Department of Economics

ISSN number 1441-5429

Discussion number 10/19

Predicting fixed effects in panel probit models*

Johannes S. Kunz¹, Kevin E. Staub², Rainer Winkelmann³

Abstract:

Many applied settings in empirical economics require estimation of a large number of fixed effects, like teacher effects or location effects. In the context of binary response variables, previous studies have been limited to the linear probability model, citing perfect prediction and incidental parameter biases as reasons. We explain why these problems arise and present an appropriate solution for the probit model. In contrast to other estimators, it ensures that predicted fixed effects exist for all units. We illustrate the approach in simulation experiments and an application to health care utilization.

Keywords: Perfect prediction, Incidental parameter bias, Fixed Effects, Panel data, Binary response, Bias reduction

Acknowledgments: We thank Alex Ballantyne, Bob Breunig, Stefan Bruder, Joseph Doyle, Denzil Fiebig, Joe Hirschberg, Andrew Jones, Maarten Lindeboom, Jenny Lye, Janina Nemitz, Dennis Petrie, Carol Propper, Chris Skeels, Jenny Williams, Frank Windmeijer and Tiemen Woutersen for helpful comments on an earlier version of the paper, which was circulated under the title “Estimating fixed effects: Perfect prediction and bias in binary response panel models”. Staub acknowledges funding from the Australian Research Council through grant DE170100644. Stata code for the bias-reduced estimators described in this paper is available from the authors’ websites or through ssc under the name brglm.

¹Monash University

²University of Melbourne, IZA

³University of Zurich

1 Introduction

This paper addresses estimation and prediction of individual-specific fixed effects (FE) in the panel probit model when the individual dimension N is large and the time dimension T is small. The view of fixed effects not as nuisance parameters but rather as features of interest has received increased attention in applied work lately, for example in the context of neighborhood FEs (Chetty and Hendren, 2018), teacher FEs (Chetty, Friedman and Rockoff, 2014), and judge FEs (Abrams, Bertrand and Mullainathan, 2012), to name but a few. All these studies employ the linear model to obtain fixed effects, although the dependent variable is frequently binary.

In such cases, it clearly would be preferable to extract the FEs from a proper (non-linear) binary response model, such as the logit or probit model. However, this approach faces a major obstacle since conventional estimation methods, such as maximum likelihood or non-linear least squares, often fail to yield finite predictions for the FEs. In short panels there are typically many individual units with identical outcomes in all time periods (either all zeros or ones). Maximum likelihood estimation will set the fitted probabilities to either zero or one, which requires the absolute value of the FE to go to infinity. The literature often describes this situation as “perfect prediction” (see Maddala, 1983). In order to emphasize the distinction between prediction of fitted probabilities and those of FEs, the main objective in this paper, we refer to such “perfectly predicted” panel units instead as *concordant*.

The degree of concordance in typical applications can be very high. For instance, Autor, Duggan and Gruber (2014) explain that they avoid using a non-linear binary response model in their panel analysis of U.S. long-term disability policies specifically because 25 per cent of their sample is concordant and would be omitted in estimation. In our application to doctor visits with representative German survey data, between 29 to 45 per cent of the respective samples are concordant. Thus, for a sizeable share of individuals it would be impossible to obtain FE predictions using conventional binary response models, and units would necessarily be dropped from the estimation sample. Such approaches discard informative variation stemming from what is often a rich set of covariates.

The concordance problem has, to the best of our knowledge, not been addressed directly by the previous literature on panel data models for binary responses. Instead, the focus has been on the well-known incidental parameters problem stating that for fixed T and $N \rightarrow \infty$, the FE dummy-variables maximum likelihood estimator is inconsistent for the structural parameters of the model (such as the vector “ β ”, the coefficients on variables; e.g., Abrevaya, 1997). The reason is that in

non-linear models, the non-vanishing estimation error of the fixed effects contaminates the estimator for the regression parameters. Two solution approaches have been proposed, neither addressing the concordance problem nor the quality of the estimated fixed effect. First, a consistent conditional maximum likelihood estimator for β exists (only) for the logit (but not the probit) model, but that is of no help here as it does not provide estimates of the FEs. Second, other approaches abandon the fixed- T assumption in favor of an alternative, where both N and T increase (Hahn and Newey, 2004, Arellano and Hahn, 2006). Just having T increase, however, does not automatically solve the incidental parameters problem, but consistent estimators are available if N does not grow faster than T^3 , and analytical or jackknife bias corrections of the first-order bias in the structural parameter such as β (e.g. Bester and Hansen, 2009, Fernandez-Val, 2009) or functionals of the fixed effects such as average marginal effects (e.g. Fernandez-Val, 2009, Dhaene and Jochmans, 2015) are applied.

Our approach builds on the second approach but focusses instead directly on the estimation of the FEs. We obtain the first-order bias of the maximum likelihood estimator for the FEs, analytically under simplifying assumptions, as well as in simulation experiments, in the context of the panel probit model for binary outcome. In order to remove the first-order, $O(1/T)$ bias, we follow Kosmidis and Firth (2009), who derive a general modified score function for models in the linear exponential family. Their approach can be tailored to the panel probit model. The resulting estimator is referred to as “bias-reduced (BR) probit”. Importantly, we show that this estimator has finite support for all model parameters, including the fixed effects of concordant observation units, by shrinking the FEs towards zero. Shrinkage also reduces the variance of the predictions and unambiguously improves the mean squared error. Relatedly, all observations contribute to estimation of the structural parameters, and the incidental parameters problem is overcome. In addition, it can be easily implemented by using iteratively re-weighted least squares, and it allows for straightforward inference using standard methods.

In the next section, we formally introduce the problems of first-order bias and concordance in the context of binary response fixed effects panel data models. We focus in our discussion on the probit model where not even a conditional maximum likelihood solution exists. We then present the BR estimator, and show how it addresses these problems. In Section 3, we set up simulation experiments for predicting the FEs across a number of differently shaped distributions from which the true FEs are drawn. The BR estimator performs well in these simulations in terms of bias as well as mean squared error. The simulations also indicate that the BR estimator delivers reliable estimates of β in short panels. Overall, there is a clear recommendation for the use of the BR

estimator in applications, in particular when there is a high degree of concordance.

In Section 4, we present an illustrative application related to health care utilisation. Using panel data from the German Socio-Economic Panel for the period 2000–2014, we obtain predictions of the individual-specific fixed effects in a model where the binary variable “any doctor visit during the last three months (yes/no)” is regressed on a number of indicators of socio-economic status and health status. Estimating FEs for separate subperiods, e.g. 2000–2004 and 2010–2014, makes it possible to study their stability over time. We show that the percentile rank of the estimated FEs for the first-period doctor visits are also very predictive for a different outcome, the incidence of hospital admission ten years later. Thus, the individuals unobserved component is not only very persistent over time but also across different domains of health care utilisation. Section 5 concludes.

2 Econometric methods

2.1 Maximum likelihood estimation of the FE probit model

Consider a panel probit model with individual-specific intercepts, or fixed effects, α_i ,

$$\Pr(y_{it} = 1 | \alpha_i, x_{it}) = \Phi(\alpha_i + x'_{it}\beta), \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where $y_{it} \in \{0, 1\}$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, x_{it} is a vector of covariates and β a conformable vector of coefficients. N is assumed to be large and T is small. Observations are independent between individuals and, conditional on fixed effects α_i , serially uncorrelated. Regressors are strictly exogenous: $\Pr(y_{it} = 1 | x_{i1}, \dots, x_{iT}, \alpha_i) = \Pr(y_{it} = 1 | x_{it}, \alpha_i)$. This model is prominent among practitioners, since it does not make any assumption on the distribution of the FEs other than that they are finite, nor does it require the α_i ’s to be exogenous (uncorrelated with x_{it}).

It is well-known that the maximum likelihood (ML) estimator, $(\hat{\alpha}, \hat{\beta}) = (\hat{\alpha}_1, \dots, \hat{\alpha}_N, \hat{\beta})$ has a number of deficiencies in this case. First, $\hat{\beta}$ is inconsistent. This is a consequence of the incidental parameters problem. Abrevaya (1997) shows for the panel logit model with $T = 2$, that $\text{plim} \hat{\beta} = 2\beta$. Greene (2004) provides Monte Carlo simulation results for the probit model showing that the upward bias persists for $T = 8$ and even $T = 20$. Second, $\hat{\alpha}_i$ is inconsistent for fixed T and $N \rightarrow \infty$, and biased for small T . Third, $\hat{\alpha}_i$ does not exist due to concordance, if $\bar{y}_i = 0$ or if $\bar{y}_i = 1$,

where $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$.

We are here mostly concerned with the second and third issues, the small sample bias and the potential non-existence of $\hat{\alpha}_i$. Our approach is based on an estimator developed by Kosmidis and Firth (2009) (see also Firth, 1993) for cross-section data and adapts it to the estimation of fixed effects in a probit panel data model. We show that the resulting estimator does not suffer from the concordance problem. It also is relatively easy to compute, as it can be obtained using an iteratively weighted least squares estimator as shown by Kosmidis and Firth (2009).

2.2 First-order bias

Non-linear maximum likelihood estimators have a finite sample bias. A formal derivation of the first-order bias of ML estimators is given in Cox and Snell (1971). For an illustration, consider a simple panel probit model with time-invariant regressors:

$$\Pr(y_{it} = 1 | \tilde{\alpha}_i, \bar{x}_i) = \Phi(\tilde{\alpha}_i + \bar{x}_i' \gamma) = \Phi(\alpha_i) \quad (2)$$

where $\alpha_i = \tilde{\alpha}_i + \bar{x}_i' \gamma$. In this case, $\bar{y}_i$ is the unbiased ML estimator for $\mu_i = \Phi(\alpha_i)$ and $\hat{\alpha}_i = \Phi^{-1}(\bar{y}_i)$. Expanding $\hat{\alpha}_i$ around α_i gives

$$\hat{\alpha}_i - \alpha_i \approx \frac{\partial \alpha_i(\mu_i)}{\partial \mu_i} (\bar{y}_i - \mu_i) + \frac{1}{2} \frac{\partial^2 \alpha_i(\mu_i)}{(\partial \mu_i)^2} (\bar{y}_i - \mu_i)^2$$

where $\alpha_i(\mu_i) = \Phi^{-1}(\mu_i)$, and

$$\frac{\partial \alpha_i}{\partial \mu_i} = \frac{1}{\partial \Phi(\alpha_i) / \partial \alpha_i} = \frac{1}{\phi(\alpha_i)}, \quad \frac{\partial^2 \alpha_i}{\partial \mu_i^2} = -\frac{1}{\phi(\alpha_i)^2} \times [-\alpha_i \phi(\alpha_i)] \times \frac{\partial \alpha_i}{\partial \mu_i} = \frac{\alpha_i}{\phi(\alpha_i)^2}.$$

Since $E[(\bar{Y}_i - \mu_i)^2] = \mu_i(1 - \mu_i)/T = \Phi(\alpha_i)(1 - \Phi(\alpha_i))/T$, it follows that

$$E(\hat{\alpha}_i - \alpha_i) = \frac{1}{2T} \frac{\alpha_i \Phi(\alpha_i)(1 - \Phi(\alpha_i))}{\phi(\alpha_i)^2} + O(T^{-2}). \quad (3)$$

The first-order bias is positive if $\alpha_i > 0$, and hence $\Phi(\alpha_i) > 0.5$. It is negative for $\alpha_i < 0$. As $|\alpha_i|$ goes to infinity, so does the product of Mills ratios $\Phi(\alpha_i)(1 - \Phi(\alpha_i))/\phi^2(\alpha_i)$ and hence the bias, both absolute and relative. Note that $\alpha_i \rightarrow \infty$ means that $\Pr(y_{it} = 1 | \alpha_i) \rightarrow 1$, while for $\alpha_i \rightarrow -\infty$, $\Pr(y_{it} = 1 | \alpha_i) \rightarrow 0$. Hence, first-order bias and concordance are two sides of the same coin.

2.3 Concordance

Concordance in the general model (1) means that the first-order conditions for the ML estimator do not have a finite solution due to lack of variation of the binary outcome in the panel probit model. The $K + N$ first-order conditions of the panel probit model are:

$$s^{ML}(\alpha_i) = \frac{\partial \log L}{\partial \alpha_i} = \sum_{t=1}^T (y_{it} - \Phi(\eta_{it})) \frac{\phi(\eta_{it})}{\Phi(\eta_{it})(1 - \Phi(\eta_{it}))} = 0, \quad i = 1, \dots, N, \quad (4)$$

$$s^{ML}(\beta_k) = \frac{\partial \log L}{\partial \beta_k} = \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \Phi(\eta_{it})) \frac{\phi(\eta_{it})}{\Phi(\eta_{it})(1 - \Phi(\eta_{it}))} x_{k,it} = 0, \quad k = 1, \dots, K, \quad (5)$$

where $\eta_{it} = \alpha_i + x'_{it}\beta$, and K is the number of regressors in x_{it} . Suppose that $\bar{y}_i = 0$ for some i . Then (4) simplifies to

$$\sum_{t=1}^T \frac{\phi(\eta_{it})}{1 - \Phi(\eta_{it})} = 0, \quad (6)$$

which does not have a solution since the inverse Mills ratio $\lambda_{it} = \phi(\eta_{it})/(1 - \Phi(\eta_{it})) > 0$ for finite η_{it} . Similarly, if $\bar{y}_i = 1$ for some i , (4) simplifies to

$$\sum_{t=1}^T \frac{\phi(\eta_{it})}{\Phi(\eta_{it})} = 0, \quad (7)$$

which does not have a solution either. In the first case, $\hat{\alpha}_i$ will tend to minus infinity, while it will tend to plus infinity in the second. Units i where observations are either all equal to zero or all equal to one are called *concordant*.

Note that the estimator for β still exists. As long as there are some panel units with variation in y_{it} (i.e., some *discordant* units), β can be estimated using those observations, based on (5). Concordant observations do not contribute to the (concentrated) score, since

$$\begin{aligned} \lim_{\hat{\alpha}_i(\beta) \rightarrow -\infty} \left(- \sum_t \frac{\phi(\hat{\alpha}_i(\beta) + x'_{it}\beta)}{1 - \Phi(\hat{\alpha}_i(\beta) + x'_{it}\beta)} x_{it} \right) &= 0 & \text{if } \bar{y}_i = 0, \\ \lim_{\hat{\alpha}_i(\beta) \rightarrow +\infty} \left(- \sum_t \frac{\phi(\hat{\alpha}_i(\beta) + x'_{it}\beta)}{\Phi(\hat{\alpha}_i(\beta) + x'_{it}\beta)} x_{it} \right) &= 0 & \text{if } \bar{y}_i = 1. \end{aligned}$$

The problem of concordance is most severe for short panels: as T increases, it becomes less and less likely to obtain panel units with $\bar{y}_i = 0$ or $\bar{y}_i = 1$, provided that $0 < \Pr(y_{it} = 1) < 1$. For example,

in the simple time-invariant model (2),

$$\Pr(\bar{y}_i = 0) + \Pr(\bar{y}_i = 1) = \Pr\left(\sum_{t=1}^T y_{it} = 0\right) + \Pr\left(\sum_{t=1}^T y_{it} = T\right) = (1 - \Phi(\alpha_i))^T + \Phi(\alpha_i)^T, \quad (8)$$

Hence, the probability of concordance decreases in T . For a given T , it has a minimum at $\alpha_i = 0$. A larger absolute value of α_i leads to both a larger first-order bias and a higher incidence of concordance. Discarding concordant units from an analysis of the distribution of the FEs, α_i , might lead to flawed conclusions.

2.4 A bias-reduced panel probit estimator

Firth (1993) showed that for linear exponential family models with canonical link function, the first-order bias of the maximum likelihood estimator can be removed by maximising a modified log-likelihood function that includes a penalty term based on the log-determinant of the information matrix (see also Ehm, 1991). The probit model is a linear exponential family model with non-canonical link where such a modified likelihood function does not exist. However, Kosmidis and Firth (2009) derived a related adjustment to the score function that achieves the same first-order bias reduction for general cross-sectional models.

Applying and extending this approach to the panel probit models with fixed effects achieves three things: first, it resolves the incidental parameters problem, in the sense that the estimator of the structural parameter is asymptotically normal and centered at the truth, contingent on asymptotics where T grows faster than $N^{1/3}$ (Hahn and Newey, 2004). Sartori (2002) establishes the same sufficient condition for the absence of the incidental parameters problem in the context of general profile likelihood functions, that extend to the Firth correction in exponential families (see also Lunardon, 2018). Second, as the focus is here on estimation of α_i , removing its first order bias has the obvious direct benefit of providing better estimates of these parameters of interest. Third, and equally important, it ensures that finite estimates $\hat{\alpha}_i$ exist even for concordant observation units, which makes it possible to study the whole distribution of FEs rather than a selected subsample.

For the probit panel model, the adjusted score function with respect to α_i is given by

$$\begin{aligned} s^{BR}(\alpha_i) &= \sum_{t=1}^T \left[y_{it} - \Phi(\eta_{it}) - \frac{1}{2} h_{it} \eta_{it} \frac{\Phi(\eta_{it})(1 - \Phi(\eta_{it}))}{\phi(\eta_{it})} \right] \frac{\phi(\eta_{it})}{\Phi(\eta_{it})(1 - \Phi(\eta_{it}))} \\ &= s^{ML}(\alpha_i) - \sum_{t=1}^T \frac{1}{2} h_{it} \eta_{it}, \end{aligned} \quad (9)$$

where h_{it} are the it -th diagonal elements of the $NT \times NT$ projection matrix

$$H = W^{1/2} X (X' W X)^{-1} X' W^{1/2}, \quad (10)$$

with X the $NT \times K$ matrix of the K regressors, and W is the $NT \times NT$ diagonal matrix with typical element $w_{it} = \phi(\eta_{it})^2 / [\Phi(\eta_{it})(1 - \Phi(\eta_{it}))]$. The β_k -terms of the score vector are adjusted accordingly as

$$s^{BR}(\beta) = s^{ML}(\beta) - \sum_{t=1}^T \frac{1}{2} h_{it} \eta_{it} x_{it} \quad (11)$$

From (9) and (11), it can be seen that if we define

$$y_{it}^* = y_{it} - \frac{1}{2} h_{it} \eta_{it} \frac{\Phi(\eta_{it})(1 - \Phi(\eta_{it}))}{\phi(\eta_{it})}, \quad (12)$$

then (9) and (11) are in the form of the standard ML scores $s^{ML}(\alpha_i)$ and $s^{ML}(\beta)$, where y_{it} is replaced by the pseudo-response y_{it}^* .

The BR panel probit estimator $(\hat{\beta}, \{\hat{\alpha}_i\})$ is obtained by jointly solving the $N + K$ first-order conditions $s^{BR}(\alpha_i) = 0$ and $s^{BR}(\beta) = 0$. In practice, this can be done using iteratively re-weighted least squares common for generalized linear models (as in Kosmidis and Firth, 2009), or Newton-Raphson type pseudo-ML estimators for the probit model, where at iteration s , the vector $\{y_{it}^*\}$ is computed from estimated $s - 1$ values $\hat{h}_{it}(\hat{\alpha}^{s-1}, \hat{\beta}^{s-1})$ and $\hat{\eta}_{it}(\hat{\alpha}^{s-1}, \hat{\beta}^{s-1})$.

2.4.1 Obtaining finite FEs for all units

The problem of non-existing ML estimates can also arise in cross-sections. An example is a model for the probability of treatment with a binary instrument Z and no “always takers”, i.e., when $Z = 0$ implies non-treatment. Heinze and Schemper (2002) show that using the logit model with Firth penalty solves this problem.

Here, we are concerned with estimating FEs in a panel probit model. To prove that the BR panel probit estimator is always finite, consider the case where all observations of unit i are equal to one, $\bar{y}_i = 1$. Then, we can write the bias-reduced score (9) as

$$s^{BR}(\alpha_i) = \left(\sum_t \frac{\phi(\eta_{it})}{\Phi(\eta_{it})} \right) - \frac{\alpha_i}{2} \left(\sum_t h_{it} \right) - \frac{1}{2} \left(\sum_t h_{it} x'_{it} \beta \right) = g_1(\alpha_i) - \alpha_i g_2(\alpha_i) - g_3(\alpha_i). \quad (13)$$

When α_i becomes large, the first term in the score, $g_1(\alpha_i)$, approaches zero, because each inverse Mills ratio in the sum approaches zero. Because h_{it} is an element of the diagonal of a projection matrix, we have that $0 < h_{it} \leq 1$ for each h_{it} , so that $g_2(\alpha_i)$ is bounded. Thus, as α_i tends to plus infinity, the second term, $-\alpha_i g_2(\alpha_i)$, tends to minus infinity. The third term, $g_3(\alpha_i)$, tends to some finite constant because it is a sum of T finite summands. Thus, the whole score tends to minus infinity when α_i tends to plus infinity. When α_i tends to minus infinity, $g_1(\alpha_i)$ grows without bound, and so does $-\alpha_i g_2(\alpha_i)$, while $g_3(\alpha_i)$ tends to some other finite constant. Thus, the whole score tends to plus infinity. Since the score is continuous, a finite solution must exist. Similar arguments can be made to show that a solution exists for the other concordance case, $\bar{y}_i = 0$, as well.

Without time varying regressors, i.e. in the constants-only model, $\sum_t h_{it} = 1$. From (9), we see that if $y_{i1} = \dots = y_{iT} = 1$, $\hat{\alpha}_i$ is obtained as solution to the non-linear equation

$$\hat{\alpha}_i = 2T \frac{\phi(\hat{\alpha}_i)}{\Phi(\hat{\alpha}_i)}, \quad (14)$$

whereas for $y_{i1} = \dots = y_{iT} = 0$,

$$\hat{\alpha}_i = -2T \frac{\phi(\hat{\alpha}_i)}{1 - \Phi(\hat{\alpha}_i)}, \quad (15)$$

The two cases differ only in sign. For $T = 2, 4, 8, 12$ the estimates for α_i are about ± 1.06 , ± 1.37 , ± 1.67 , and ± 1.84 , respectively.

When $\bar{y}_i = 0$, the associated predicted probabilities $\Pr(\widehat{y_{it}} = 0) = 1 - \Phi(\hat{\alpha}_i)$ are 0.144, 0.086, 0.048 and 0.033. These values illustrate how the BR estimator shrinks the predicted probabilities away from the perfect-prediction probability of 0. The amount of shrinkage is inversely related to the number of individual observations T . The ML solution, of course, is a probability of exactly zero, which, while unbiased, might be an unreasonable prediction for many applications: it means that an event that has not occurred in two, three or four periods is deemed impossible. Note that in models with time-varying covariates x_{it} , the ML solution remains an exact zero irrespective of the value of the covariates. In contrast, the BR estimator adjusts the predicted probabilities according

to the value of x_{it} allowing distinctions between different underlying propensities.

2.4.2 Mean squared error

After having established the finiteness of the BR estimator for $\hat{\alpha}_i$, we now briefly consider the quality of the estimation. For the simple model without covariates, the mean square error (MSE) of the estimated fixed effects in the BR panel probit model can be written as

$$\begin{aligned} \text{MSE}(\hat{\alpha}_i, \alpha_i) &= E_{\bar{y}_i} (\hat{\alpha}_i - \alpha_i)^2 \\ &= \sum_{k=0}^T \Pr(\bar{y}_i = k/T) (\hat{\alpha}_i(k/T) - \alpha_i)^2 \\ &= \sum_{k=0}^T \frac{T!}{k!(T-k)!} \Phi(\alpha_i)^k (1 - \Phi(\alpha_i))^{T-k} (\hat{\alpha}_i(k/T) - \alpha_i)^2, \end{aligned} \quad (16)$$

The first and last terms of the sum represent concordant observations. Hence, the MSE of the ML estimator $\hat{\alpha}_i^{ML}$ cannot be finite. With first-order bias reduction, the MSE of $\hat{\alpha}_i^{BR}$ is always finite.

— — — — — Figure 1 about here — — — — —

Figure 1 visualizes the relationship between the MSE of $\hat{\alpha}_i$ and the severity of concordance. The graph plots the MSE of $\hat{\alpha}_i$ against the share (or probability) of concordant units for $T=4$ and $T=12$, respectively. For example, if $\alpha_i = 0$, it follows that $\Phi(0) = 0.5$, and for $T = 4$ and $N \rightarrow \infty$, $2 \times 0.5^4 = 12.5\%$ of all observations can be expected to be concordant (see also equation (8)). Other values for α_i necessarily lead to a higher degree of concordance.

In these graphs, the MSE is decomposed into a part attributable to concordant observations (the summands corresponding to $k = 0$ and $k = T$ on the right-hand-side of Equation (16)) and a part attributable to discordant observations (the summands corresponding to $0 < k < T$). The relative contribution of concordant observations to the total MSE first decreases and then increases, as concordance becomes more prevalent. Eventually, the MSE is almost exclusively driven by concordant observations and becomes large. But we know from results in section 2.4.1 that it remains finite. In contrast, the ML estimator has infinite bias and infinite MSE and is hence clearly inferior to BR for the purpose of predicting fixed effects. Introducing covariates to the

model is likely to improve the estimation of the FEs due to the increase in available information. We assess this using simulations with varying distributions of the FEs.

3 Simulation experiments

3.1 Set-up

The focus of this study is the performance of the BR approach for estimating fixed effects in panel probit models with a small to moderate number of time periods and a varying prevalence of concordance. The simulation experiments contrast the behaviors of the ML and BR estimators for different true underlying distributions of the FEs, in terms of bias and mean squared error.

— — — — — Figure 2 about here — — — — —

In our simulations, the time-invariant individual effects α_i are drawn from four alternative distributions: uniform, beta, Gaussian, and Bernoulli, as plotted in Figure 2. The distributions have been rescaled and shifted to make them more comparable. All distributions have a mean of zero, or close to zero, and all, or most, of their probability mass lies within the interval $[-1,1]$. The distributions vary starkly, however, in their shape. The data generating processes correspond to a “random effects” model as the distribution of α_i does not depend on the regressor. This allows us to focus on biases purely related to small samples and the concordance problem, whereas additional dependence on regressors would exacerbate or attenuate those biases.

Below, we report simulation results for $N = 100$ and $T \in \{2, 4, 8, 12\}$. For each of the four distributions from Figure 2, we draw one hundred values of α_i first, and keep them fixed through all Monte Carlo replications. There is a single regressor, x_{it} , which is drawn from a uniform distribution with support $[-1,1]$. Again, this is done once for each T and kept fixed over replications. Finally, the binary dependent variables y_{it} are obtained as

$$y_{it}^{(r)} = \mathbb{1}(\alpha_i + \beta x_{it} + \varepsilon_{it}^{(r)} > 0), \quad i = 1, \dots, 100 \quad t = 1, \dots, T,$$

where $\varepsilon_{it}^{(r)}$ has a standard normal distribution, $\beta = 1$, and $r = 1, \dots, 500$ denotes Monte Carlo replications.

In each of the 500 replications, we keep track of the fraction of concordant, or perfectly predicted,

observations, i.e., the fraction of cross-sectional units for which $\bar{y}_i^{(r)} = 0$ or $\bar{y}_i^{(r)} = 1$. For instance, with $T = 4$ and a uniformly distributed α_i , the average fraction of concordant observations over the 500 replications amounts to 24 per cent. This fraction is somewhat lower for the beta (15 per cent) and Bernoulli (20 per cent) distributions, and higher for the normal distribution (28 per cent). Plots and summary statistics of our results are based on all finite estimates: since the maximum likelihood estimator of α_i does not exist for concordant observations, the effective replication sample size is below 500 in these cases.

3.2 Results

Table 1 lists, for each of the four distributions of α_i , the estimated mean and standard deviations of the $N = 100$ predictions, averaged over 500 replications. These can be benchmarked against the mean and standard deviation of the (once) simulated α_i 's, for instance -0.030 and 0.451 in the case of the Bernoulli data generating process.

— — — — — Table 1 about here — — — — —

For the longer time horizons, $T = 12$ and to a lesser extent also $T = 8$, there is not much of a difference between ML and BR. Both estimators are centered roughly at the true mean with the right variance. For the shorter panels, the BR estimator performs much better than ML, where only finite fixed effects predictions can be used. For example, the mean of the ML predictions has a sign opposite to that of the true mean in 7 out of 8 cases (for $T = 2$ and $T = 4$). Regarding variation, we find that for short panels (in particular for $T = 2$), the BR estimator underestimates the variance of the generated fixed effects somewhat, reflecting the shrinkage property of the BR estimator.

Table 2 presents means and standard deviations of the estimated $\hat{\beta}$ across different distributions of α_i and different number of time periods. The true value is 1. The corresponding entries in the table confirm that the ML estimator for the common parameter suffers from incidental parameters bias. The bias is sizeable regardless of the distribution of α_i , and it amounts to about 110, 40, 15 and 10 per cent for T equal to 2, 4, 8 and 12, respectively. In contrast, BR effectively removes much of the bias in $\hat{\beta}$. Even for $T = 2$, only a bias of about -10 per cent is left. At $T = 4$ the bias falls to between 0.6 per cent (Bernoulli) and 2.3 per cent (normal)

— — — — — Table 2 about here — — — — —

One of the differences between BR and ML estimation is that concordant observations contribute to estimation of β in the former but not in the latter case. This affects the relative precision of the two estimators. For $T = 2$, where the degree of concordance is most severe, the standard deviation of $\hat{\beta}_{BR}$ is much smaller than that of $\hat{\beta}_{ML}$, by a factor varying between 2.5 and almost 4, depending on data generating process. For $T = 12$, the difference is reduced to around 10 per cent.

3.3 Predicting fixed effects

We have seen in Table 1 that the BR estimates $\hat{\alpha}_i$ are approximately centered at the true mean $\bar{\alpha}_i$ for each of the four distributions (normal, Bernoulli, uniform and beta). However, from the point of view of estimating individual α_i 's, this results is not very informative. For example, an upward bias for large α_i 's and a downward bias for small α_i 's might simply offset each other.

Figures 3, 5 and 6 therefore plot the mean prediction, over the $r = 500$ replications, separately for each true value of α_i . All results are based on a cross-section size of $N = 100$. Points on the 45-degree line indicate the absence of bias. Figure 4, showing the results from the Bernoulli DGP, is an exception. Here, α_i can only take two values, and we compare the entire empirical distribution function of $\hat{\alpha}_i$ to the step function implied by the true Bernoulli DGP (with jumps at -.75 and .25, see Figure 2).

Each figure has a 2×2 format, one for the estimation method (ML versus BR) and one for the time dimension ($T = 4$ versus $T = 8$).

— — — — — Figure 3 about here — — — — —

Figure 3 shows results for the normally distributed individual effect. The scale on the y-axis indicates that the true draws of α_i varied from -2.0 to 1.6 in this case. Naturally, the density of observations increases as we move to the center of the distribution. In order to compute the mean of the ML estimates, we had to drop all concordant observations which are more prevalent for α_i 's located in the tail of the distribution. Given this necessary adjustment, and for $T = 4$, we find that the ML means display a substantial bias. The bias is positive for α_i ' below zero, and negative for α_i 's above zero, and increasing, in absolute terms, as α_i 's move towards the tails. Doubling the sample size from $T = 4$ to $T = 8$ improves the ML estimates, although bias remains in the tails.

The BR means are computed on the full sample, as they include values for concordant observations. They are close to the 45-degree line, even for the small panel, although some bias emerges for α_i 's more than one standard deviation above or below its mean. Here, the shrinkage towards zero starts having a noticeable effect, with an upward bias being observed for large negative α_i 's and a downward bias for large positive α_i 's. The pattern is similar when the sample size is doubled, although the magnitude of the biases in the tails decreases. The fit in the center of the distribution is even tighter.

— — — — — Figure 4 about here — — — — —

A comparison of the empirical distribution functions in Figure 4 shows a good recovery of the binary true distribution function through the BR estimator, with two separate clusters of estimates closely centered around the true values -0.75 and 0.25; The fit is already quite good for $T = 4$ and further improves for $T = 8$. By contrast, the ML estimates in the small sample are quasi-continuously spread over the range from -1 to 1; the binary nature of the true α_i 's only emerges for the large sample.

— — — — — Figure 5 about here — — — — —

Finally, Figures 5 and 6 are based on bounded but continuous distributions for the individual effect, a uniform(-1,1) distribution in case of figure 5 and a scaled beta distribution in case of Figure 6, the latter with values between -0.5 and 0.7; In the context of the probit model, this means that the probability of a success is bounded, as well, for the average x_{it} between $\Phi(-0.5) = 30.9\%$ and $\Phi(0.7) = 75.8\%$ for the beta DGP. As a consequence, the beta simulations produce less concordance (overall 15 per cent) compared to the normal simulations (28 per cent), and avoiding the tails improves the fit substantially. Although to a lesser degree, this is true also for the uniform DGP. The main message from these simulation results is that the BR method gives approximately unbiased estimates of the fixed effects regardless of sample size, whereas the ML method is unreliable for $T = 4$, while it works satisfactorily for the larger sample, where estimated means are mostly lined up along the 45 degree line.

— — — — — Figure 6 about here — — — — —

4 Application to the determinants of doctor visits

To illustrate the benefits of predicting fixed effects for the complete distribution of individuals, we use data for a 2000–2014 subsample of the Socio-Economic Panel, a large representative household panel survey for Germany (SOEP, see Wagner, Frick and Schupp 2007). The dependent variable is an indicator variable, stating whether a visit to a physician took place ($anyvisit = 1$), or did not take place ($anyvisit = 0$), during the three-months period prior to the annual interview. We apply the BR estimator to obtain predictions of the fixed effects in the panel probit model. Our results relate to the previous literature on the demand for health services based on the number of doctor visits, a count variable (see, e.g., Cameron and Trivedi, 1986; Winkelmann, 2004). Specifically, we zoom in on the extensive margin decision, and correspondingly on the first step of a possible hurdle count data model (Mullahy, 1986; Pohlmeier and Ulrich, 1995; Kunz and Winkelmann, 2017).

We focus on two advantages of our approach that are likely to be important to the study of persistence in health care utilisation. First, we assess whether fixed effects are in fact “fixed”, i.e., capture *time-invariant* individual heterogeneity, such as genetic makeup or childhood experiences. One way to address this issue is by allowing the fixed effects to be time varying, and then studying their behavior over time. There is no unique way of doing this. With 15 years of data, we can split the overall period into three five-year subperiods, 2000–2004, 2005–2009, and 2010–2014, and estimate separate fixed effects for each period. Second, we assess the predictive power of these doctoral visit fixed effects onto other domains of future health care use, such as the propensity of future hospitalization. In this case, the idea is that FEs obtained from current outcomes capture long-term health capital, which can be assessed by considering their correlation with future health outcomes.

Table 3 provides selected summary statistics, separately for men and women and the three time periods. We only include individuals who are aged between 20–65 at the time of interview, drop observations with missing values on any of the variables, and retain a balanced panel of 55,230 person-year observations, representing 1,997 women and 1,685 men. The use of a balanced sample is not necessary but it simplifies the analysis of changes in the distribution of fixed effects. As a consequence of the way the sample is constructed, the average age increases by exactly five years for each five year sub-period. Hence, it is not surprising that the share of women and men with at least one visit increases over time, and self-assessed health (SAH) worsens. SAH is measured on a five-point scale, where the best outcome [1] means “Very good” and the worst outcome [5] means “Bad” (the intermediate outcomes are “Good”, “Satisfactory”, and “Poor”, respectively).

The main difference between males and females is the latter’s overall higher prevalence of doctor visits.

— — — — — Table 3 about here — — — — —

The last row of the sub-panels of Table 3 states the proportion of concordant outcomes by gender and period. This proportion varies from a minimum of 29 percent to a maximum of 46 percent. It tends to increase over time for both men and women, which is due to the fact that overall utilization increases, making it more likely that individuals have outcome “one” (i.e., at least one visit).

4.1 Estimation results

A total of three models were estimated, separately by gender, as displayed in Table 4. First, a pooled cross-sectional probit that does not include individual fixed effects (Columns 1 and 4). Second, a fixed effect ML probit (2,5), and, third, the fixed effects BR probit model (3,6). Both ML and the BR model estimate three sets of individual fixed effects: three indicator variables are included for each individual, one for each five year period.

One important takeaway is that the fixed effects ML probit coefficients, based on a model with dummies for each person without further adjustment, tend to be biased upward: they are always at least as large, in absolute values, as the BR probit estimates, and often substantially larger. This reflects the incidental parameters bias for a set-up where each of the fixed effects needs to be estimated based on $T = 5$ observations only. Due to the substantial proportion of concordant observations, the effective ML sample size is reduced by 33% for men, and by 42% for women. The reduced sample size leads to estimated standard errors that are correspondingly higher for the ML probit model relative to the BR probit model.

— — — — — Table 4 about here — — — — —

In terms of substantive results (based on the BR probit model), better self-assessed health reduced the probability of any doctor visit for both men and women. We calculate a simple approximate upper bound for the effect size: for a man at the margin, i.e., with a 50/50 chance of seeing a doctor, a one-point worsening in SAH increases the probability of seeing a doctor by about $0.36 \times 0.4 = 14.4$ percentage points. Married women are more likely to see a doctor than non-married ones, and having more education reduces the probability only for men.

4.2 Analysing fixed effects

The BR probit model directly provides predictions for the individual-specific fixed effects. These are available for further analyses, e.g. for ranking of individuals by their underlying propensity of experiencing the event. Figure 7 depicts a few examples of the distributions of fixed effects. Panel A compares the fixed effects from the first period (2000-2004) between males and females. Panel B compares the fixed effects of males over time (2000-2004 vs. 2010-2014) and C, analogously, the ones of females.

— — — — — Figure 7 about here — — — — —

The shapes of the fixed effects distributions differ remarkably. For males, the time constant propensity net of covariates has a larger mean, is more variable, and appears to be unimodal. In contrast, the females' fixed effects distribution is shifted to the left and exhibits three modal points, which could indicate a finite mixture type of distribution. This shows the benefits of the fixed effects approach, which does not restrict the shape of the unobserved heterogeneity, relative to, for example, random effects models with parametric distributional assumptions.

Predictions can also be used to assess their stability over time (Panels B, C in Figure 7). For each person, e.g. for $N = 1,685$ males (1,997 females), we obtain $T = 3$ distinct predictions. The total variance for the $NT = 5,055$ (5,991) fixed effects is 0.601 (0.954). The individual-specific component is fairly stable over time: variation within individuals contributes not more than 36 (26) percent to the overall variance.

The bivariate scatter plots in Figure 8 provide another perspective on stability over time. We find a substantial positive correlation in estimated fixed effects for the same person between adjacent sample periods (black dots and linear fit). Yet, the fit is far from the 45 degree line, which would imply complete stability. When comparing the first to the last period, ten years later (lightly-colored dots and linear fit), the correlation is again lower (indicated by the flatter linear fit), although only slightly so. For women, the five-year fixed effects change somewhat more over the longer time frame, implying a slightly lower stability over time.

— — — — — Figure 8 about here — — — — —

In Figure 9, we assess the predictive power of the first period fixed effects for another measure

of health care utilisation: The ten-year-ahead probability to visit any hospital (in the last three months before the annual survey elicitation). To assess this we group the first period fixed effects (2000-2004) into ten percentile bins and plot them against the unconditional mean hospital visit probability in the last period of our sample (2010-2014). The percentile ranks of the time-constant doctor visit fixed effects are clearly predictive of the overall probability to visit any hospital ten years later, and this predictive ability is stronger for females than for males. Taken together, these results show the persistence in unobserved heterogeneity not only over a long period of an individual's life cycle, but also the spillover into other (more costly) domains of health care use, findings which might need to be better accounted for in general models of health care demand.

— — — — — Figure 9 about here — — — — —

Finally, we use predictions to explore any time-invariant relationships between the observed and unobserved components of the demand for health services. For men, the correlation between $\hat{\alpha}_i$ and $\bar{x}_i'\hat{\beta}$ (where $\bar{x}_i$ are the average characteristics) equals -0.403 , -0.414 , and -0.409 in the first, second, and third period, respectively; Hence, individual differences in observed factors tend to be associated with unobservables that move in the opposite direction. Ignoring this correlation (such as in the pooled probit model) would understate the importance of either of the two. Interestingly, the correlation among females, in contrast, is almost non-existent (0.018 , 0.031 , and 0.062 , respectively), implying a potentially less problematic assessment of the cross-sectional variation.

5 Conclusions

This paper discusses the problem of estimating a panel probit model with individual-specific constants or fixed effects. A useful estimator should address the incidental parameters problem, provide finite estimates for all fixed effects, and ideally estimate them with small bias. We argue that a specific first-order bias reduction method, based on a modified joint score function of the structural parameters and the fixed effects as in Kosmidis and Firth (2009), solves all three of the above issues.

An extended simulation study confirms that the theoretical properties of this estimator materialize already in quite small samples (e.g. $T = 4$ and $N = 100$). The incidental parameter bias disappears, and the estimated fixed effects are nearly unbiased for the true heterogeneity parameters. While other first-order bias correction methods have been proposed to overcome the incidental parameters

problem (e.g. Hahn and Newey, 2004, or Bester and Hansen, 2009), these do not remove the concordance problem that is so common in applications, especially in short panels. In the data of our application—balanced panels covering five-year periods—about 40 percent of the observations were concordant and would have led to infinite estimates for the corresponding fixed effects had we used the conventional ML estimators or other bias-corrected estimators.

One could argue that the technical problem of binary response models leading to infinite estimates of α_i is not necessarily a problem; that some α_i might actually be equal to positive or negative infinity. However, for many applications, having infinite FEs is clearly unattractive. From a theoretical point of view, an infinite α_i implies that no unobserved or observed shock could possibly produce a change in the outcome, a very strong assumption. In our application, this would imply a person that has not visited a doctor in five years will under no circumstance ever visit a doctor. From a practical point of view, discarding the information contained in time-varying covariates seems inefficient and leads to a clumping of predictions with the same values. Moreover, as seen in the simulations and application, the exclusion of a potentially large share of the data from the estimation is also likely to lead to larger standard errors for the parameter β .

Of course one can focus on predicted probabilities instead of directly on the FEs. In that case, the infinite FEs of concordant units translate to fitted probabilities of zero and one. Obviously, this does not solve the problem of a clump in the predictions stemming from ignoring the information in the covariates. In addition, from an empirical point of view, making extreme predictions that assign probability of one and zero based on a small number of observations might often be undesirable. Finally, obtaining predicted probabilities of zero and one can be problematic when used in further analyses that require probabilities to be strictly bounded away from these values; examples include using the predictions for inverse probability weighting (such as propensity score and marginal treatment effects methods) or as inputs in some structural dynamic discrete choice models. This last problem also affects linear probability models estimated by OLS. While they obtain finite estimates for the fixed effects of concordant units, linear probability models do not restrict the range of predictions to the unit interval, and concordant units frequently result in predictions greater than one or smaller than zero. The advocated BR probit does not suffer from any of these problems. It produces consistent and meaningful fitted probabilities which lie strictly within the $(0,1)$ interval.

Because of its advantages over other approaches across a number of domains, coupled with the simplicity of implementing it, we believe that the fixed effect BR probit estimator has the essential

features of a new workhorse for the estimation of panel probit models. We focused in our discussion on the probit model as it is a common choice in empirical work, but the advocated approach is applicable to a number of other binary response models as well, such as logit or cloglog. More broadly, the estimator can be extended to other nonlinear fixed effects panel models which suffer from perfect prediction, such as models for ordered and count data.

References

Abrams, David S., Marianne Bertrand and Sendhil Mullainathan. 2012. Do Judges Vary in Their Treatment of Race? *Journal of Legal Studies* 41(2):347-383.

Abrevaya, Jason. 1997. The equivalence of two estimators of the fixed-effects logit model. *Economics Letters* 55(1):41-43.

Arellano, Manuel and Jinyong Hahn. 2016. A likelihood-Based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects. *Global Economic Review* 45(3):251-274.

Autor, David, Mark Duggan, and Jonathan Gruber. 2014. Moral hazard and claims deterrence in private disability insurance. *American Economic Journal: Applied Economics*, 6(4):110-41.

Bester, Alan C. and Christian Hansen. 2009. A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics* 27(2):131-148.

Cameron, A Colin and Pravin K Trivedi. 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1:29-53.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2014. Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review* 104(9):2633-2679.

Chetty, Raj, and Nathaniel Hendren. 2018. The Impact of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. *Quarterly Journal of Economics* 113 (3), 1163-1228.

Cox, David R. and Emily J. Snell. 1971. On test statistics calculated from residuals. *Biometrika* 58(3):589-594.

Dhaene, Geert and Koen Jochmans. 2015. Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82(3):991-1030.

Ehm, Werner. 1991. Statistical problems with many parameters: Critical quantities for approximate normality and posterior density based inference. *Habilitationsschrift*, University of Heidelberg.

Fernandez-Val, Ivan. 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71-85.

Firth, David. 1993. Bias reduction of maximum likelihood estimates. *Biometrika* 80(1):27-38.

Greene, William H. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal* 7(1):98-119.

Hahn, Jinyong and Whitney Newey. 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72(4):1295-1319.

Heckman, James J. and Bo E. Honore. 1990. The empirical content of the Roy model. *Econometrica* 58(5):1121-1149.

Heinze, Georg and Michael Schemper. 2002. A solution to the problem of separation in logistic regression. *Statistics in Medicine* 21(16):2409-2419.

Kosmidis, Ioannis and David Firth. 2009. Bias reduction in exponential family nonlinear models. *Biometrika* 96(4):793-804.

Kunz, Johannes S. and Rainer Winkelmann. 2017. An Econometric Model of Healthcare Demand With Nonlinear Pricing. *Health Economics* 26(6):691-702

Lunardon, N. 2018. On bias reduction and incidental parameters. *Biometrika* 105:233-238.

Maddala, Gangadharrao S. 1983. Qualitative and limited dependent variable models in econometrics. Cambridge: Cambridge University Press.

McCullagh, Peter. and John A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. London, UK: Chapman and Hall/CRC.

Mullahy, John (1986) Specification and testing in some modified count data models, *Journal of Econometrics* 33:341-365.

- Pohlmeier, Winfried and Volker Ulrich. 1995. An Econometric Model of the Two-Part Decision-making Process in the Demand for Health Care. *Journal of Human Resources* 30(2):339-361.
- Sartori, N. 2003. Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* 90:533-49.
- Wagner, G G, J R Frick and J Schupp. 2007. The German Socio-Economic Panel Study (SOEP): Scope, evolution and enhancements. *Schmollers Jahrbuch* 127:139169.
- Winkelmann, Rainer. 2004. Co-payments for prescription drugs and the demand for doctor visits: Evidence from a natural experiment. *Health Economics* 13:1081-1089.

Figures and Tables

Figure 1: SHARE OF CONCORDANT UNITS AND MSE IN BR PROBIT MODELS WITHOUT COVARIATES

Notes: The graphs in the figure plot $\text{MSE}(\hat{\alpha}_i)$ against the share of concordant units, $P(\bar{y}_i = 0) + P(\bar{y}_i = 1)$, in panel probit models without covariates, $P(y_{it} = 1) = \Phi(\alpha_i)$. The decomposition of MSE into contributions from concordant and discordant observations refer to Eq. (16), where the concordant contribution is equal to the sum of the first ($t = 0$) and last ($t = T$) addends of the right-hand-side of Eq. (16), and the discordant contribution is equal to the remainder.

Figure 2: DISTRIBUTIONS OF α_i

Notes: Distributions from which α_i were drawn for the Monte Carlo simulation: “uniform” corresponds to a uniform distribution on the interval $[-1,1]$; “beta”, to a Beta distribution with shape parameters 2 and 5, rescaled to the interval $[-1,1]$ by multiplying the variable by 2 and subtracting 0.5; “bernoulli”, to a modified Bernoulli distribution taking the value -0.75 with probability 0.25, and the value 0.25 with probability 0.75; and “normal”, to a Normal distribution with mean 0 and variance 0.5.

(A) $T = 4$

(B) $T = 8$

Figure 3: MEAN OF ESTIMATED FIXED EFFECTS BY TRUE α_i ($\alpha_i \sim normal$)

Notes: Graphs show average estimates of $\alpha_1, \dots, \alpha_{100}$ over 500 replications against their true values.

(A) $T = 4$

(B) $T = 8$

Figure 4: ESTIMATED VERSUS TRUE DISTRIBUTIONS OF α_i ($\alpha_i \sim \text{Bernoulli}$)

Notes: Graphs show the empirical cdf of the hundred true α_i against the empirical cdf of the hundred average $\hat{\alpha}_i$ estimated over 500 replications.

(A) $T = 4$

(B) $T = 8$

Figure 5: MEAN OF ESTIMATED FIXED EFFECTS BY TRUE α_i ($\alpha_i \sim \text{uniform}$)

Notes: Graphs show average estimates of $\alpha_1, \dots, \alpha_{100}$ over 500 replications against their true values.

(A) $T = 4$

(B) $T = 8$

Figure 6: MEAN OF ESTIMATED FIXED EFFECTS BY TRUE α_i ($\alpha_i \sim \text{beta}$)

Notes: Graphs show average estimates of $\alpha_1, \dots, \alpha_{100}$ over 500 replications against their true values.

Figure 7: HISTOGRAMS OF PREDICTED FIXED EFFECTS

Notes: Predicted fixed effects based on separate regressions for male and females presented in Table 4. Regressions are pooled across years and include 3 sets of individual $\times$ period fixed effects: period 1 (2000–2004), period 2 (2005–2009), period 3 (2010–2014). The left panel compares the distributions of fixed effects in the first period of males and females; the middle panel, the distributions of males in period 1 and period 3; the right panel, those of females.

Source: SOEP v33, 2000–2014, own calculations.

Figure 8: SCATTERPLOTS OF PERIOD-SPECIFIC FIXED EFFECTS

Notes: Plot depicts correlation of the fixed effects based on BR Probit results Table 4, from first period fixed effect to second (blue) and third period (red). Lines denote the linear fit. Left panel males, right panel females.

Source: SOEP v33, 2000–2014, own calculations.

Figure 9: PREDICTING FUTURE HOSPITAL VISITS

Notes: Figure plots the average probability to visit a hospital (in the last 3 month) in 2010-2014 against bins of percentile rank in the individual unobserved time constant component of doctor visits (fixed effects) ten years earlier (2000-2004), based on the BR Probit results Table 4. Left panel males, right panel females, dashed lines depict the linear fit.

Source: SOEP v33, 2000–2014, own calculations.

Table 1: MC SIMULATION: ESTIMATES OF $E(\alpha_i)$ [MEAN] AND $SD(\alpha_i)$ [SD]; $N = 100, 500$ REPLICATIONS

	$T = 2$		$T = 4$		$T = 8$		$T = 12$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\alpha_i \sim \text{Bernoulli}$								
True	-0.030	0.451						
ML	0.025	0.778	-0.030	0.361	-0.019	0.452	-0.032	0.482
BR	-0.018	0.353	-0.023	0.420	-0.031	0.446	-0.034	0.453
$\alpha_i \sim \text{Uniform}$								
True	-0.045	0.585						
ML	0.025	0.819	-0.048	0.445	-0.047	0.573	-0.053	0.620
BR	-0.044	0.431	-0.041	0.540	-0.047	0.576	-0.050	0.588
$\alpha_i \sim \text{Beta}$								
True	0.034	0.296						
ML	-0.147	0.836	-0.014	0.317	0.025	0.305	0.034	0.318
BR	0.007	0.229	0.027	0.290	0.032	0.295	0.033	0.296
$\alpha_i \sim \text{Normal}$								
True	0.045	0.733						
ML	-0.138	0.831	-0.000	0.473	0.028	0.638	0.047	0.710
BR	0.009	0.502	0.038	0.629	0.040	0.696	0.043	0.715

Notes: Rows labelled “True” contain the (true) mean and standard deviation of the 100 drawn α_i for each of the four distributions (Bernoulli, uniform, beta, and normal). Cells in rows ML and BR contain the average, over 500 replications, of the mean and standard deviation of the estimated α_i for each of the two estimators.

Table 2: MC SIMULATION: MEAN AND STANDARD DEVIATION [SD] OF $\hat{\beta}$ ($\beta = 1$, 500 REPLICATIONS)

	$T = 2$		$T = 4$		$T = 8$		$T = 12$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\alpha_i \sim \text{Bernoulli}$								
ML	2.105	0.673	1.400	0.256	1.154	0.122	1.092	0.089
BR	0.953	0.240	1.006	0.169	1.007	0.103	1.002	0.080
$\alpha_i \sim \text{Uniform}$								
ML	2.206	0.747	1.427	0.272	1.163	0.122	1.098	0.091
BR	0.928	0.242	0.997	0.173	1.005	0.102	1.004	0.082
$\alpha_i \sim \text{Beta}$								
ML	2.075	0.716	1.364	0.231	1.143	0.125	1.084	0.086
BR	0.942	0.268	1.013	0.159	1.004	0.107	0.999	0.078
$\alpha_i \sim \text{Normal}$								
ML	2.195	0.990	1.410	0.263	1.163	0.126	1.103	0.090
BR	0.889	0.250	0.977	0.165	0.997	0.105	1.001	0.080

Notes: Cells contain the average and standard deviation, over 500 replications, of the estimated β for each of the two estimators, ML and BR. The true value of β is 1.

Table 3: SELECTED DESCRIPTIVE STATISTICS, BY PERIOD

	2000-2004	2005-2009	2010-2014	All periods
	(1)	(2)	(3)	(4)
<i>Men</i>				
Any doctor, last 3 month (Yes/No)	0.559	0.595	0.638	0.597
Any hospital, last 3 month (Yes/No)	0.073	0.079	0.102	0.085
Age/10	3.954	4.454	4.954	4.454
Self-assessed health	2.385	2.555	2.671	2.537
Disability (Yes/No)	0.055	0.085	0.127	0.089
Years of education	12.478	12.567	12.595	12.546
Share of concordant obs.	0.291	0.312	0.400	0.335
<i>Women</i>				
Any doctor, last 3 month (Yes/No)	0.722	0.729	0.742	0.731
Any hospital, last 3 month (Yes/No)	0.122	0.101	0.103	0.109
Age/10	3.904	4.404	4.904	4.404
Self-assessed health	2.437	2.560	2.675	2.557
Disability (Yes/No)	0.044	0.067	0.104	0.072
Years of education	12.359	12.457	12.485	12.434
Share of concordant obs.	0.389	0.418	0.456	0.421

Source: SOEP v33, 2000–2014, own calculations.

Table 4: ML PROBIT AND BR PROBIT RESULTS, ALL PERIODS

	<i>Men</i>			<i>Women</i>		
	Pooled	<i>Fixed Effects Probit</i>		Pooled	<i>Fixed Effects Probit</i>	
	Probit	ML	BR	ML	Probit	BR
	(1)	(2)	(3)	(4)	(5)	(6)
Self-assessed health	0.40 (0.01)	0.49 (0.02)	0.36 (0.02)	0.36 (0.01)	0.45 (0.02)	0.31 (0.01)
Disability (Yes/No)	0.68 (0.04)	0.26 (0.14)	0.15 (0.09)	0.79 (0.05)	0.61 (0.17)	0.28 (0.08)
Married (Yes/No)	0.05 (0.02)	0.13 (0.08)	0.10 (0.06)	0.02 (0.02)	0.14 (0.08)	0.10 (0.05)
Age/10	-0.53 (0.08)	-0.94 (0.40)	-0.67 (0.31)	-0.65 (0.08)	-0.89 (0.38)	-0.62 (0.26)
Age ² /100	0.06 (0.01)	0.13 (0.04)	0.09 (0.03)	0.07 (0.01)	0.08 (0.04)	0.06 (0.03)
Years of education	0.02 (0.00)	-0.24 (0.07)	-0.17 (0.05)	0.01 (0.00)	-0.01 (0.05)	-0.01 (0.04)
Fulltime work (Yes/No)	0.03 (0.04)	0.03 (0.09)	0.03 (0.06)	0.09 (0.02)	0.16 (0.06)	0.12 (0.04)
Parttime work (Yes/No)	0.09 (0.04)	0.14 (0.09)	0.10 (0.06)	0.06 (0.02)	0.06 (0.05)	0.04 (0.03)
Log household income	0.06 (0.02)	-0.07 (0.05)	-0.05 (0.04)	0.05 (0.02)	-0.01 (0.05)	-0.01 (0.03)
Nr. of observations	25,275	16,820	25,275	29,955	17,345	29,955
Nr. of individuals	1,685	1,194	1,685	1,997	1,221	1,997
Individual×5-year-period fixed effects		✓	✓		✓	✓

Notes: Table contains main estimation results (coefficients and standard errors), regressing any doctor visit in last 3 month, separately for males (Columns 1-3) and females (Columns 4-6) using pooled ML probit (1,4), ML fixed effect probit (2,5) including three individual indicator variables for individual×period (2000-2004, 2005-2009, 2010-2014), and the BR Probit fixed effects estimator (3,6).

Source: SOEP v33, 2000–2014, own calculations.