(1 - \Lambda_{i2}) + (\Lambda_{i1} - \Lambda_{i2})^2] \right] \tag{17} \\ &= (\Lambda_{i1} - \Lambda_{i2})^2. \end{aligned}$$

Hence, the expected value of  $z_i$  is zero under the null hypothesis of binomial dispersion as long as  $x'_{i1}\beta = x'_{i2}\beta$ .

One possible alternative to the null of a binomial variance is given by the beta-binomial model, where the variance is

$$\text{Var}(Y_i) = K\Lambda_{it}(1 - \Lambda_{it})(1 + \eta) \tag{18}$$

and  $\eta$  is equal to  $\eta = \frac{K-1}{\phi+1} > 0$ . In this case, overdispersion originates from random unobserved heterogeneity.

### 3.1 Case I: discrete covariates

Define the set of individuals with the same expectations over time,  $A = \{i : \Lambda_{i1} = \Lambda_{i2}\}$ , for which  $E(z_i|i \in A) = 0$  holds. With time-invariant fixed effect  $\alpha_i$  and a single regressor, the set  $A$  is equal to  $\{i : x_{i1} = x_{i2}\}$ . In general, the set  $A$  is broader, including all cases where  $x'_{i1}\beta = x'_{i2}\beta$ . In most cases, it will be possible to find such a set  $A$  if all covariates are finite discrete variables, assuming that the  $x$ -values are drawn from a stationary distribution. The test term for discrete  $x_{it}$  is defined as:

$$\tau_A = \hat{E}(z_i|i \in A) = \frac{\sum_{i \in A} z_i}{|A|}, \tag{19}$$

where  $|A|$  represents the number of elements in  $A$ . Under  $H_0$ ,  $\tau_A \xrightarrow{p} 0$ . Further, by the central limit theorem (CLT), the statistic  $\tau_A$  converges to a normal distribution,

$$\sqrt{|A|}(\tau_A - 0) \xrightarrow{d} N(0, \sigma_A^2), \tag{20}$$

where  $\sigma_A^2 = \text{Var}(z_i|i \in A)$ . In practice,  $\sigma_A^2$  is replaced by the sample variance  $\hat{\sigma}_A^2$ . So we reject the binomial distribution assumption at the  $\alpha\%$  significance level if  $\left| \frac{\tau_A}{\hat{\sigma}_A/\sqrt{|A|}} \right| \geq c_{1-\frac{\alpha}{2}}$ , where the critical value  $c$  is the  $1 - \alpha/2$ -percentile of the standard normal distribution.

Individuals in the set  $A$  do not contribute to the estimation of the Blogit CML model, since  $x_{it}$  are cancelled out as fixed effects. Nonetheless, they are needed for generating our dispersion test. This nonparametric method to build a test is similar to finding proper cell estimators in matching theory, but likewise faces the curse of dimensionality. It is hard to find the set  $A$  when the dimension of  $x_{it}$  becomes larger. If  $|A|$  shrinks, the convergence rate  $\sqrt{|A|}$  will decrease and the estimator  $\tau_A$  will converge more slowly.

### 3.2 Case II: continuous covariates

The set  $A = \{i : \Lambda_{i1} = \Lambda_{i2}\}$  is empty or very small when  $x_{i1}$  and  $x_{i2}$  are continuous. A more general method uses a kernel estimator for the conditional expectation  $E(z_i | \Lambda_{i1} - \Lambda_{i2} = 0)$ . The main idea is to put more weight on individuals with smaller  $|\Lambda_{i1} - \Lambda_{i2}|$ . Since we do not observe the underlying expectations  $\Lambda_{it}$  directly, we find the set  $A$  by using observables  $x_{it}$ . Under the assumption of a single scalar regressor and time-invariant unobserved heterogeneity, we can decompose the conditional expectation (17) by a Taylor expansion at  $x_{i2}$ ,

$$\begin{aligned} (\Lambda_{i1} - \Lambda_{i2})^2 &= [\Lambda(x_{i1}\beta + \alpha_i) - \Lambda(x_{i2}\beta + \alpha_i)]^2 \\ &= [\Lambda'(x_{i2}\beta + \alpha_i)\beta (x_{i1} - x_{i2}) \\ &\quad + \frac{\Lambda''(x_{i2}\beta + \alpha_i)}{2!}\beta^2 (x_{i1} - x_{i2})^2 + o((x_{i1} - x_{i2})^2)]^2 \\ &= [\Lambda'(x_{i2}\beta + \alpha_i)\beta (x_{i1} - x_{i2})]^2 + o(\beta^2(x_{i1} - x_{i2})^2), \end{aligned}$$

Denote  $\Delta_i = (x_{i1} - x_{i2})\beta$ ,

$$E(z_i | \Lambda_{i1} - \Lambda_{i2})^2 = (\Lambda_{i1} - \Lambda_{i2})^2 = (\Lambda'_{i2}\Delta_i)^2 + o(\Delta_i^2).$$

As the fixed effect  $\alpha_i$  is cancelled out, an alternative conditional expectation function is given by  $\Delta_i$ ,

$$\tau(\Delta) = E(z_i | \Delta_i = \Delta, X_i) = (\Lambda'_{i2}\Delta)^2.$$

Then, under the binomial assumption,

$$E(z_i | \Lambda_{i1} - \Lambda_{i2} = 0) = \tau(0) = 0.$$

The result generalizes to a vector-valued  $x$ , in which case  $\Delta_i = (x_{i1} - x_{i2})'\beta$ .

The next step is to build a kernel estimator for  $\tau(0)$ . One conditional moment estimator for  $\tau(\Delta)$  is  $\hat{\tau}(\Delta) = \frac{\sum_{i=1}^N K(\frac{\Delta_i - \Delta}{h})z_i}{\sum_{i=1}^N K(\frac{\Delta_i - \Delta}{h})}$ , where  $h$  is the kernel bandwidth for  $\Delta_i$  and  $K(\frac{\Delta_i - \Delta}{h})$  is the kernel function. For a given sample,  $\Delta_i$  needs to be replaced by  $\hat{\Delta}_i = (x_{i1} - x_{i2})\hat{\beta}$ , where  $\hat{\beta}$  is estimated. We can use the Blogit CML estimator for

estimation, as it is consistent under the binomial null hypothesis. We construct a local estimate  $\hat{\tau}$  for the object of interest  $\tau(0)$  (see Pagan and Ullah 1999):

$$\hat{\tau} = \frac{\sum_{i=1}^N K\left(\frac{\hat{\Delta}_i}{h}\right)z_i}{\sum_{i=1}^N K\left(\frac{\hat{\Delta}_i}{h}\right)} = \sum_{i=1}^N w_{ni}z_i, \quad w_{ni} = \frac{K\left(\frac{\hat{\Delta}_i}{h}\right)}{\sum_{i=1}^N K\left(\frac{\hat{\Delta}_i}{h}\right)},$$

The Gaussian function  $K\left(\frac{\hat{\Delta}_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\hat{\Delta}_i/h)^2}{2}\right)$  is chosen for simplicity.

### 3.3 Asymptotic properties

Let  $f = f(\Delta = 0)$  denote the continuous density function of the random variable  $\Delta$  at point 0. The kernel density estimator  $\hat{f}$  for  $f$  is

$$\hat{f} = \sum_{i=1}^N \frac{K\left(\frac{\hat{\Delta}_i}{h}\right)}{nh}.$$

In addition, rewrite  $z_i$  as the sum of its conditional expectation  $E(z_i|\Delta_i) = \tau(\Delta_i)$  and an error term  $u_i$ , such that

$$z_i = \tau(\Delta_i) + u_i = (\Lambda'_{i2}\Delta_i)^2 + u_i$$

where  $E(u_i|\Delta_i, X_i) = 0$  and  $\text{Var}(u_i|\Delta_i, X_i) = \sigma^2$ .

The estimator  $\hat{\tau}$  is a combination of  $\hat{f}$  and  $z_i$

$$\hat{\tau} = \frac{\sum_{i=1}^N \frac{1}{nh} K\left(\frac{\hat{\Delta}_i}{h}\right)z_i}{\sum_{i=1}^N \frac{1}{nh} K\left(\frac{\hat{\Delta}_i}{h}\right)} = \frac{1}{\hat{f}} \sum_{i=1}^N \frac{1}{nh} K\left(\frac{\hat{\Delta}_i}{h}\right)z_i = \frac{1}{\hat{f}} \sum_{i=1}^N \frac{1}{nh} K\left(\frac{\hat{\Delta}_i}{h}\right) (\Lambda'_{i2}\Delta_i)^2 + u_i.$$

The expectation of  $\hat{\tau}$  is

$$\begin{aligned} E(\hat{\tau}) &= E\left(\frac{1}{\hat{f}} \sum_{i=1}^N \frac{1}{nh} K\left(\frac{\hat{\Delta}_i}{h}\right) (\Lambda'_{i2}\Delta_i)^2 + \frac{1}{\hat{f}} \sum_{i=1}^N \frac{1}{nh} K\left(\frac{\hat{\Delta}_i}{h}\right) u_i\right) \\ &= \int \int \frac{1}{h\hat{f}} K(v)(\Lambda'_{i2})^2 (hv)^2 f(hv, \Lambda_{i2}) h dv d\Lambda_{i2} \\ &\quad + \hat{E}(u_i|\hat{\Delta} = 0), \quad \text{where we replace } \Delta = hv \\ &= h^2 \int \int K(v)(v)^2 (\Lambda'_{i2})^2 \frac{f(v, \Lambda_{i2})}{\hat{f}} dv d\Lambda_{i2} \\ &= h^2 \mu_2 E[(\Lambda'_{i2})^2|\Delta = 0], \quad \text{where } \mu_2 = \int K(v)(v)^2 dv \end{aligned}$$

We therefore obtain a bias

$$\text{Bias}(\hat{\tau}) = E(\hat{\tau}) - \tau(0) = E(\hat{\tau}) - h^2\mu_2 E[(\Lambda'_{i2})^2 | \Delta = 0], \tag{21}$$

that is proportional to  $h^2$ .

To guarantee consistency of the estimator  $\hat{\tau}_n$ , convergence of the mean square error to zero is required. The MSE is equal to  $\text{MSE}(\hat{\tau}) = \text{Bias}(\hat{\tau})^2 + \text{Var}(\hat{\tau})$ . So the bias for  $\tau_n$  should decrease to zero, as  $n$  increases:

$$h^2 \longrightarrow 0, \quad \text{as } n \longrightarrow \infty. \tag{22}$$

Besides the convergence condition for bias, we also consider the asymptotic performance of the variance of  $\hat{\tau}$ . Using a result on the variance of conditional expectations from Pagan and Ullah (1999), we obtain:

$$\text{Var}(\hat{\tau}) = \frac{\sigma^2}{nhf} \int K^2(v)dv, \quad \text{Var}(\hat{\tau}) \propto \frac{1}{nh}. \quad \text{If } n \longrightarrow \infty, \quad \frac{1}{nh} \longrightarrow 0. \tag{23}$$

To make sure that the MSE converges at the fastest speed, bias<sup>2</sup> and variance should converge at the same rate:  $h^4 \propto \frac{1}{nh}$ . Otherwise, the slower speed dominates the convergence rate. Thus,  $h$  is of order  $h \propto n^{-\frac{1}{5}}$  and by the central limit theorem,

$$\sqrt{nh}(\hat{\tau} - E(\hat{\tau})) \xrightarrow{d} N(0, f^{-1}\sigma^2 \int K^2(v)dv) \tag{24}$$

Here,  $\sigma^2 = \text{Var}(z_i^2 | \Delta = 0)$ , with the same definition as in the discrete case (Eq. 20). In practice, we standardize  $\Delta_i$  at first and set bandwidth  $h' = 0.9n^{-\frac{1}{5}}$ . The approximate bias is calculated by  $\hat{E}(\hat{\tau}) = \sum_{i=1}^N w_{ni} (y_{i2}(1 - y_{i2})\hat{\Delta})^2$ ,  $\sigma^2$  is replaced by  $\hat{\sigma}^2 = \sum w_{ni} (z_i - \tau(\hat{\Delta}_i))^2$  and  $\widehat{\text{Var}}(\hat{\tau}) = \frac{\hat{\sigma}^2 \sum_{i=1}^N K^2(v)}{f\hat{f}^2 n^2 h^2}$ . Hence,  $\frac{\hat{\tau} - \hat{E}(\hat{\tau})}{\sqrt{\widehat{\text{Var}}(\hat{\tau})}}$  can be used as a  $t$ -test.

### 3.4 Multiple periods

The test can be extended to multiple time periods. With  $T = 2$ , there is a single moment condition for  $E(z_i | \Delta_i = 0)$  that can be tested. For  $T > 2$ , one possibility is to combine  $T - 1$  such moment conditions into a single test statistic.

In the discrete case, for set  $A_t = \{i : x_{i,t} = x_{i,t+1}\}$ ,

$$g_{i,t} = \frac{(Y_{i,t} - Y_{i,t+1})^2 - K(M_{i,t} - M_{i,t+1})^2}{K(K - 1)}, \quad t = 1, \dots, T - 1.$$

$g_{i,t}$  is empty if  $x_{i,t} \neq x_{i,t+1}$ . As we derived before,  $E(g_{i,t}) = 0$ .

In matrix form,

$$g_i = \begin{pmatrix} g_{i,1} \\ \dots \\ g_{i,T-1} \end{pmatrix}, \text{ and the sample mean is } \bar{g}_n = \begin{pmatrix} \frac{1}{n_1} \sum_{i=1}^{n_1} g_{i,1} \\ \dots \\ \frac{1}{n_{T-1}} \sum_{i=1}^{n_{T-1}} g_{i,T-1} \end{pmatrix},$$

with  $n_t = |A_t|$ , the cardinality of set  $A_t$ . Denote  $n = (n_1, \dots, n_{T-1})'$ . To calculate the sample variance-covariance matrix  $\hat{S}$ , we replace off-diagonal elements with pairwise sample covariances and diagonal ones with  $g_t$  sample variances. A test statistics can be derived

$$J = (\sqrt{n} \circ \bar{g}_n)' \hat{S}^{-1} (\sqrt{n} \circ \bar{g}_n).$$

In the continuous case, moment conditions are

$$g_{i,t} = \frac{K\left(\frac{\hat{\Delta}_{i,t}}{h}\right) (z_{i,t} - \tau(\hat{\Delta}_{i,t}))}{\sum_{i=1}^N \frac{1}{n} K\left(\frac{\hat{\Delta}_{i,t}}{h}\right)}, \quad t = 1, \dots, T - 1$$

where  $\hat{\Delta}_{it} = (x_{it} - x_{i,t+1})\hat{\beta}$  and  $\tau(\hat{\Delta}_{i,t}) = (y_{i,t+1}(1 - y_{i,t+1})\hat{\Delta}_{it})^2$ . Under the null hypothesis,  $E(g_{i,t}) = 0$ .

These moment conditions can be written in matrix form for individual  $i = 1, \dots, N$  as:

$$g_i = \begin{pmatrix} g_{i,1} \\ \dots \\ g_{i,T-1} \end{pmatrix}, \text{ and the sample mean is } \bar{g}_N = \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N g_{i,1} \\ \dots \\ \frac{1}{N} \sum_{i=1}^N g_{i,T-1} \end{pmatrix}.$$

Let  $\hat{S}$  denote the sample variance:

$$\hat{S} = \frac{1}{N} \sum_{i=1}^N g_i g_i' - \bar{g}_N \bar{g}_N',$$



Since  $\bar{g}_N \xrightarrow{p} E(g_i) = 0$ , a test statistic is given by

$$J = N \bar{g}'_N \hat{S}^{-1} \bar{g}_N = (\sqrt{N} \cdot \bar{g}_N)' \hat{S}^{-1} (\sqrt{N} \cdot \bar{g}_N)$$

Therefore,  $\sqrt{N} \cdot \bar{g}_N \xrightarrow{d} N(0, S)$ ,  $\hat{S} \xrightarrow{p} S$ , and  $J \xrightarrow{d} \chi^2_{T-1}$ .

The Chi-square test rejects the binomial distribution assumption at the  $\alpha\%$  significance level if  $J \geq \chi^2_{\alpha}(T - 1)$ .

### 3.5 Simulation study

We conduct a number of simulation experiments to examine the performance of these tests under two scenarios. In the first setting, explanatory variables are discrete (in fact, there is a single binary regressor, to keep things as simple as possible), while the explanatory variable is continuous in the second. The remaining aspects of the DGP regarding fixed effects, expectation functions and parameters setting are the same as those in Sect. 2.3.

Table 2 presents rejection rates, i.e. the relative number of times that our test rejects the binomial assumption over 1000 replications, when  $x$  is discrete.  $x_{it}$  is either 0 or 1 with equal probability. In this case,  $\Pr(x_{i1} = x_{i2}) = 50\%$ , and on average half of the observations will be in the set  $A$  of individuals with the same expectations over time and thus informative for computing the test statistic. As before, the number of time periods increases from  $T = 2$  to  $T = 10$ , and binomial parameter from  $K = 2$  to  $K = 10$ .

The first row of each subpanel shows results without overdispersion, i.e. sampling from a binomial DGP applies. In this case, the rejection rates are equivalent to the proportion of type-I errors and ideally should be close to the nominal size of the test, in this case 5%. The lower part of each subpanel shows the rejection rates under  $H_0$  when  $H_0$  is false, i.e. the power.

When we implemented the multi-period discrete test as described in Sect. 3.4, we found that the size of the test was seriously distorted when  $T$  was large. For  $T = 10$ , the rejection rates under the binomial assumption were 44.7% for  $N = 100$ , and 12.9% for  $N = 500$ . For larger  $N$ , there is a convergence to the nominal size, but it is rather slow (e.g. 8.5% rejection rate for  $N = 1000$ ). The reason for this test behaviour is the poor estimation of the covariance elements of the weighing matrix  $\hat{S}$ . For example, when  $T = 3$ , the covariance between  $g_{i,1}$  and  $g_{i,2}$  is estimated based on the small subset of observations for which  $x_{i1} = x_{i2} = x_{i3}$ . The imprecise estimation of the covariances for small  $N$  leads to a large sampling variability of  $\hat{S}$ , and this problem increases with  $T$ . As an alternative, we therefore show in Table 2 simulation results, where all off-diagonal elements of  $\hat{S}$  were set to zero, leading to a better performance of the test in small samples. We also note that the continuous, kernel-based test does not suffer from this problem.

Reassuringly, we find that the test has some power against the alternative of rather modest overdispersion (10%), in particular for  $N = 500$ ,  $K = 2$  and  $T = 10$ , where around 36% of wrong null hypotheses are rejected. As the dispersion degree increases,

**Table 2** Simulation results for rejection rates when  $x$  is discrete

$N = 100$	Overdispersion (%)	$T = 2$	$T = 5$	$T = 10$	$N = 500$	Overdispersion (%)	$T = 2$	$T = 5$	$T = 10$
$K = 2$	0	0.062	0.062	0.057	$K = 2$	0	0.032	0.062	0.055
	10	0.080	0.093	0.130		10	0.167	0.258	0.358
	50	0.502	0.817	0.927		50	0.992	1	1
$K = 5$	0	0.056	0.064	0.053	$K = 5$	0	0.059	0.058	0.066
	10	0.044	0.068	0.066		10	0.102	0.147	0.187
	50	0.225	0.449	0.579		50	0.885	0.999	1
$K = 10$	100	0.603	0.929	0.990	$K = 10$	100	0.999	1	1
	200	0.938	0.999	1		200	1	1	1
	0	0.052	0.053	0.069		0	0.053	0.055	0.055
$K = 10$	10	0.052	0.060	0.071	$K = 10$	10	0.075	0.118	0.185
	50	0.204	0.395	0.497		50	0.836	0.993	1
	100	0.492	0.878	0.972		100	1	1	1
200	0.869	1	1	200	1	1	1		

$x_{it}$  is a binary variable with 50% probability equal to 0 or 1. The remaining DGP is the same as in Table 1; the null hypothesis is that of binomial dispersion

the power of the test also grows, and it reaches 100% for DGPs where overdispersion, the number of observations, and the number of time periods are large.

In Table 3, we show the results for the kernel weighted test statistics for continuous regressors.  $x_{it}$  is drawn from a uniform distribution with positive support between  $-1$  and  $1$ , with mean  $0$  and variance  $1/3$ . The general patterns regarding type-I errors and power of the tests are mostly similar to those of Table 2. As in Table 2, the power of the test tends to decrease in  $K$ , for a given overall degree of overdispersion, but this tendency is more uniform in the continuous version of the test. This indicates that the power of the test reacts differently to the two parameters driving overdispersion, and in particular that it is more sensitive to increases in  $\phi$  rather than  $K$ . The combined results from our simulation experiments are reassuring: on the one hand, modest amounts of overdispersion cause only minor bias of the Blogit CML estimator; on the other hand, the test we derive has good power properties against medium- or high-dispersion alternatives to the binomial assumption.

## 4 Application to labour supply

In this illustrative application of the binomial estimator and overdispersion test, we re-consider the association between fertility and female labour supply. Data are from the Swiss Household Panel (SHP) for the years 2012–2016. The SHP is an ongoing longitudinal survey of people residing in Switzerland. It collects information on a large range of topics on living conditions, both objective and subjective, including work, fertility and health. We restrict the analysis to women aged 25–45, who participated in the survey at least twice during the 5-year period. This gives us a panel of 5854 person-year observations for 1712 different women.

There exists a huge literature modelling female labour supply, a large part of which is devoted to the endogeneity of the fertility decision. We want to make here a different point, namely that the labour supply outcome, i.e. the amount of time a women decides to spend in market work, fits in principle into the empirical framework discussed in this paper and hence can be analysed using the methods proposed in this paper: empirically, the amount of days or hours worked is discrete, and it has a lower bound of zero, as well as an upper bound, and can thus be expressed as a proportion.

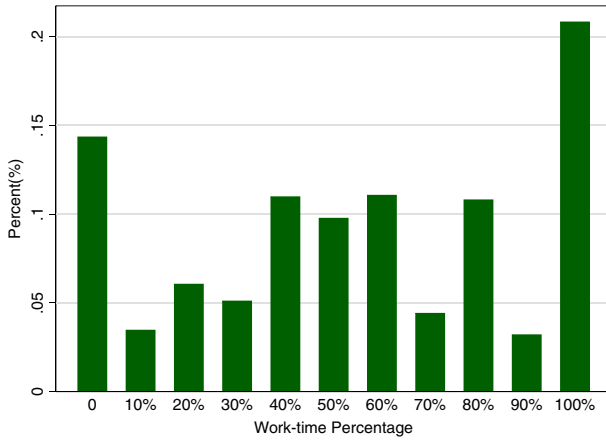
Modelling labour supply as a fraction of time may be promising in particular in institutional settings, where employment contracts offer various part-time options. A case in point is Switzerland, where vacancies are advertised, and work contracts are written using full-time fractions. For instance, 60% work-time means that the worker is employed for the equivalent of 3 days per week and also is paid 60% of a full-time salary. In practice, the large majority of agreed-upon work-time percentages are multiples of 10%.

Figure 3 shows the distribution of work-time percentages for the sample of women extracted from the Swiss Household Panel. Here, the data are pooled over the five years. The relative frequency of zeros is 14.4%, meaning that the estimated participation rate in our sample for this age group is 85.6%, a number very close to the official statistic published by the Federal Statistical Office (BfS 2016). Although there are peaks for non-work and for full-time work, all intermediate values are present in the data.

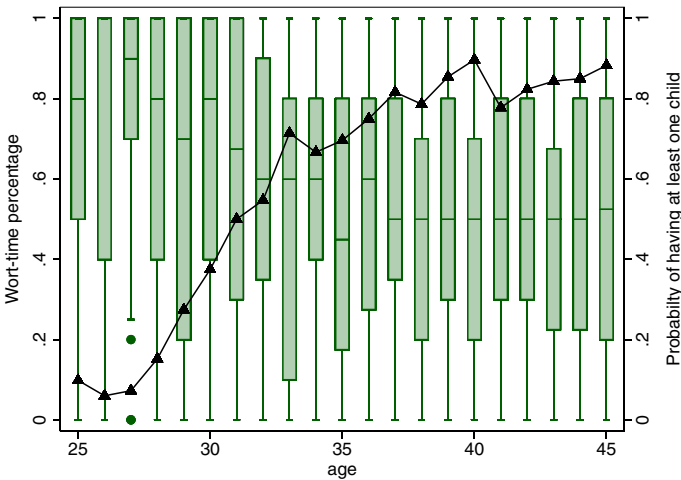
**Table 3** Simulation results for rejection rates when  $x$  is continuous

$N = 100$	Dispersion (%)	$T = 2$	$T = 5$	$T = 10$	$N = 500$	Dispersion (%)	$T = 2$	$T = 5$	$T = 10$
$K = 2$	0	0.053	0.063	0.097	$K = 2$	0	0.062	0.047	0.056
	10	0.070	0.095	0.185		10	0.138	0.223	0.394
	50	0.424	0.684	0.899		50	0.948	1	1
$K = 5$	0	0.05	0.053	0.078	$K = 5$	0	0.042	0.05	0.057
	10	0.044	0.063	0.110		10	0.083	0.105	0.190
	50	0.239	0.512	0.805		50	0.772	0.992	1
$K = 10$	100	0.608	0.958	0.999	$K = 10$	100	0.996	1	1
	200	0.927	0.991	0.995		200	1	1	1
	0	0.041	0.062	0.091		0	0.052	0.056	0.043
$K = 10$	10	0.041	0.057	0.086	$K = 10$	10	0.073	0.094	0.179
	50	0.190	0.402	0.673		50	0.677	0.988	1
	100	0.482	0.908	0.997		100	0.991	1	1
200	0.873	0.999	0.999	200	1	1	1		

Results for 1000 Monte Carlo replications; standard deviations in parentheses. In each period, the number of observation is 100.  $x_{it} \sim U[-1, 1]$ .  $\beta_1 = 2$  and  $\alpha_i = \sqrt{T} \bar{x}_i + N(0, 1)$ . Overdispersion factor  $\frac{k-1}{\phi+1}$  represents dispersion degree of variance. Overdispersion degree 0% is generated by binomial distribution, and positive dispersion degree is generated by a beta-binomial DGP



**Fig. 3** Distribution of Work-time percentage in 2016. *Source:* Swiss Household Panel 2012–2016. Women aged 25–45. “Work-time percentage” denotes the contracted work-time. For example, 60% = 3 full days per week; 0 indicates non-work, 100% full-time work



**Fig. 4** Work-time percentage and motherhood by age in 2016. *Notes:* The green bar is the box plot of the work-time percentage by age in 2016; the black triangle denotes the probability of having at least one child in 2016. (Color figure online)

In particular, Fig. 3 documents that for Switzerland, the vast majority of women does not work full-time. A question one can then ask is: How does the work-time percentage vary with the presence of children in the household? Box-plots in Fig. 4 show, for our data, a clear negative association between work and children. The median work-time percentage drops from 80% or higher for those aged 30 or below to 50% for women in their early 40s. At the same time, older women are more likely to have children.

**Table 4** Descriptive statistics ( $NT = 5854$ ). *Source:* Swiss Household Panel 2012–2016, own calculations

	Mean	SD
Work-time percentage	0.557	0.346
Age	36.30	6.01
Children (yes = 1)	0.631	0.482
Partner (yes = 1)	0.584	0.492
Self-rated health (0: worst to 4: best)	3.114	0.610
French speaking (yes = 1)	0.293	0.455
Italian speaking (yes = 1)	0.043	0.204

The key assumption of the following analysis is that we can treat 10-times the work-time percentage as a binomial variable with outcomes  $0, 1, \dots, 10$ , where the mean will be modelled as a function of covariates as well as individual-specific time-invariant fixed effects.

It is of course difficult to imagine the work-time decision as literally arising from an underlying sequence of  $K$  independent Bernoulli trials. Nevertheless, the binomial model can provide a useful approximation to the distribution of work-time percentages, in particular since it conditions on fixed effects and hence is compatible with an observed unconditional “W-shaped” outcome distribution as observed in Fig. 3. Our illustrative application abstracts from additional complexities, such as potential differences between the participation decision (i.e. the extensive margin) and the intensive margin, and, beyond the inclusion of individual fixed effects, the endogeneity of the fertility decision. Clearly, these are important concerns, and they should be addressed in further extensions of the approach.

## Results

Table 4 provides some descriptive statistics (means and standard deviations) for both the dependent and the explanatory variables used in the estimation. The average work-time percentage is 56%, with a standard deviation of 0.34. Under the binomial assumption, the standard deviation for a fraction with a mean of 0.56 is equal to  $\sqrt{0.56(1 - 0.56)/10} = 0.157$ , substantially below the observed standard deviation of 0.346. Hence, there is evidence of overdispersion at the marginal level.

Women have an average age of 36.3 years, and 63.1% report having at least one child in the year they are surveyed. For 58.4% of person-year observations, there is a partner present in the household. The health status is captured by a 5-point scale for self-assessed health, where 0 means “not well at all” and 4 means “very well”. We treat it as a cardinal scale for simplicity, and also abstract from its potential endogeneity to working or having children. Finally, we include information on language region. There is quite a bit of evidence that work-norms differ between the French and the German-speaking parts of Switzerland, with some stigma attached to working mothers, in particular during the first years of the child’s life (see Steinhauer 2018). This stigma

seems to be stronger in the German-speaking part of Switzerland (65% of our sample) but less so in the French-speaking part (29% of our sample).

Our final estimation model includes four year dummies, age-squared (the linear age term is dropped; alternatively, one could identify the linear age effect by setting a second year effect equal to zero), indicators for the presence of a child and partner, and the health variable. Since language region is mostly constant over time, it is near-collinear with the fixed effects when applying the Blogit CML or Blogit DV estimators, and we therefore only include its interaction with the child-indicator variable.

As is the case for the binary logit model with fixed effects, DV estimation of the binomial model is subject to the perfect prediction problem (see, for example, Kunz et al. 2018). Outcomes for women, whose work-time percentage is either zero or one in each year, are perfectly predicted, meaning that the associated dummy coefficient will tend to minus or plus infinity, respectively. For the Blogit CML, perfect prediction formally does not arise as the  $\alpha_i$ 's are not estimated. However, all such observations have mechanically a log-conditional likelihood value of zero and thus do not contribute to estimation of  $\beta$  either. To use the same estimation sample everywhere, we right away drop all perfectly predicted outcomes, leading to a final sample size of 4661 person-year observations for the work-time percentage model.

Regression results are given in Table 5. The first column shows the estimated coefficients from the Blogit CML and the second those from the Blogit DV model. The last two columns add corresponding (binary) logit models for the extensive margin model (work yes/no), again using alternatively the CML or DV estimators. Standard errors are clustered at the individual level. The linear age term has been dropped due to collinearity in a model with individual and year fixed effects.

When interpreting magnitudes, we note the recent suggestion by Kemp and Santos Silva (2016) and focus on expected (semi-) elasticities. These can be estimated without knowledge of  $\alpha_i$  and are thus very suitable for our conditional maximum likelihood approach. For the binomial proportion model with  $E(y_{it}|x_{it}, \alpha_i) = \Lambda_{it}$ , we obtain

$$\partial \log E(y_{it}|x_{it}, \alpha_i) / \partial x_{it} = \beta(1 - \Lambda_{it})$$

A good estimator of the overall mean of  $\Lambda_{it}$  is the sample mean of the outcome,  $\bar{\Lambda} = \bar{y} = 0.55$ , so that the CML estimators  $\hat{\beta}$  can be multiplied by 0.45 to obtain an estimate of the population average semi-elasticities with respect to changes in the associated covariate.

From columns (1) and (2) of Table 5, we find a large negative association between having a child and the amount of work. The point estimate of the main effect is about -2, which means that not having a child increases the expected work-time percentage by about 90 percent. This effect is highly statistically significant, as are two of the three interaction effects: having a child reduces the work-time percentage more if a partner is present than otherwise, underlining the relevance of pecuniary motives for work, and the need to “make ends meet”. The labour supply response of women to having children is about half as large for French-speaking women as it is for German speak-

**Table 5** Determinants of female labour supply (2012–2016). *Source:* Swiss Household Panel

	Work-time percentage		Work (yes/no)	
	Blogit CML	Blogit DV	Logit CML	Logit DV
Age-squared	−0.001 (0.002)	−0.001 (0.002)	0.005 (0.005)	0.007 (0.008)
Self-rated health	0.071 (0.038)	0.073 (0.039)	0.448 (0.153)	0.642 (0.220)
Partner (yes=1)	0.322 (0.253)	0.333 (0.260)	1.371 (0.658)	1.927 (0.952)
Children (yes=1)	−2.097 (0.276)	−2.160 (0.286)	−1.975 (0.956)	−2.771 (1.381)
Children × Partner	−0.824 (0.260)	−0.848 (0.268)	−2.216 (1.056)	−3.188 (1.510)
Children × French	1.152 (0.405)	1.194 (0.418)	1.876 (1.089)	2.777 (1.619)
Children × Italian	−0.247 (0.720)	−0.266 (0.754)		
Year 2013	0.143 (0.147)	0.147 (0.151)	−0.018 (0.465)	−0.018 (0.647)
Year 2014	0.246 (0.278)	0.253 (0.285)	−0.147 (0.820)	−0.265 (1.143)
Year 2015	0.338 (0.412)	0.347 (0.423)	−0.247 (1.205)	−0.408 (1.677)
Year 2016	0.387 (0.545)	0.397 (0.560)	−0.449 (1.577)	−0.717 (2.204)
Number of person-years	4661	4661	1071	1071
Number of persons	1334	1334	295	295
Log pseudo-likelihood	−23,183.6	−1838.3	−358.8	−595.9
Fixed effects	Yes	Yes	Yes	Yes

*Blogit CML* denotes the binomial logit conditional maximum likelihood estimator; *Blogit DV* is the binomial logit estimator with dummy variables for each individual; *Logit CML* and *Logit DV* are the corresponding estimators for the binary logit model

ers, corroborating the social norm results found in the earlier literature (Steinhauer 2018).

In this application, the Blogit CML and the Blogit DV results are very similar. The DV results are always a bit larger in absolute value, but the difference never exceeds 5%. This resonates with our simulation results, because both  $T$  and  $K$  are relatively large. Nevertheless, the joint test for the binomial assumption derived in Sect. 3.3 indicates a clear rejection (test value of 37.7 with a  $\chi^2_{0.95}$  critical value of 9.5). This rejection result due to overdispersion was already foreshadowed, although not logically implied because of the conditional nature of the test, by the high proportion of no work (zero) and full-time work (100%) as evident in Fig. 1. However, we know from the simulation



results (Tables 1, 2) that even with 50% overdispersion, the bias of the Blogit CML is small for  $K = 10$  and  $T = 5$ , a setting similar to the current application. At the same time, the probability of rejecting the wrong  $H_0$  is very close to 1 (see Table 3). On a practical note, the CML estimator can be computed much faster than the DV estimator, by a factor of about 10 in our case. The speed problem of DV models would be exacerbated in applications with more cross-sectional units, to the point where computation of the Blogit DV estimator may become infeasible in the current Stata/R setting.

In the last two columns of Table 5, we allow for a comparison with results from a more conventional binary logit extensive margin estimator. A first point to note is that the effective sample becomes much smaller, since all observations with variation in the positive range only, i.e. percentages between 10 and 100%, are now coded as “1” and thus become perfectly predicted. Their variation does not contribute to estimation, the usable sample size drops by 3/4, and the standard errors of the estimated coefficients increase accordingly. We had to drop the interaction between speaking Italian and having children, as it could not be estimated in the reduced sample.

The estimated coefficients tend to be substantially larger, but they are not directly comparable. To obtain the implied expected semi-elasticities for the probability of work, coefficients need to be multiplied by the non-participation rate, 0.145 in this case, compared to a factor of 0.45 applicable in the first two columns. Based on the CML estimates, some of the extensive margin semi-elasticities are smaller than the overall semi-elasticities (like the main effect of having a child), and some of them larger (such as self-rated health). In terms of statistical significance, we find that the health and partner coefficients were not significant in the work-time percentage model, but they are in the participation model. And in terms of point estimates, the interaction between speaking French and having children just offsets the main effect of having at least one child, meaning that there is no difference in participation probabilities for French-speaking mothers and non-mothers, although some labour supply responsiveness was found in the work-time percentage model for the combined extensive and intensive margin effect. Also, the participation model suffers from a massive incidental parameters bias, since the point estimates for the DV estimator exceed those of the CML estimator by 50% on average.

## 5 Concluding remarks

Although Machado (2004) introduced the fixed effects binomial model as a method for proportions of successes in a sequence of Bernoulli trials, it can be used for discrete bounded outcomes, or fractions, more generally. However, it remained an open question whether or not the conditional binomial logit maximum likelihood estimator is robust to misspecification. In this paper, we focus on the consequences of overdispersion as it originates, for instance, from neglected unobserved heterogeneity. We show in simulation experiments that the Blogit CML estimator maintains a rather good performance even if the binomial model is misspecified, as long as the length of the panel  $T$  is sufficiently large, or the degree of overdispersion is modest.

We then derive a test of the null hypothesis that the binomial assumption is valid, based on departures from the implied binomial variance function. The test computes the variance of within-individual outcome differences. For the subset of observations whose regressors do not change over time, the mean difference is zero (and close to zero if regressors do not differ “too much”) and it is possible to construct and compare two variance estimates, one with and one without the binomial assumption, that both do not depend on fixed effects. This is essential, as fixed effects are not estimated by the Blogit CML estimator. Our simulation experiments show that the test has good power properties against the alternative of medium or large degrees of overdispersion. But these are exactly the cases where the bias of the Blogit CML estimator becomes noticeable.

We study in our empirical application an outcome related to women’s work behaviour, namely the contracted work-time percentage. In our sample of mid-aged women obtained from the Swiss Household Panel, 65% of all women report working part-time, i.e. a percentage between 10 and 90%. The empirical analysis using the fixed effects binomial logit model predicts substantially different work-time percentages for mothers and non-mothers. Having a partner makes the difference more pronounced, whereas speaking French reduces it. We show how these coefficients can be interpreted in terms of expected semi-elasticities even if the fixed effects are not estimated. In comparison with the fixed effects logit estimation for the participation model, much fewer observations are lost in the work-time percentage model due to perfect prediction, contributing to a much more precise estimation of the model parameters.

In future work, we will consider alternative estimators that could be pursued if the binomial null hypothesis is rejected. If the logit conditional expectation function is to be kept, a binomial logit correlated random effects model is a possible approach. Such a model would explicitly account for overdispersion, by assuming for instance that random, time-varying unobserved heterogeneity follows a normal distribution with mean depending on the regressors.

**Funding** Open Access funding provided by University of Zurich.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A

See Table 6.

**Table 6** Simulation results  $N = 100$

$N = 100$	Overdispersion (%)	$T = 2$			$T = 5$			$T = 10$		
		Blogit CML	Blogit DV	Pooled logit	Blogit CML	Blogit DV	Pooled logit	Blogit CML	Blogit DV	Pooled logit
$K=2$	0	2.049 (0.419)	2.880 (0.621)	2.242 (0.255)	2.003 (0.178)	2.280 (0.211)	1.986 (0.145)	2.006 (0.118)	2.134 (0.128)	1.877 (0.101)
	10	2.100 (0.419)	2.968 (0.632)	2.225 (0.263)	2.036 (0.192)	2.320 (0.227)	1.988 (0.156)	2.010 (0.121)	2.138 (0.131)	1.876 (0.105)
	50	2.516 (0.822)	3.674 (1.283)	1.898 (0.399)	2.179 (0.360)	2.484 (0.429)	1.908 (0.279)	2.073 (0.246)	2.204 (0.266)	1.937 (0.222)
$K = 5$	0	2.012 (0.233)	2.279 (0.272)	2.243 (0.190)	2.000 (0.111)	2.101 (0.118)	1.990 (0.104)	2.002 (0.073)	2.051 (0.075)	1.872 (0.076)
	10	2.033 (0.252)	2.303 (0.295)	2.239 (0.184)	2.013 (0.120)	2.115 (0.128)	1.990 (0.115)	2.010 (0.079)	2.059 (0.081)	1.875 (0.078)
	50	2.139 (0.300)	2.431 (0.354)	2.205 (0.214)	2.054 (0.139)	2.159 (0.149)	1.971 (0.122)	2.032 (0.086)	2.081 (0.088)	1.882 (0.083)
100	2.320 (0.402)	2.657 (0.485)	2.140 (0.253)	2.111 (0.176)	2.219 (0.188)	1.958 (0.147)	2.052 (0.116)	2.102 (0.119)	1.898 (0.104)	
200	2.721 (0.736)	3.167 (0.910)	1.887 (0.354)	2.231 (0.306)	2.345 (0.326)	1.891 (0.246)	2.115 (0.208)	2.167 (0.214)	1.948 (0.186)	

Table 6 continued

N = 100	Overdispersion (%)	T = 2			T = 5			T = 10		
		Blogit CML	Blogit DV	Pooled logit	Blogit CML	Blogit DV	Pooled logit	Blogit CML	Blogit DV	Pooled logit
K = 10	0	2.005 (0.157)	2.129 (0.169)	2.241 (0.153)	2.001 (0.078)	2.050 (0.081)	1.989 (0.091)	2.000 (0.052)	2.025 (0.052)	1.871 (0.063)
	10	2.017 (0.174)	2.142 (0.187)	2.237 (0.158)	2.004 (0.082)	2.053 (0.084)	1.986 (0.091)	2.002 (0.055)	2.026 (0.055)	1.868 (0.063)
	50	2.065 (0.202)	2.195 (0.219)	2.241 (0.175)	2.026 (0.098)	2.076 (0.101)	1.984 (0.100)	2.012 (0.064)	2.036 (0.065)	1.869 (0.072)
	100	2.139 (0.239)	2.275 (0.260)	2.214 (0.183)	2.051 (0.109)	2.102 (0.113)	1.979 (0.106)	2.025 (0.074)	2.049 (0.075)	1.876 (0.076)
	200	2.293 (0.325)	2.443 (0.354)	2.146 (0.215)	2.105 (0.148)	2.157 (0.153)	1.963 (0.124)	2.049 (0.095)	2.073 (0.096)	1.892 (0.087)

Results for 1000 Monte Carlo replications; standard deviations in parentheses. In each period, the number of observation is 100.  $x_{it} \sim U[-1, 1]$ .  $\beta_1 = 2$  and  $\alpha_i = \sqrt{T} \tilde{x}_i + N(0, 1)$ . Overdispersion factor  $\frac{k-1}{\phi+1}$  represents dispersion degree of variance. Data for overdispersion degree 0 are generated from the binomial model, and data with positive dispersion degree are generated by a beta-binomial DGP

Appendix B

See Table 7

Table 7 Simulation results  $N = 500$

$N = 500$	Overdispersion (%)	$T = 2$			$T = 5$			$T = 10$		
		Blagit CML	Blagit DV	Pooled logit	Blagit CML	Blagit DV	Pooled logit	Blagit CML	Blagit DV	Pooled logit
$K = 2$	0	2.013 (0.170)	2.826 (0.254)	2.234 (0.112)	2.002 (0.082)	2.278 (0.097)	1.987 (0.068)	2.002 (0.053)	2.130 (0.057)	1.870 (0.046)
	10	2.078 (0.195)	2.936 (0.293)	2.214 (0.120)	2.026 (0.084)	2.307 (0.099)	1.975 (0.070)	2.011 (0.058)	2.139 (0.062)	1.870 (0.050)
	50	2.361 (0.311)	3.424 (0.489)	1.849 (0.175)	2.142 (0.151)	2.438 (0.180)	1.881 (0.120)	2.062 (0.102)	2.191 (0.110)	1.926 (0.092)
$K = 5$	0	2.003 (0.102)	2.268 (0.119)	2.232 (0.078)	2.001 (0.051)	2.102 (0.054)	1.986 (0.050)	2.000 (0.033)	2.049 (0.034)	1.867 (0.033)
	10	2.032 (0.108)	2.302 (0.126)	2.237 (0.088)	2.007 (0.052)	2.109 (0.055)	1.984 (0.051)	2.006 (0.035)	2.055 (0.036)	1.868 (0.034)
	50	2.143 (0.132)	2.435 (0.156)	2.204 (0.090)	2.052 (0.063)	2.157 (0.068)	1.973 (0.055)	2.023 (0.041)	2.073 (0.043)	1.871 (0.040)
$K = 10$	100	2.289 (0.168)	2.617 (0.203)	2.121 (0.107)	2.103 (0.080)	2.211 (0.085)	1.951 (0.063)	2.050 (0.049)	2.100 (0.051)	1.895 (0.046)
	200	2.619 (0.292)	3.034 (0.363)	1.852 (0.152)	2.225 (0.137)	2.339 (0.146)	1.882 (0.109)	2.105 (0.093)	2.156 (0.095)	1.932 (0.083)
	0	1.997 (0.071)	2.121 (0.076)	2.235 (0.067)	1.999 (0.034)	2.049 (0.035)	1.984 (0.042)	2.000 (0.023)	2.024 (0.023)	1.867 (0.029)
$K = 10$	10	2.014 (0.075)	2.140 (0.081)	2.233 (0.073)	2.003 (0.038)	2.053 (0.039)	1.986 (0.043)	2.002 (0.025)	2.026 (0.025)	1.869 (0.031)
	50	2.066 (0.092)	2.195 (0.099)	2.228 (0.076)	2.023 (0.043)	2.073 (0.044)	1.980 (0.044)	2.013 (0.029)	2.037 (0.029)	1.868 (0.033)
	100	2.134 (0.108)	2.269 (0.117)	2.206 (0.083)	2.050 (0.050)	2.100 (0.051)	1.974 (0.048)	2.024 (0.035)	2.048 (0.035)	1.871 (0.035)
200	2.285 (0.143)	2.443 (0.156)	2.136 (0.095)	2.105 (0.066)	2.157 (0.068)	1.959 (0.057)	2.049 (0.043)	2.074 (0.043)	1.890 (0.040)	

see "Appendix A"

## References

- Abrevaya J (1997) The equivalence of two estimators of the fixed-effects logit model. *Econ Lett* 55(1):41–43
- Barmby T, Nolan M, Winkelmann R (2001) Contracted workdays and absence. *Manchester Sch* 69(3):269–275
- BfS Bundesamt für Statistik (2016) Arbeitsmarktindikatoren der Schweiz. Neuchâtel, Switzerland
- Chamberlain G (1980) Analysis of covariance with qualitative data. *Rev Econ Stud* 47:225–238
- Dean CB (1992) Testing for overdispersion in Poisson and binomial regression models. *J Am Stat Assoc* 87:451–457
- Gourieroux C, Monfort A, Trognon A (1984) Pseudo maximum likelihood methods: theory. *Econometrica* 52:681–700
- Kemp G, Santos Silva J (2016) Partial effects in fixed-effects models, United Kingdom Stata Users' Group Meetings 2016
- Kunz J, Staub K, Winkelmann R (2018) Predicting fixed effects in panel probit models. York Health, Econometrics and Data Group Working Paper #18/23
- Machado MP (2004) A consistent estimator for the binomial distribution in the presence of incidental parameters: an application to patent data. *J Econom* 119:73–98
- Maddala GS (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge
- Mroz TA (1987) The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55:765–799
- Pagan A, Ullah A (1999) Nonparametric econometrics. Cambridge University Press, Cambridge
- Papke LE, Wooldridge JM (1996) Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *J Appl Econom* 11:619–632
- Papke LE, Wooldridge JM (2008) Panel data methods for fractional response variables with an application to test pass rates. *J Econom* 145:121–133
- Ramalho EA, Ramalho JJS, Coelho LMS (2016) Exponential regression of fractional-response fixed-effects models with an application to firm capital structure. *J Econom Methods* 7(1):20150019
- Steinhauer A (2018) Working moms, childlessness, and female identity. CEPR Discussion Paper Nr. 12929
- Wooldridge JM (2002) Econometric analysis of cross section and panel data. MIT Press, Cambridge

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.