Correcting the bias in the estimation of a dynamic ordered probit with fixed effects of self-assessed health status.*

Jesus M. Carro† Alejandra Traferri‡

April 2010

Abstract

This paper considers the estimation of a dynamic ordered probit with fixed effects, with an application to self-assessed health status. The estimation of nonlinear panel data models with fixed effects by MLE is known to be biased when T is not very large. The problem is specially severe in our model because of the dynamics and because it contains two fixed effects: one in the linear index equation, interpreted as unobserved health status, and another one in the cut points, interpreted as heterogeneity in reporting behavior. The contributions of this paper are twofold. Firstly this paper contributes to the recent literature on bias correction in nonlinear panel data models by applying and studying the finite sample properties of two of the existing proposals to the ordered probit case. The most direct and easily applicable correction to our model is not the best one and still has important biases in our sample sizes. Secondly, we contribute to the literature that studies the determinants of Self-Assessed Health measures by applying the previous analysis on estimation methods to the British Household Panel Survey.

Keywords: dynamic ordered probit, self-assessed health, reporting bias, panel data, unobserved heterogeneity, incidental parameters, bias correction.

JEL classification: C23, C25, I19

---

*We thank Raquel Carrasco, Matilde Machado and seminar participants at Universidad Carlos III and Boston University for useful comments and suggestions. The first author gratefully acknowledges that this research was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme, and by grant number SEJ2006-05710/ECON from the Spanish Minister of Education. The second author would like to thank the Institute for Economic Development at Boston University, in which she conducted part of this research as a visiting scholar.

†Department of Economics, Universidad Carlos III de Madrid. E-mail: jcarro@eco.uc3m.es

‡Department of Economics, Universidad Carlos III de Madrid. E-mail: atraferr@eco.uc3m.es

1 Introduction

The estimation of nonlinear panel data models with fixed effects is known to be problematic with the panels usually available, since they do not have a very large number of periods. This is even more severe when estimating dynamic models, like the dynamic ordered probit model. This incidental parameters problem is reflected in the inconsistency of standard estimators like the maximum likelihood estimator (MLE) when the number of individuals N goes to infinity and T is fixed. Even when T goes to infinity, if it does at a smaller or the same rate as N, the asymptotic normal distribution is not centered at zero due to the bias coming from the incidental parameters. Moreover, this problem results in large finite sample biases of the MLE when using panels where T is not very large. The dynamic ordered probit model is not an exception to this, specially if it contains more than one individual specific parameter, as in our case.

An important part of the research on microeconometrics in recent years has been concerned with finding a solution to this problem, by developing bias-adjusted methods to estimate those models. Given this fast growing literature, there are several bias correction methods we could consider to estimate our model. These methods can be grouped in three approaches.[1] The first one is to construct an analytical or numerical bias correction of a fixed effect estimator. Hahn and Newey (2004), Hahn and Kuersteiner (2004) and Fernandez-Val (2009), for example, take this approach to the problem. The second approach is to correct the bias in moment equations. An example of this is Carro (2007), which uses an estimator of this type to correct the bias in dynamic binary choice models. The third approach is to correct the objective function. Arellano and Hahn (2006) and Bester and Hansen (2009) take this approach, with the latter including an application to a dynamic ordered probit model.

Asymptotically all of the above methods reduce the order of the bias of the MLE from the standard $O(T^{-1})$ to $O(T^{-2})$. Therefore, from this perspective we could use any of the methods developed for dynamic models. A second criteria to choose among the several alternatives is to check the easiness of implementation to our model. From this criteria the estimator that corrects the objective function using a penalty term based on a product of the sample scores and Hessian can be directly applied without modification to our specific model. Bester and Hansen (2009) refer to it as the HS penalty. In contrast with the direct applicability of this estimator, others are computationally more difficult and require some transformation to be

[1] See Arellano and Hahn (2007) for a good review of this literature, detailed references and a general framework in which the various approaches can be included.

applied to our model, specially because our model contains two fixed effects instead of one as usually is the case in binary choice models. This does not mean that other methods cannot be applied nor that we do not know their theoretical properties. They have been developed for a quite general class of nonlinear panel data models with fixed effects.

Notwithstanding, a third and more important criteria is the finite sample performance of the method when estimating our model with the sample size we have. The incidental parameters problem can be seen as a finite sample bias problem in panel data context. The incidental parameters problem is not very important when T is large. However, since our panel does not have a very large number of periods it is reasonable to wonder whether the good asymptotic properties when T goes to infinity are a good approximation to our finite sample. Given this, we should evaluate the finite sample performance of the available methods we could use to estimate our model. As usual, this comparison is done through Monte Carlo experiments. Bester and Hansen (2009) do not compare the finite sample properties of the method they use with others for the ordered probit case because many of the other methods will require some derivation to get the specific correction for this case. They, however, make such a comparison using a static and a dynamic logit model. Also, Carro (2006) and Fernandez-Val (2009) make Monte Carlo experiments for logit and probit models with different sample sizes. The Monte Carlo experiments made in these three papers allow us to compare a wide range of methods for the dynamic logit and probit models. From all these comparisons we can conclude that the HS penalty approach is clearly not the best one. We can also conclude that for sample sizes with T smaller than 13 the reminding bias when using HS could still be significant, specially for the ordered probit Bester and Hansen (2009) simulate. This result is also confirmed in our simulations. Given this and that our empirical application has $T = 13$, some other of the proposed methods should be considered, in addition to the HS penalty approach. Interesting candidates are the corrections discussed by Fernandez-Val (2009) and Carro (2006) since they are both equally superior to other methods in the relevant existing monte Carlo experiments. In this paper we derive explicit formulas of the modified MLE used in Carro (2007) for the model considered here, evaluate its finite sample performance and compare it with the HS penalty estimator. This exercise is a main contribution of this paper since, as Arellano and Hahn (2007) point out in their conclusions, more research is needed to know "how well each of the methods recently proposed work for other specific models and data set of interest in applied econometrics." Also, Greene and Henshen (2008) comment on the lack of studies about the applicability to ordered choice models of the recent proposals for bias reduction estimators in binary choice models.

Self-assessed health (SAH) has been used as a proxy for true overall individual health status in many socioeconomic studies. Also, it has been shown to be a good predictor of mortality and of subsequent demand of medical care. Motivated by this importance and the high observed persistence in health outcomes, Contoyannis, Jones and Rice (2004) study the dynamics and effects of socioeconomic variables on SAH for the British Household Panel Survey. Among other aims, they try to know the relative contribution of state dependence and unobserved heterogeneity in explaining the observed persistence in SAH. Given that SAH is a categorical variable this is a case where the use of a dynamic order probit model is appropriate.

In addition to accounting for unobserved factors that affect health status (index shift), here we also have to take into account the possible heterogeneity in reporting behavior (cut-point shift). The cut-point shifts occur if individuals use different thresholds when assessing their health and reporting it in the SAH categorical variable, so that they report a different value of SAH even though having the same level of true health.[2] To control for these two unobserved factors, which are possibly correlated with other explanatory variables and between each other, we include individual effects not only in the levels of the order probit but also in the cut points. As it happens with one individual effect, we could take a 'random effects' approach. However, this approach has the drawback of imposing either independence, or a specific and restrictive functional form for the relation between the unobserved heterogeneity and other explanatory variables. It also has the drawback of having to deal with the so-called initial conditions problem. Taking a 'fixed effects' approach we leave unrestricted (i.e. nonparametric) the joint distribution of the two kind of individual effects and their correlation with the explanatory variables. Moreover, there is not initial conditions problem. Despite these advantages, there have been only few applications in health economics of nonlinear panel models with fixed effects, as can be seen by reading Jones' (2007) handbook chapter. This is due to the difficulty of solving the incidental parameters problem addressed by this paper and the related literature.

The rest of the paper proceeds as follows. We first present our model and its estimation problems. We comment on the possible solutions from the nonlinear bias correction literature for nonlinear panel data models with fixed effects. We use simulations to evaluate the finite sample performance of two of the alternatives

[2]See Lindeboom and van Doorslaer (2004) for a test about the existence of these two different kinds of shifts.

and use this as final criteria for choosing our estimator. In Section 3, we apply all that to the study self-assessed health status in the British Household Panel Survey. There we first present the data and variables we include in our model. The estimates and comments on them follow. Last section concludes.

2 The Model and Estimation Method

We consider a dynamic panel data ordered probit with fixed effects:

$$h_{it}^* = \alpha_i + \rho_1 \mathbf{1}\left(h_{i,t-1} = 1\right) + \rho_{-1}\mathbf{1}\left(h_{i,t-1} = -1\right) + x_{it}'\beta + \varepsilon_{it};\ i = 1, ..., N,\ t = 0, .., T \tag{1}$$

h_{it}^* is the latent variable (e.g. health status), and the observed variable (h_{it}) is determined according to the following thresholds:

$$h_{it} = \begin{cases} -1 & \text{if} \quad h_{it}^* < -c_i \\ 0 & \text{if} \quad -c_i < h_{it}^* \leq 0 \\ 1 & \text{if} \quad h_{it}^* > 0 \end{cases} \tag{2}$$

For instance, in our empirical application, $h_{it} = -1$ corresponds to poor health, $h_{it} = 0$ to fair health and $h_{it} = 1$ to good health. α_i and c_i are the model's fixed effects, and $\varepsilon_{it} \underset{iid}{\sim} N(0, 1)$. Note that in addition to the usual scale normalization in discrete choice models, here we are also normalizing one of the two cut points to be zero. The, somehow more conventional, normalization of setting the intercept in the linear index equal to zero is not available to us because with the fixed effects approach the distribution of the intercept, including its mean, is unrestricted. An alternative normalization is to put the two fixed effects in the two cut point and leave the linear index equation without any intercept.

From this discussion on normalization it is clear that it is not possible to separately identify individual effects affecting only h_{it}^* from the individual effects affecting the cut points. Having only the fixed effect in the linear index (α_i) will also allow for heterogeneity in the cut points, but in a very restrictive way. In particular, by introducing only one individual effect (α_i), we would be assuming that the unobserved heterogeneity must have effects of opposite sign in $\Pr(h_{it} = 1)$ and $\Pr(h_{it} = -1)$; and also we would be restricting how these two effects differ in magnitude for all individuals. Having two fixed effects as in (2), we are not imposing any restrictions on the cut-point shifts as well as on the index shift.

From (1), (2) and the assumption about ε_{it}, we have that

$$\begin{aligned}
&\Pr(h_{it}=-1|x_{it},h_{it-1},c_i,\alpha_i)=1-\Phi\left(c_i+\mu_{it}\right)\\
&\Pr(h_{it}=0|x_{it},h_{it-1},c_i,\alpha_i)=\Phi\left(c_i+\mu_{it}\right)-\Phi\left(\mu_{it}\right)\\
&\Pr(h_{it}=1|x_{it},h_{it-1},c_i,\alpha_i)=1-\Pr(h_{it}=-1|.)-\Pr(h_{it}=0|.)=\Phi\left(\mu_{it}\right) \qquad (3)
\end{aligned}$$

where

$$\mu_{it}=\alpha_i+\rho_1\mathbf{1}\left(h_{i,t-1}=1\right)+\rho_{-1}\mathbf{1}\left(h_{i,t-1}=-1\right)+x_{it}'\beta \qquad (4)$$

Conditioning on the first observation, the log-likelihood is:

$$\begin{aligned}
l(\rho_1,\rho_{-1},\beta,\alpha,\mathbf{c})=\sum_{i=1}^{N}\sum_{t=1}^{T-1}\{&\mathbf{1}\left\{y_{it}=-1\right\}\log\left[1-\Phi\left(c_i+\mu_{it}\right)\right]+\\
&\mathbf{1}\left\{y_{it}=0\right\}\log\left[\Phi\left(c_i+\mu_{it}\right)-\Phi\left(\mu_{it}\right)\right]+\mathbf{1}\left\{y_{it}=1\right\}\log\left[\Phi\left(\mu_{it}\right)\right]\},
\end{aligned} \qquad (5)$$

2.1 Estimation problem and possible solutions

Using standard MLE to estimate models like (2) is well known to be biased, since we do not have a large number of periods. The MLE is inconsistent when T is not going to infinity because the fixed effects are acting as incidental parameters. Furthermore, existing Monte Carlo experiments with nonlinear models similar to this shows that the MLE has large bias. In fact, simulations of a dynamic ordered probit in Bester and Hansen (2009) and, in following sections, we show that the bias is non-negligible even with T as large as 20. As mentioned in the introduction, several bias-correction methods have been recently developed that could overcome this problem. Arellano and Hahn (2007) summarize the different approaches.

The methods can be grouped in three approaches based on the object that is corrected. The first one is to construct an analytical or numerical bias correction of a fixed effect estimator. Fernandez-Val (2009), among others, takes this approach to the problem and applies his analytical bias correction and a jackknife automatic correction to dynamic binary choice models. The second group are those that correct the bias in moment equations. An example of this is Carro (2007) that uses an estimator of this type to correct the bias in dynamic binary choice models. The third group are those that correct the objective function. Arellano and Hahn (2006) and Bester and Hansen (2009) take this approach, with the latter including an application to a dynamic ordered probit model. Given that our model of interest is also a dynamic ordered probit, and that other alternatives will require some sort of transformation or derivations to be applied to our case, the HS-penalty estimator studied in Bester and Hansen (2009) is the first option we should consider. In addition to that, this estimator has the advantages of being simpler to compute than the Modified MLE in Carro (2007) and than the Bias Correction in Fernandez-Val (2009) because the HS does not require the calculation of expectations and the other two do. This advantage is more relevant in our case, because it has two fixed effects. The HS is also obviously much less computationally costly than a jackknife automatic correction.

Arellano and Hahn (2007) shows the relations between the different type of approaches. Asymptotically all the methods and approaches are always reducing the order of the bias of the MLE from the standard $O(T^{-1})$ to $O(T^{-2})$ for the general classes of models they were developed. However there may be differences when they are applied to specific cases . The following very simple example, used in Carro (2007), Arellano and Hahn (2007), and Bester and Hansen (2009), illustrates this point. Consider the model where $y_{it} \underset{iid}{\sim} N(\eta_i, \sigma_0^2)$. The ML estimator of σ_0^2 is $\widehat{\sigma}^2_{MLE} = \frac{1}{NT} \sum_i \sum_t (y_{it} - \widehat{\eta}_i)^2$. It is well known that $\widehat{\sigma}^2_{MLE}$ is not a consistent estimator of σ_0^2 when $N \to \infty$ with fixed T, since it converges to $\frac{T-1}{T}\sigma_0^2$. In this case the whole problem is very easy to fix. $\frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \widehat{\eta}_i)^2$ is the fixed T consistent estimator of σ_0^2. The MMLE from Carro (2007) produces this very same estimator, correcting not only the $O(T^{-1})$ term of the bias, but all the asymptotic bias in this special example. The jackknife automatic correction gives the fixed T consistent estimator too. The HS removes the $O(T^{-1})$ term of the bias, but it does not attain the fixed-T consistent estimator. The one-step bias correction to the ML estimator from Fernandez-Val (2009) does not produce a fixed-T consistent estimator either, but its iterated form does. So, differences may appear between the different approaches when applied to specific models.

On the other hand, the incidental parameters problem can be seen as a finite sample bias problem in panel data context. The problem is not very important when T is large. However, since our panel does not have a large number of periods it is reasonable to wonder whether the good asymptotic properties when T goes to infinity are a good approximation to our finite sample. As a matter of fact, our problem is that the MLE has large biases when T is not very large. It seems from simulations that we would need panels with a much larger number of time periods than those usually found in practice. This also implies that we should look at the finite sample performance of the estimators for our model and sample sizes. In the methods considered here this is done through Monte Carlo experiments. Unfortunately, Bester and Hansen (2009) do not compare the finite sample properties of the method they use with others for the ordered probit case because many of the other methods will require some derivation to get the specific correction for this case. They, however, make such a comparison using a binary choice (probit and logit) models. Also, Carro (2006) and Fernandez-Val (2009) make Monte Carlo experiments for logit and probit models with different sample sizes (both in T and N), allowing us to compare a wide range of methods for these models. From these comparisons we can conclude that the HS penalty approach is clearly not the best one and for sample sizes with T smaller than 13 the reminding bias can still be significant. Given this result, we should consider other of the proposed methods to estimate our ordered probit and evaluate its finite sample properties. Interesting candidates are the corrections discussed by Fernandez-Val (2009) and Carro (2006) since they are equally superior to other alternatives in finite sample performance in the relevant existing comparisons. In the next subsection we derive explicit formulas of the modified MLE used in Carro (2007) for the model considered here and evaluate its finite sample performance.

2.2 MMLE for a dynamic order probit with two fixed effects

The model to be estimated is defined in (1) and (2), and its log-likelihood is (5). Let $\gamma = (\beta, \rho_1, \rho_{-1})$ and $\eta_i = (\alpha_i, c_i)$. Partial derivatives will be denoted by the letter d, so the first order conditions will be $\mathbf{d}_{\eta i}(\gamma, \eta_i) \equiv \frac{\partial l_i(\gamma,\eta_i)}{\partial \eta_i}$ and $\mathbf{d}_{\gamma i}(\gamma, \eta_i) \equiv \frac{\partial l_i(\gamma,\eta_i)}{\partial \gamma}$. Bold letters represent vectors.

The MLE of η_i for given γ, $\eta_i(\gamma)$, solves $\mathbf{d}_{\eta i}(\gamma, \eta_i) = 0$. Then, the MLE of γ is obtained by maximizing the concentrated log-likelihood ($\sum_{i=1}^{N} l_i(\gamma, \eta_i(\gamma))$), i.e. by solving the following first order condition:

$$\frac{1}{T\,N} \sum_{i=1}^{N} \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) = 0 \tag{6}$$

where $\mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) = \left.\frac{\partial l_i(\gamma,\eta_i)}{\partial \gamma}\right|_{\eta_i=\eta_i(\gamma)}$.

To reduce the bias of the estimation, we follow Carro (2006) in modifying the score of the concentrated log-likelihood adding a term that takes away the first order term of the asymptotic bias in T. By doing this, we get that the MMLE of the γ parameters of model (2) is the value that solves the following score equation:

$$\mathbf{d}_{\gamma Mi}(\gamma) = \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) - \frac{1}{2}\frac{1}{d_{\alpha\alpha i}d_{cci} - {d_{\alpha ci}}^2}\left[d_{\alpha\alpha i}\left(\mathbf{d}_{\gamma cci} + d_{\alpha cci}\frac{\partial\widehat{\alpha}_i}{\partial\gamma} + d_{ccci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)\right.$$
$$\left. + d_{cci}\left(\mathbf{d}_{\gamma\alpha\alpha i} + d_{\alpha\alpha\alpha i}\frac{\partial\widehat{\alpha}_i}{\partial\gamma} + d_{\alpha\alpha ci}\frac{\partial\hat{c}_i}{\partial\gamma}\right) - 2d_{\alpha ci}\left(\mathbf{d}_{\gamma\alpha ci} + d_{\alpha\alpha ci}\frac{\partial\widehat{\alpha}_i}{\partial\gamma} + d_{\alpha cci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)\right]$$
$$- \frac{\partial}{\partial\alpha_i}\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{\alpha ci}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{\alpha ci})]^2}\right)\Bigg|_{\eta_i=\eta_i(\gamma)}$$
$$- \frac{\partial}{\partial c_i}\left(\frac{E(\mathbf{d}_{\gamma\alpha i})E(d_{\alpha ci}) - E(d_{\alpha\alpha i})E(\mathbf{d}_{\gamma ci})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{\alpha ci})]^2}\right)\Bigg|_{\eta_i=\eta_i(\gamma)} = 0 \qquad (7)$$

where $\mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma))$ is the standard first order condition from the concentrated log-likelihood, as in (6). $\mathbf{d}_{\gamma ci} = \frac{\partial^2 l_i}{\partial\gamma\partial c_i}$, $d_{\alpha\alpha i} = \frac{\partial^2 l_i}{\partial\alpha_i^2}$, $\mathbf{d}_{\gamma\alpha ci} = \frac{\partial^3 l_i}{\partial\gamma\partial c_i\partial\alpha_i}$, and so on. $\widehat{\alpha}_i(\gamma)$ and $\widehat{c}_i(\gamma)$ are obtained from the first order conditions of α_i and c_i, as it is done in order to concentrate the log-likelihood.

We show in appendix A how this modification on the score of the concentrated log-likelihood in (7) is a first order adjustment on the asymptotic bias of the ML score, so the first order condition is more nearly unbiased and the order of the bias of the estimator is reduced from $O(T^{-1})$ to $O(T^{-2})$.

2.3 Simulations

We simulate model (1 - 2) with following value of the parameters: $\beta = 1$, $\rho_1 = 0.5$, and $\rho_{-1} = -0.5$. The error follows a normal distribution: $\varepsilon_{it} \sim N(0,1)$. The fixed effects are constructed as follows:

$$\alpha_i = \frac{1}{4}\sum_{t=1}^{4} x_{it} \qquad (8)$$

$$c_i = |z_i|, \quad \text{where } z_i \sim N(x_{i0}, 1). \qquad (9)$$

so that they are correlated with the explanatory variables.[3] x_{it} follows a Gaussian AR(1) with autoregressive parameter equal to 0.5. Initial conditions are $x_{i0} \sim N(0,1)$ and $h^*_{i0} = a_i + \beta_0\, x_{i0} + \varepsilon_{i0}$. We perform 1000 replications, with a population of $N = 250$ individuals. For each simulation we estimate the MLE, the MMLE given by equation (7) and the HS estimator defined in Bester and Hansen (2009).

[3]Note that Bester and Hansen (2009) only consider in their simulations of an order probit the case where the fixed effects are independent of the covariates. Correlation of the unobserved heterogeneity, as here, makes the problem more severe. The estimators are likely to perform worse, but we consider this situation to be more realistic. We want to evaluate the alternatives in a realistic setting.

Table 1: Monte Carlo Results. Dynamic Order Probit parameters

Parameter	β		ρ_1		ρ_{-1}	
True value	1		0.5		-0.5	
Estimator	Mean Bias	RMSE	Mean Bias.	RMSE	Mean Bias	RMSE
			$T = 4$			
MLE	0.816	0.828	-0.474	0.516	0.551	0.586
HS	0.796	0.809	-0.392	0.443	0.467	0.509
MMLE	0.172	0.182	-0.254	0.282	0.280	0.305
			$T = 8$			
MLE	0.335	0.341	-0.188	0.216	0.189	0.216
HS	0.247	0.254	-0.115	0.153	0.119	0.154
MMLE	0.073	0.086	-0.062	0.108	0.067	0.109
			$T = 10$			
MLE	0.257	0.263	-0.145	0.171	0.154	0.179
HS	0.170	0.178	-0.083	0.119	0.093	0.127
MMLE	0.052	0.067	-0.036	0.086	0.050	0.093
			$T = 12$			
MLE	0.210	0.215	-0.127	0.152	0.127	0.151
HS	0.127	0.134	-0.072	0.106	0.074	0.106
MMLE	0.040	0.054	-0.030	0.079	0.036	0.081
			$T = 16$			
MLE	0.154	0.159	-0.093	0.118	0.096	0.119
HS	0.081	0.088	-0.048	0.083	0.054	0.085
MMLE	0.026	0.041	-0.017	0.068	0.022	0.069
			$T = 20$			
MLE	0.122	0.127	-0.072	0.095	0.078	0.101
HS	0.058	0.065	-0.034	0.067	0.042	0.074
MMLE	0.019	0.034	-0.009	0.058	0.016	0.062

That is, the HS estimator is the value of the parameters that maximize the following penalized objective function:

$$\sum_{i=1}^{N} lk_i\left(\beta, \rho_1, \rho_{-1}, \alpha_i, c_i\right) - \sum_{i=1}^{N} \frac{1}{2} trace\left(\widehat{I}_{\alpha c_i}^{-1} \widehat{V}_{\alpha c_i}\right) - \frac{k}{2} \tag{10}$$

where lk_i is the log likelihood of i, $\widehat{I}_{\alpha c_i}$ is the sample information matrix for $e_i = (\alpha_i, c_i)'$, $\widehat{V}_{\alpha c_i}$ is a HAC estimator of $Var\left(\frac{1}{\sqrt{T}} \frac{\partial l_i}{\partial e_i}\right)$, and $k = \dim(e_i)$

Results from this experiment for different T are reported in Table 1, which shows the mean bias and the Root Mean Squared Error (RMSE). We find that for all T, the MMLE performs much better than the other two estimators. Comparing it with the HS, the differences are of greater magnitude for $T = 4$ and $T = 8$, where the HS is closer to the MLE than to the MMLE. When using the MMLE the bias is small than 10% of the true values with $T = 10$ for all but for one of the ρ parameters. With $T = 12$ the bias when using the MMLE is already negligible whereas the HS contain biases and RMSE larger than the MMLE with $T = 10$. Even with $T = 16$ the HS exhibit mean biases greater than the MMLE with $T = 10$. It is not until $T = 20$ that the HS has small biases and RMSE. So HS needs a larger number of periods (at least larger than 16) to have small finite sample biases. Given this and the fact that the sample sizes we have in the empirical application of this paper are smaller than $T = 14$, we will use MMLE.

3 Empirical application: self-assessed health status in the British Household Panel

Self-assessed health (SAH) measures have been used as a proxy for true overall individual health status in many socioeconomic studies. Also, it has been shown to be a good predictor of mortality and of subsequent demand of medical care. This motivates the study of dynamics and potential explanatory factors of SAH. Moreover, SAH measures exhibit high persistence and it is interesting to know the relative contributions of state dependence and heterogeneity to it. In this section we estimate a dynamic ordered probit of SAH with two fixed effects, using MMLE whose properties has been studied in previous sections.

Our model, in contrast with previous studies like Contoyannis, Jones and Rice (2004), includes two fixed effects: one in the linear index equation and another one in the cut points. The motivation for doing this is to account for heterogeneity in reporting behavior (cut-points) among individuals, in addition to accounting for unobserved factors that affect health status (index shift). The cut-point shifts occur if individuals use different thresholds when assessing their health and reporting it in the SAH categorical variable, so that they report a different value of SAH even though having the same level of true health. To control for these two, possibly correlated with other explanatory variables and between each other, unobserved factors, we include individual effects not only in the levels of the order probit but also in the cut points.

The model we estimate is as in (1) and (2):

$$h_{it}^{*} = \alpha_i + \rho_1 \mathbf{1}\left(h_{i,t-1} = 1\right) + \rho_{-1} \mathbf{1}\left(h_{i,t-1} = -1\right) + x_{it}'\beta + \varepsilon_{it} \qquad (11)$$

where h_{it}^* is the unobserved true health status of person i at period t, and the observed variable (h_{it}) is determined according to the following thresholds:

$$h_{it} = \begin{cases} -1 & \text{if} \quad h_{it}^* < -c_i \\ 0 & \text{if} \quad -c_i < h_{it}^* \leq 0 \\ 1 & \text{if} \quad h_{it}^* > 0 \end{cases} \tag{12}$$

where, $h_{it} = -1$ corresponds to the situation where poor health is reported, $h_{it} = 0$ to fair health and $h_{it} = 1$ to good health. α_i and c_i are the model's fixed effects, and $\varepsilon_{it} \underset{iid}{\sim} N(0, 1)$. The explanatory variables included in the model are described in the following subsection.

3.1 Data and variables

For our empirical analysis, we use the British Household Panel Survey (BHPS). This is a longitudinal survey of private households in Great Britain, and was designed as an annual survey of each adult (16+) member of a nationally representative sample of more that 5,000 households, with a total of approximately 10,000 individual interviews. The same individuals are re-interviewed in successive waves and, if they split off from their original households are also re-interviewed along with all adult members of their new households. Similarly, new members joining sample households become eligible for interview and children are interviewed as they reach the age of 16. Currently, sixteen waves of data for the years 1991 - 2006 are available. We take into account individuals who gave a full interview at each wave. An unbalanced panel of individuals who were interviewed in at least 8 subsequent waves is used. Our sample consists of 74,451 observations from 6,255 individuals.

SAH is defined for waves 1-8 and 10-16 as the response to the question "Compared to people of your own age, would you say your health over the last 12 months on the whole has been: excellent, good, fair, poor, very poor?" At wave 9 the SAH question and categories were reworded. This makes the comparison with other waves difficult and wave 9 is not used in our empirical analysis.

The original five SAH categories were collapsed to a three-category variable, creating a new SAH variable, that will be our dependent variable, with the following codes: poor ($h_{it} = -1$) for individuals who reported either "very poor" or "poor" health; fair ($h_{it} = 0$) for individuals who reported "fair" health; and Good ($h_{it} = 1$) for individuals who reported "good" or "excellent" health.

The explanatory x variables in (11) can be grouped in three categories:

1. Socioeconomic variables: three dummy variables representing marital status (Married, Widowed, Divorced/Separated), with Single as the reference category; six dummy variables representing employment (Self employed, In paid employment, Unemployed, Retired, Looking after family or home, Long term sick or disabled), with Other (On maternity leave, On a government training scheme, Full-time student/at school, Something else) as the reference category; and size of the household (the number of people living in the same household). The income variable is the logarithm of equivalised real income, adjusted using the Retail Price Index and equivalised by the McClement's scale to adjust for household size and composition, and consists on the sum of non-labour income and labour income in the reference year.

2. Health variables: Among the explanatory variables of overall self-assessed heath status, we include information on objective health problems. The BHPS contains several questions about health problems and health care demand, but many of them can be induced by a self valuation that might differ from true health as much as SAH, and in an unobserved way. For example the number of visits to the doctor can be determined by a perception of a health problem rather than a true health problem. To avoid this endogeneity bias, we have selected only those questions that we regard as measuring more objective health situations and, therefore, are not affected by personal health assessments. We introduce the following variables:

   - Health problems: This is a dummy variable, which takes the value 1 if the individual reports he/she has at least one of the following *permanent* health problems or disabilities: arthritis or rheumatism, difficulty in hearing, allergies, asthma, bronchitis, blood pressure, diabetes, migraine or frequent headaches, cancer and stroke, among others.

   - Health limits daily activities: This is a dummy variable, which takes the value 1 if the individual answers 'yes' to the following question: does your health in any way limit your daily activities, compared to most people of your age? Examples of daily activities included are: doing the housework, climbing stairs, dressing yourself, walking for at least 10 minutes, etc.

   - Health limits ability to work: Similar to previous question.

   - Number of days in a Hospital as an in-patient in the reference year.

   - Number of cigarettes smoked per day.

3. Other controls: We include year dummies (excluding the necessary number to

avoid prefect colinearity), age and age square. Note that the question about SAH that we use to construct our dependent variable asks for a comparison with the health of people with the same age as the respondent. However, there is a trend for SAH to become worse over time in the raw sample data that may indicate that the age effect over health is not being totally discounted by the respondents. This can be seen in table 3.[4] This is the reason for including age as explanatory variable.

Table 2: Number of individuals that reports each category of SAH by number of times it is reported.

Number	Excellent or good		Fair		Poor or very poor	
of times	Freq.	%	Freq. (N)	%	Freq. (N)	%
0	245	3.92	2062	32.97	4346	69.48
1	161	2.57	1105	17.67	878	14.04
2	173	2.77	842	13.46	360	5.76
3	186	2.97	628	10.04	202	3.23
4	227	3.63	468	7.48	130	2.08
5	268	4.28	364	5.82	91	1.45
6	386	6.17	261	4.17	68	1.09
7	454	7.26	196	3.13	45	0.72
8	670	10.71	144	2.3	39	0.62
9	554	8.86	79	1.26	30	0.48
10	523	8.36	57	0.91	28	0.45
11	482	7.71	20	0.32	13	0.21
12	539	8.62	20	0.32	8	0.13
13	670	10.71	5	0.08	8	0.13
14	717	11.46	4	0.06	9	0.14
Total	6255	100	6255	100	6255	100

Variables that are time-constant and specific for individuals, like the level of education and gender are not included in the set of explanatory variables since they can not be separately identified from the permanent unobserved heterogeneity. Therefore, the fixed effects account for these variables as well as for unobserved characteristics, and we cannot separate their effects. Sometimes this is seen as a drawback of the fixed effects approach. However, the random effects approach only separately identifies the effect of these variables because of the unrealistic assumption that the unobserved characteristics are independent from them (for example that unobserved healthy life style is independent of education). Even with

[4]See Contoyannis, Jones and Rice (2004) for further discussion on this.

Table 3: Proportion (in %) of each category of SAH by several characteristics

Characteristics and their Sample Proportions		SAH categories: Excellent or good	Fair	Poor or very poor
	All	73.74	19.23	7.03
By age group				
39.90	<40	78.81	16.32	4.87
43.80	40-64	73.77	18.67	7.56
16.29	65+	61.24	27.87	10.88
By sex				
47.23	Male	75.43	18.32	6.25
52.77	Female	72.23	20.04	7.72
Smoke				
23.95	Yes	68.06	22.22	9.72
76.05	No	75.53	18.29	6.18
By marital status				
63.56	Married	74.55	18.67	6.78
8.81	Divorced	70.80	19.26	9.94
6.46	Widowed	59.33	28.68	11.99
21.17	Single	76.94	18.02	5.05
Health problems				
58.13	Yes	61.31	27.13	11.56
41.87	No	91.00	8.26	0.73
Health limits daily activities				
12.80	Yes	23.08	39.55	37.37
87.20	No	81.18	16.25	2.58
Health limits work				
15.84	Yes	30.50	38.63	30.87
84.16	No	81.88	15.58	2.54

a correlated random effects approach, if correlation is allowed in a Mundlak (1978) and Chamberlain (1984) style and initial conditions are controlled for following the proposal in Wooldridge (2005), it is not possible to separately identify the effect of these time constant variables from the effect of the unobserved factors correlated with them. For instance, Contoyannis, Jones and Rice (2004) follows Wooldridge (2005) proposal and they comment about this impossibility of separating the effect of variables like education from the effect of the unobservables correlated with them.

Tables 2, 3 and 4 contain some descriptive numbers of the self-assessed heath reported in our sample. The most frequent category is excellent or good with more than 70% of the answers corresponding to this category. Also, there is high per-

Table 4: Sample transition probabilities from SAH in t-1 to SAH in t

		SAH in t			
		Excellent or good	Fair	Poor or very poor	Total
SAH	Excellent	86.03	11.76	2.21	100
in	Fair	43.60	45.18	11.22	100
$t-1$	Poor or very poor	18.45	32.05	49.50	100
	Proportion	73.34	19.51	7.15	100

sistence in SAH reported, as can be seen in table 4, which shows the transition probabilities. In this table, the largest numbers are in the diagonal for all three values of SAH_{t-1}. Table 3 presents the variation on SAH across different characteristics and health variables. People that smokes tend to select worse self-assessed health categories than those that do not smoke. Married or single people respond the excellent or good health category more frequently than widows or divorced. The three objective health measures in table 3 alter the SAH responses in the expected direction and in greater magnitude than the socioeconomic variables also presented in the table.

Although there are clear connections, this empirical application does not substitute Contoyannis, Jones and Rice (2004) since the latter contains a more detailed data description, makes further discussion of the estimated model and address other issues, like sample attrition, that are not considered in this paper. However our paper complements Contoyannis, Jones and Rice (2004) in several ways:

(i) We use more periods from the BHPS than them. They only use the first eight waves because the ninth contains a different question and categorization about SAH. While we drop the 9th wave too, we incorporate the waves after the 9th in our estimation. Since the model specified includes only one lag of h_{it}, we have all the variables we need for the 11th to 16th waves. For the 10th wave we have all the variables but h_{it-1} as it happens with the first wave. We treat the 10th wave like an initial observation and we condition it out in our likelihood leaving the probability of that observation totally unrestricted. Contoyannis, Jones and Rice (2004) can not do this because of their way of solving the initial conditions problem and the use of random effects.

(ii) In our model we have two individual specific effects: one in the linear index and one in the cut points. Lindeboom and van Doorslaer (2004) tests the existence of cut-point shifts and find clear evidence of different reporting behavior (cut-point shifting) for gender and age. Given that Contoyannis, Jones and Rice (2004) are imposing homogeneous cut points, they estimate different models by gender to allow for that differing reporting behavior, but they do not allow unrestricted different behavior by age. Our approach is robust to heterogenous cut points freely correlated with any of the determinants of SAH.

(iii) Use of fixed effects instead of random effects approach. The main advantages of this are that no arbitrary restriction is imposed in the correlation between the permanent unobserved heterogeneity and the observable variables, and that there is no initial conditions problem.

(iv) As an additional complement, our study includes some objective health measures, so we can see how much it is explained by the socioeconomic variables and by state dependence even after these measures are included.

3.2 Estimates

Table 5 presents the coefficient estimates for the dynamic ordered probit model based on three different estimators, that also includes different specification of the heterogeneity. The first estimated model (column I) is a pooled model without individual specific effects. The second (column II) is a correlated random effects specification with an individual effect in the linear index equation (the α_i parameter in (11), but with homogeneous cut points. Here, $\alpha_i = \alpha_0 + \alpha_1' h_{i1} + \alpha_2' \overline{x}_i + u_i$, where $\overline{x}_i$ is the average over the sample period of the exogenous variables, and $u_i \sim N(0, \sigma_u^2)$ independently of everything else. This is the kind of specification estimated in Contoyannis, Jones and Rice (2004) that accounts for the correlated heterogeneity and the initial condition following Wooldridge (2005). The last specification (column III) is the specification described in previous subsections, that is the model in (11) and (12) treating α_i and c_i as fixed effects. It is estimated by MMLE. The estimated value of the coefficients is not directly comparable. To compare magnitudes of the effects of different variables and estimates we look at the relative effects (i.e. ratio of coefficients), and at the average and median marginal effects reported in tables 6 and 7 for the variables with a coefficient significantly different from zero.[5]

[5]The marginal effects are averaged (or calculated their median) across the first eight waves of the panel as well as across individuals to obtain summary measures of the marginal effect representative of the situation of the population.

Table 5: Estimates

Variables	I Pooled	II Random Effects	III MMLE
Health in t-1: Good	0.8329*** (0.0131)	0.3396*** (0.0234)	0.3739*** (0.0231)
Health in t-1: Poor	-0.5704*** (0.0212)	-0.3005*** (0.0343)	-0.2732*** (0.0307)
Age	0.0087*** (0.0023)	0.0033 (0.0210)	-0.0180 (0.0291)
Age square	0.0000 (0.0000)	-0.0003** (0.0001)	-0.0003*** (0.0001)
Married	-0.0349* (0.0185)	0.1043 (0.0752)	0.0600 (0.0699)
Separated/Divorced	-0.0572** (0.0246)	0.1141 (0.1028)	0.0480 (0.0847)
Widowed	-0.0443 (0.0288)	0.2136 (0.1329)	0.0478 (0.1134)
Self employed	0.0652 (0.0410)	0.0353 (0.0839)	-0.0058 (0.0885)
In paid employment	0.0186 (0.0357)	0.0137 (0.0639)	0.0832 (0.0691)
Unemployed	-0.0094 (0.0478)	0.0485 (0.0786)	0.0949 (0.0907)
Retired	-0.0088 (0.0426)	-0.0645 (0.0891)	0.1111 (0.0864)
Looking after family	-0.0161 (0.0403)	-0.0768 (0.0784)	-0.0470 (0.0795)
Household size	-0.0124* (0.0064)	0.0538*** (0.0189)	0.0071 (0.0157)
Household Income	0.0355*** (0.0082)	-0.0233 (0.0191)	-0.0033 (0.0170)
Male	0.0035 (0.0120)	-0.0370 (0.0265)	
Non-white	-0.1306*** (0.0327)	-0.1057 (0.0709)	
Higher/1st degree	0.2082*** (0.0216)	0.2490*** (0.0466)	
HND/A level	0.1460*** (0.0171)	0.1862*** (0.1862)	
CSE/O level	0.1382*** (0.0156)	0.1933*** (0.0327)	
Long term sick or disa.	-0.2683*** (0.0493)	-0.2510** (0.1093)	-0.2315** (0.0999)
Health problems	-0.6181*** (0.0140)	-0.6244*** (0.0281)	-0.7780*** (0.0340)
Health limits daily acti.	-0.6462*** (0.0196)	-0.6067*** (0.0341)	-0.6837*** (0.0303)
Health limits work	-0.4403*** (0.0186)	-0.4337*** (0.0331)	-0.4949*** (0.0310)
Cigarettes per day	-0.0077*** (0.0007)	0.0034 (0.0026)	0.0042* (0.0023)
Hospital days	-0.0312*** (0.0013)	-0.0372*** (0.0021)	-0.0351*** (0.0008)
Cut point 1	-1.2934*** (0.0885)	-1.2519*** (0.2169)	
Cut point 2	0.0344 (0.0882)	0.2623 (0.2165)	

Standard errors are reported in parenthesis.
* significant at 10% ; ** significant at 5% ; *** significant at 1%
Estimates of year dummies in all models and within means of variables in random effects are not reported.

Table 6: Average Marginal Effects on Probability of reporting good and poor health for significant variables.

(a) Good

	I		II		III	
			Random			
	Pooled	St.Err.	Effects	St.Err.	MMLE	St.Err.
Health in t-1: Good	0.2376	0.0043	0.0742	0.0050	0.1138	0.0080
Health in t-1: Poor	-0.1974	0.0073	-0.0765	0.0093	-0.0820	0.0227
Age	0.0018	0.0001	-0.0040	0.0004	-0.0127	0.0083
Long term sick or disa.	-0.0659	0.0121	-0.0535	0.0244	-0.0661	0.0292
Health problems	-0.1430	0.0032	-0.1279	0.0066	-0.2287	0.0524
Health limits daily act.	-0.1766	0.0063	-0.1451	0.0095	-0.2043	0.0337
Health limits work	-0.1132	0.0054	-0.0985	0.0085	-0.1472	0.0141
Cigarettes per day	-0.0017	0.0002	0.0007	0.0005	0.0012	0.0007
Hospital days	-0.0070	0.0003	-0.0075	0.0004	-0.0100	0.0003

(b) Poor

	I		II		III	
			Random			
	Pooled	St.Err.	Effects	St.Err.	MMLE	St.Err.
Health in t-1: Good	-0.0702	0.0016	-0.0186	0.0012	-0.0678	0.0927
Health in t-1: Poor	0.1041	0.0047	0.0226	0.0030	0.0635	0.0676
Age	-0.0006	0.0001	0.0012	0.0001	0.0082	0.0161
Long term sick or disa.	0.0237	0.0044	0.0151	0.0071	0.0483	0.0567
Health problems	0.0413	0.0001	0.0291	0.0016	0.1206	0.1747
Health limits daily act.	0.0635	0.0025	0.0403	0.0030	0.1486	0.1703
Health limits work	0.0387	0.0019	0.0264	0.0024	0.1009	0.1216
Cigarettes per day	0.0006	0.0001	-0.0002	0.0001	-0.0008	0.0010
Hospital days	0.0024	0.0001	0.0021	0.0001	0.0065	0.0079

The pooled model exacerbates the state dependence effect due to the lack of permanent unobserved heterogeneity. It also interesting to note that smoking more cigarettes per day has a negative and significant effect over SAH (i.e. reduces the probability of reporting good health) in the pooled estimates. That correspond with the sample correlation between smoking and SAH in Table 3. However that effect is positive and significant once we allow for unobserved heterogeneity (columns II and III). This means that, once we have controlled for unobserved heterogeneity and everything else equal, smoking increases the probability of reporting good health as self-assessed health measure. This indicates that we should interpret the effect of smoking over SAH not as an objective health impact, but as an effect over the subjective perception over health. Though not reported, we also estimated by MLE

Table 7: Median Marginal Effects on Probability of reporting good and poor health for significant variables.

(a) Good

	I Pooled	II Random Effects	III MMLE
Health in t-1: Good	0.2418	0.0744	0.1194
Health in t-1: Poor	-0.2090	-0.0793	-0.0882
Age	0.0018	-0.0032	-0.0122
Long term sick or disabled	-0.0671	-0.0536	-0.0702
Health problems	-0.1382	-0.1184	-0.2413
Health limits daily activities	-0.1817	-0.1483	-0.2183
Health limits work	-0.1160	-0.0997	-0.1573
Cigarettes per day	-0.0017	0.0007	0.0013
Hospital days	-0.0071	-0.0074	-0.0105

(b) Poor

	I Pooled	II Random Effects	III MMLE
Health in t-1: Good	-0.0516	-0.0072	-0.0633
Health in t-1: Poor	0.0973	0.0119	0.0631
Age	-0.0003	0.0003	0.0074
Long term sick or disabled	0.0131	0.0054	0.0471
Health problems	0.0194	0.0085	0.1057
Health limits daily activities	0.0421	0.0190	0.1501
Health limits work	0.0233	0.0110	0.0989
Cigarettes per day	0.0003	-0.0001	-0.0007
Hospital days	0.0012	0.0006	0.0061

model in (11) and (12). As seen in the simulations it is severely biased, and that bias implies estimating much lower state dependance effects and higher effect of the other explanatory variables.

More interesting it is the comparison between the correlated random effects model (column II) and the MMLE (column III). In the MMLE case the effect of all explanatory variables (with a significant effect) increases in absolute value with respect to the random effects model. That includes also the effect of the state dependence (effect of h_{it-1}). Comparing columns II and III we can also see that the effect of h_{it-1} increases proportionally less than the effect of the other relevant explanatory variables. In the Random effects specification the ratio of the coefficient of 'health problems' over the coefficient of $\mathbf{1}\,(h_{i,t-1} = good)$ is around 1.8, whereas in the MMLE that ratio is 2.1. In any case, this increase in the effect of the explanatory variables, specially in the effect of state dependence, is remarkable because in the model in column III we are allowing for more, and more flexible permanent unobserved heterogeneity than in column II.[6] This is an indication that ignoring the added dimension of heterogeneity and the flexibility in the distribution of the fixed effects matters when estimating the model and the marginal effects of variables. It is not only a matter of the amount of heterogeneity but also a matter of the other restrictions being imposed in the model in column II.

Table 8: Proportion of individuals with marginal effects (on the probability of reporting good and poor) that are significantly different from zero at 10%.

Variable	Proportion	
	Good	Poor
Health in t-1: Good	51.41%	12.64%
Health in t-1: Poor	50.54%	21.73%
Age	30.97%	5.13%
Long term sick or disabled	37.33%	14.01%
Health problems	49.24%	11.05%
Health limits daily activities	51.34%	19.42%
Health limits work	50.47%	17.83%
Cigarettes per day	23.54%	7.51%
Hospital days	49.46%	18.12%

Focusing on the MML estimates, the two indicators of h_{it-1} and the variables that capture objective health problems have a significant effect over SAH, with the expected signs. As in Contoyannis, Jones and Rice (2004) we also find evidence of strong positive state dependence, even after including more heterogeneity and the objective health measures. Apart from age and the interpretation of the effect of smoking already commented, no socioeconomic variable has a significant effect. This is in contrast with apparent correlation in the sample between these variables and SAH described in table 3.

In addition to looking at the average and median marginal effects reported in tables 6 and 7, we look at how many individuals have a significant marginal effect in the sample given their particular situation and unobserved characteristics. Table 8 presents the proportion of individuals with a significant (at 10%) marginal effects over the probability of reporting good and bad health, for the same variables as in

[6]Remember here that permanent unobserved heterogeneity, state dependence and persistence in observable variables are alternative explanations of the observed high persistence in h_{it}.

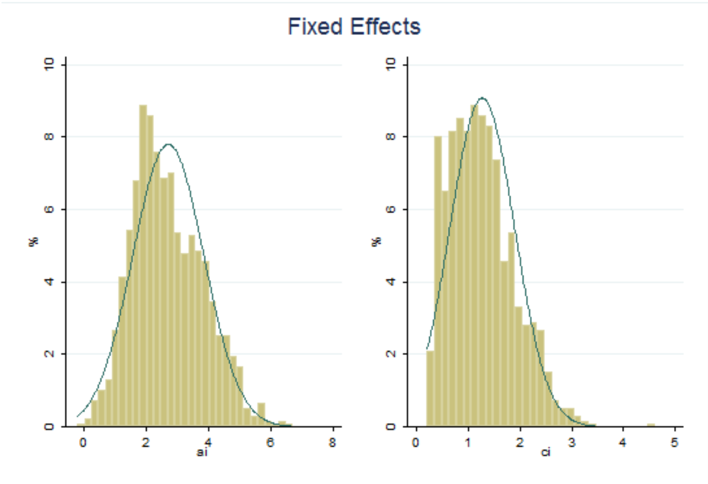

Figure 1: Distribution (histogram) of the fixed effects from MML estimates.

table 6. Notice that although the average marginal effects are significant, there is a great deal of heterogeneity so that for half of the population the marginal effects over the probability of reporting good health is not significantly different from zero for many of these variables.

Lastly, we look at the unobserved heterogeneity both in the linear index equation and in the cut point shift. Figure 1 displays the estimated distribution (histogram) of both fixed effects in the population. Both exhibit important variation. The average for α_i is 2.70 and 1.27 for c_i. The standard deviations are 1.14 and 0.62 respectively. Focusing on the heterogeneity on the cut points, though not a formal test, we can compare the estimated cut points in the model (12) with the estimated cut points in the random effects model. In (12) the second cut point have been normalized to be zero. Interestingly its estimate in the random effects model it is not significantly different from zero. With respect to the first cut point, the average of $-c_i$ is very close to the estimate of the first cut point in the random effects specification. However, as can be seen in the right panel of figure 1 there is important variation in c_i among individuals and the distribution is clearly asymmetric. A normal density, i.e. the continuous lines in Figure 1, does not fit the distribution of the fixed effects.

4 Conclusion

In this paper we have considered the estimation of a dynamic ordered probit with fixed effects of a self-assessed health status, which includes two fixed effects: one in the linear index equation, interpreted as unobserved health status, and another one in the cut points, interpreted as heterogeneity in reporting behavior. Based on our best estimates, the two fixed effects exhibit important variation and it is relevant to flexible account for both when estimating the effect of other variables. Our estimates show the state dependence is very important even though we have controlled for unobserved heterogeneity and some forms of objective health measures. The latter are the variables with higher marginal effects.

The recent literature in bias-adjusted methods of estimation of nonlinear panel data models with fixed effects has produced several potentially equivalent estimators. Here we find that the most directly and easily applicable correction to our model, which is the HS estimator proposed in Bester and Hansen (2009), has still important biases in our sample size. This lead us to consider the Modified MLE proposed in Carro (2007). We derive the expression of the MMLE in our case, and perform Monte Carlo experiments to evaluate its finite sample properties and compare it with the HS. The MMLE has a negligible bias in our sample size. These Monte Carlo experiments contribute to the mentioned literature on bias-adjusted methods of estimation by showing how well two of the proposed methods work for a specific model and sample size. Also, this will be useful information for other applications when having to choose among the several correction methods.

References

[1] Arellano, M. and J. Hahn (2007): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments". in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*, Volume 3, edited by Richard Blundell, Whitney Newey, and Torsten Persson. Cambridge University Press.

[2] Arellano, M. and J. Hahn (2006): "A likelihood-based approximate solution to the incidental parameter problemin dynamic nonlinear models with multiple effects", *unpublished manuscript.*

[3] Bester, C. A. and C. Hansen (2009): "A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects". *Journal of Business & Economic Statistic,* 27 (2):131-148

[4] Carro, Jesús M. (2007) "Estimating dynamic panel data discrete choice models with fixed effects". *Journal of Econometrics,* 140 (2007):503-528

[5] Chamberlain, G. (1984): "Panel Data", in Griliches, Z. and M.D. Intriligator (eds.) *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.

[6] Contoyannis, P., A. M. Jones and N. Rice (2004): "The Dynamics of Health in the British Household Panel Survey" *Journal of Applied Econometrics,* 19: 473-503

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

[8] Greene, W. H. and D. A. Henshen (2008): "Modeling Ordered Choices: A Primer and Recent Developments", Available at SSRN: http://ssrn.com/abstract=1213093.

[9] Hahn, J. and G. Kuersteiner (2004): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects", Mimeo.

[10] Hahn, J. and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", *Econometrica,* 72(4): 1295-1319.

[11] Jones, A. M. (2007): "Panel Data Methods and Applications to Health Economics", to appear in *The Palgrave Handbook of Econometrics Volume II:*

Applied Econometrics, edited by Terence C. Mills and Kerry Patterson. Basingstoke: Palgrave MacMillan.

[12] Lindeboom and van Doorslaer (2004):"Cut-point shift and index shift in self-reported health" *Journal of Health Economics*, 23: 1083-1099

[13] Mundlak, Y. (1978): "On the pooling of time series and cross-section data", *Econometrica,* 46(1): 69-85.

[14] Wooldridge (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity ", *Journal of Applied Econometrics,* 20: 39-54.

A Appendix: Reduction of the order of the bias

In this appendix we show that the modified score presented above is less biased than the original score. This follows Carro (2006), adapting it to our model with two fixed effects.

The notation used is the same as before: we denote partial derivatives by the letter d; bold letters are used to denote vectors; the derivatives evaluated at the true values of the parameters are represented by including a 0 in the sub-index (e.g. $d_{\eta i0} = d_{\eta i}(\gamma_0, \eta_{i0})$).

A.1 Deriving the leading term of the bias of the score in the MLE

We start by deriving the first term of the bias in the score of the original unmodified concentrated log-likelihood. Expanding this score around η_{i0}, and evaluating it at γ_0 we get:

$$\begin{aligned}\mathbf{d}_{\gamma i}(\gamma_0, \eta_i(\gamma_0)) = \mathbf{d}_{\gamma i0} + \; & d_{\gamma ai0}(\hat{a}_i(\gamma_0) - a_{i0}) \\ + \; & \mathbf{d}_{\gamma ci0}(\hat{c}_i(\gamma_0) - c_{i0}) \\ + \; & \frac{1}{2}\mathbf{d}_{\gamma aai0}(\hat{a}_i(\gamma_0) - a_{i0})^2 + \frac{1}{2}\mathbf{d}_{\gamma cci0}(\hat{c}_i(\gamma_0) - c_{i0})^2 \\ + \; & \mathbf{d}_{\gamma aci0}(\hat{a}_i(\gamma_0) - a_{i0})(\hat{c}_i(\gamma_0) - c_{i0}) + O_p(T^{-1/2}) + \ldots\end{aligned} \tag{A1}$$

Now we need expressions for $(\hat{a}_i(\gamma_0) - a_{i0})$ and $(\hat{c}_i(\gamma_0) - c_{i0})$, for which we do asymptotic expansions, following Rilstone, Srivastava and Ullah (1996):

$$(\hat{a}_i(\gamma_0) - a_{i0}) = b^a_{-1/2} + b^a_{-1} + O_p(T^{-3/2}) \tag{A2}$$

$$(\hat{c}_i(\gamma_0) - c_{i0}) = b^c_{-1/2} + b^c_{-1} + O_p(T^{-3/2}) \tag{A3}$$

where

$$b^a_{-1/2} = \frac{\frac{1}{T}d_{ci0}E\left(\frac{1}{T}\,d_{aci0}\right) - \frac{1}{T}d_{ai0}E\left(\frac{1}{T}\,d_{cci0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} \tag{A4}$$

$$b^c_{-1/2} = \frac{\frac{1}{T}d_{ai0}E\left(\frac{1}{T}\,d_{aci0}\right) - \frac{1}{T}d_{ci0}E\left(\frac{1}{T}\,d_{aai0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} \tag{A5}$$

It is also useful to obtain:

$$(\hat{a}_i(\gamma_0) - a_{i0})^2 = (b^a_{-1/2})^2 + O_p(T^{-3/2}) \tag{A6}$$

$$(\hat{c}_i(\gamma_0) - c_{i0})^2 = (b^c_{-1/2})^2 + O_p(T^{-3/2}) \tag{A7}$$

$$(\hat{a}_i(\gamma_0) - a_{i0})\,(\hat{c}_i(\gamma_0) - c_{i0}) = b^a_{-1/2}\,b^c_{-1/2} + O_p(T^{-3/2}) \tag{A8}$$

With respect to the squares of $b^a_{-1/2}$ and $b^c_{-1/2}$, we get:

$$(b^a_{-1/2})^2 = \frac{\left(\frac{1}{T}d_{ai0}\right)^2 E\left(\frac{1}{T}\,d_{cci0}\right)^2 + \left(\frac{1}{T}d_{ci0}\right)^2 E\left(\frac{1}{T}\,d_{aci0}\right)^2 - 2\,\frac{1}{T}d_{ai0}\frac{1}{T}d_{ci0}E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)}{\left(E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2\right)^2}$$

$$(b^c_{-1/2})^2 = \frac{\left(\frac{1}{T}d_{ci0}\right)^2 E\left(\frac{1}{T}\,d_{aai0}\right)^2 + \left(\frac{1}{T}d_{ai0}\right)^2 E\left(\frac{1}{T}\,d_{aci0}\right)^2 - 2\,\frac{1}{T}d_{ai0}\frac{1}{T}d_{ci0}E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{aci0}\right)}{\left(E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2\right)^2}$$

Substituting by expectations, and using the information matrix identity ($E(d_{aci}) = -E(d_{ai}d_{ci})$), we get:

$$(b^a_{-1/2})^2 = -\frac{1}{T}\,\frac{E\left(\frac{1}{T}\,d_{cci0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} + O_p(T^{-3/2}) \tag{A9}$$

$$(b^c_{-1/2})^2 = -\frac{1}{T}\,\frac{E\left(\frac{1}{T}\,d_{aai0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} + O_p(T^{-3/2}) \tag{A10}$$

Following the same procedure for the cross-product, we get:

$$b^a_{-1/2}\,b^c_{-1/2} = \frac{1}{T}\,\frac{E\left(\frac{1}{T}\,d_{aci0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} + O_p(T^{-3/2}) \tag{A11}$$

With respect to b^a_{-1} and b^c_{-1}, we follow the same procedure (replace by expectations and use the information matrix identity) to get:

$$b^a_{-1} = \frac{1}{2\,T}\,\frac{1}{\left(E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2\right)^2} \tag{A12}$$

$$\Bigg\{2\,E\left(\frac{1}{T}\,d_{aci0}\right)^2\left(E\left(\frac{1}{T}\,d_{acci0}\right) + E\left(\frac{1}{T}\,d_{ai0}d_{cci0}\right) + E\left(\frac{1}{T}\,d_{ci0}d_{aci0}\right)\right)$$

$$+\,E\left(\frac{1}{T}\,d_{cci0}\right)^2\left[E\left(\frac{1}{T}\,d_{aaai0}\right) + 2E\left(\frac{1}{T}\,d_{ai0}d_{aai0}\right)\right]$$

$$+\,E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[E\left(\frac{1}{T}\,d_{acci0}\right) + 2E\left(\frac{1}{T}\,d_{ci0}d_{aci0}\right)\right]$$

$$-\,E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,d_{aai0}\right)\left[E\left(\frac{1}{T}\,d_{ccci0}\right) + 2E\left(\frac{1}{T}\,d_{ci0}d_{cci0}\right)\right]$$

$$-E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[3E\left(\frac{1}{T}\,d_{aaci0}\right) + 4E\left(\frac{1}{T}\,d_{ai0}d_{aci0}\right) + 2E\left(\frac{1}{T}\,d_{ci0}d_{aai0}\right)\right]\Bigg\}$$

$$+\,O_p(T^{-3/2})$$

$$b^c_{-1} = \frac{1}{2\,T}\,\frac{1}{\left(E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2\right)^2} \tag{A13}$$

$$\Bigg\{2\,E\left(\frac{1}{T}\,d_{aci0}\right)^2\left[E\left(\frac{1}{T}\,d_{aaci0}\right) + E\left(\frac{1}{T}\,d_{ci0}d_{aai0}\right) + E\left(\frac{1}{T}\,d_{ai0}d_{aci0}\right)\right]$$

$$+\,E\left(\frac{1}{T}\,d_{aai0}\right)^2\left[E\left(\frac{1}{T}\,d_{ccci0}\right) + 2E\left(\frac{1}{T}\,d_{ci0}d_{cci0}\right)\right]$$

$$+\,E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[E\left(\frac{1}{T}\,d_{aaci0}\right) + 2E\left(\frac{1}{T}\,d_{ai0}d_{aci0}\right)\right]$$

$$-\,E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,d_{cci0}\right)\left[E\left(\frac{1}{T}\,d_{aaai0}\right) + 2E\left(\frac{1}{T}\,d_{ai0}d_{aai0}\right)\right]$$

$$-E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,d_{aai0}\right)\left[3E\left(\frac{1}{T}\,d_{acci0}\right) + 4E\left(\frac{1}{T}\,d_{ci0}d_{aci0}\right) + 2E\left(\frac{1}{T}\,d_{ai0}d_{cci0}\right)\right]\Bigg\}$$

$$+\,O_p(T^{-3/2}) \tag{A14}$$

Introducing all these expressions in (A1), and taking expectations, we get: :

$$E(d_{gi}(g_0,\hat{e}_i(g_0))) = \tag{A15}$$

$$\frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ai0}d_{ci0}\right)E\left(\frac{1}{T}d_{aci0}\right)-E\left(\frac{1}{T}\mathbf{d}_{\gamma ai0}d_{ai0}\right)E\left(\frac{1}{T}d_{cci0}\right)}{E\left(\frac{1}{T}d_{aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)-E\left(\frac{1}{T}d_{aci0}\right)^2}$$

$$+\frac{1}{2}\frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ai0}\right)}{\left(E\left(\frac{1}{T}d_{aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)-E\left(\frac{1}{T}d_{aci0}\right)^2\right)^2}$$

$$\left\{2\,E\left(\frac{1}{T}d_{aci0}\right)^2\left(E\left(\frac{1}{T}d_{acci0}\right)+E\left(\frac{1}{T}d_{ai0}d_{cci0}\right)+E\left(\frac{1}{T}d_{ci0}d_{aci0}\right)\right)\right.$$

$$+E\left(\frac{1}{T}d_{cci0}\right)^2\left[E\left(\frac{1}{T}d_{aaai0}\right)+2E\left(\frac{1}{T}d_{ai0}d_{aai0}\right)\right]$$

$$+E\left(\frac{1}{T}d_{aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)\left[E\left(\frac{1}{T}d_{accci0}\right)+2E\left(\frac{1}{T}d_{ci0}d_{aci0}\right)\right]$$

$$-E\left(\frac{1}{T}d_{aci0}\right)E\left(\frac{1}{T}d_{aai0}\right)\left[E\left(\frac{1}{T}d_{ccci0}\right)+2E\left(\frac{1}{T}d_{ci0}d_{cci0}\right)\right]$$

$$\left.-E\left(\frac{1}{T}d_{aci0}\right)E\left(\frac{1}{T}d_{cci0}\right)\left[3E\left(\frac{1}{T}d_{aaci0}\right)+4E\left(\frac{1}{T}d_{ai0}d_{aci0}\right)+2E\left(\frac{1}{T}d_{ci0}d_{aai0}\right)\right]\right\}$$

$$+\frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}d_{ai0}\right)E\left(\frac{1}{T}d_{aci0}\right)-E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}d_{ci0}\right)E\left(\frac{1}{T}d_{aai0}\right)}{E\left(\frac{1}{T}d_{aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)-E\left(\frac{1}{T}d_{aci0}\right)^2}$$

$$+\frac{1}{2}\frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)}{\left(E\left(\frac{1}{T}d_{aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)-E\left(\frac{1}{T}d_{aci0}\right)^2\right)^2}$$

$$\left\{2\,E\left(\frac{1}{T}d_{aci0}\right)^2\left[E\left(\frac{1}{T}d_{aaci0}\right)+E\left(\frac{1}{T}d_{ci0}d_{aai0}\right)+E\left(\frac{1}{T}d_{ai0}d_{aci0}\right)\right]\right.$$

$$+E\left(\frac{1}{T}d_{aai0}\right)^2\left[E\left(\frac{1}{T}d_{cccci0}\right)+2E\left(\frac{1}{T}d_{ci0}d_{cci0}\right)\right]$$

$$+E\left(\frac{1}{T}d_{aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)\left[E\left(\frac{1}{T}d_{aaci0}\right)+2E\left(\frac{1}{T}d_{ai0}d_{aci0}\right)\right]$$

$$-E\left(\frac{1}{T}d_{aci0}\right)E\left(\frac{1}{T}d_{cci0}\right)\left[E\left(\frac{1}{T}d_{aaai0}\right)+2E\left(\frac{1}{T}d_{ai0}d_{aai0}\right)\right]$$

$$\left.-E\left(\frac{1}{T}d_{aci0}\right)E\left(\frac{1}{T}d_{aai0}\right)\left[3E\left(\frac{1}{T}d_{acci0}\right)+4E\left(\frac{1}{T}d_{ci0}d_{aci0}\right)+2E\left(\frac{1}{T}d_{ai0}d_{cci0}\right)\right]\right\}$$

$$+\frac{1}{E\left(\frac{1}{T}d_{aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)-E\left(\frac{1}{T}d_{aci0}\right)^2}$$

$$\left[E\left(\frac{1}{T}\mathbf{d}_{\gamma aci0}\right)E\left(\frac{1}{T}d_{aci0}\right)-\frac{1}{2}E\left(\frac{1}{T}\mathbf{d}_{\gamma aai0}\right)E\left(\frac{1}{T}d_{cci0}\right)-\frac{1}{2}E\left(\frac{1}{T}\mathbf{d}_{\gamma cci0}\right)E\left(\frac{1}{T}d_{aai0}\right)\right]$$

$$+O(T^{-1})$$

The remainder of this expression is $O(T^{-1})$ because $O_p(T^{-1/2})$ terms have zero mean. This means that the score of the original concentrated likelihood has a bias of order $O(1)$, whose expression is in the previous formulae.

A.2 Modified Score

The modified score in (7) can be decomposed in three terms, $\mathbf{d}_{\gamma Mi}(\gamma) = A + B + C$, such that:

$$
\begin{aligned}
A &= \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) && \text{(A16)}\\
B &= -\frac{1}{2}\,\frac{1}{d_{aai}d_{cci} - {d_{aci}}^2} && \text{(A17)}\\
&\left[d_{aai}\left(\mathbf{d}_{\gamma cci} + d_{acci}\frac{\partial \hat{a}_i}{\partial \gamma} + d_{ccci}\frac{\partial \hat{c}_i}{\partial \gamma}\right)\right.\\
&+ d_{cci}\left(\mathbf{d}_{\gamma aai} + d_{aaai}\frac{\partial \hat{a}_i}{\partial \gamma} + d_{aaci}\frac{\partial \hat{c}_i}{\partial \gamma}\right)\\
&\left.-2d_{aci}\left(\mathbf{d}_{\gamma aci} + d_{aaci}\frac{\partial \hat{a}_i}{\partial \gamma} + d_{acci}\frac{\partial \hat{c}_i}{\partial \gamma}\right)\right]\\
C &= -\frac{\partial}{\partial a_i}\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{aci}) - E(d_{cci})E(\mathbf{d}_{\gamma ai})}{E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2}\right)\Bigg|_{\eta_i=\eta_i(\gamma)} && \text{(A18)}\\
&-\frac{\partial}{\partial c_i}\left(\frac{E(\mathbf{d}_{\gamma ai})E(d_{aci}) - E(d_{aai})E(\mathbf{d}_{\gamma ci})}{E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2}\right)\Bigg|_{\eta_i=\eta_i(\gamma)}
\end{aligned}
$$

A is the score of the original un-modified concentrated log-likelihood. So, we now analyze B and C.

Part B**.** We first want to derive expression for $\partial \hat{a}_i/\partial\gamma$ and $\partial \hat{c}_i/\partial\gamma$. Differentiating the score of the concentrated log-likelihood, $\mathbf{d}_{\eta i}(\gamma, \eta_i(\gamma))$, with respect to γ we get a system of two equations with two unknowns. Solving for $\partial \hat{a}_i/\partial\gamma$ and $\partial \hat{c}_i/\partial\gamma$ we get:

$$
\frac{\partial \hat{a}_i(\gamma)}{\partial \gamma} = \frac{\mathbf{d}_{\gamma ci}d_{aci} - d_{cci}\mathbf{d}_{\gamma ai}}{d_{aai}d_{cci} - d_{aci}^2} \tag{A19}
$$

$$
\frac{\partial \hat{c}_i(\gamma)}{\partial \gamma} = \frac{\mathbf{d}_{\gamma ai}d_{aci} - d_{aai}\mathbf{d}_{\gamma ci}}{d_{aai}d_{cci} - d_{aci}^2} \tag{A20}
$$

evaluating at γ_0 and replacing by expectations:

$$
\frac{\partial \hat{a}_i(\gamma_0)}{\partial \gamma} = \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}\,d_{aci0}\right) - E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} + O_p(T^{-\frac{1}{2}}) \tag{A21}
$$

$$
\frac{\partial \hat{c}_i(\gamma_0)}{\partial \gamma} = \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)E\left(\frac{1}{T}\,d_{aci0}\right) - E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} + O_p(T^{-\frac{1}{2}}) \tag{A22}
$$

Introducing in (A17) and rearranging terms:

$$B = -\frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}\,d_{aci0}\right) - E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2} \tag{A23}$$
$$\frac{d_{aai}d_{acci} + d_{cci}d_{aaai} - 2d_{aci}d_{aaci}}{2(d_{aai}d_{cci} - d_{aci}^2)}$$
$$- \frac{d_{aai}d_{acci} + d_{cci}d_{aaai} - 2d_{aci}d_{aaci}}{2(d_{aai}d_{cci} - d_{aci}^2)}\,O_p(T^{-1/2})$$
$$- \frac{E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)E\left(\frac{1}{T}\,d_{aci0}\right) - E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2}$$
$$\frac{d_{cci}d_{aaci} + d_{aai}d_{ccci} - 2d_{aci}d_{acci}}{2(d_{aai}d_{cci} - d_{aci}^2)}$$
$$- \frac{d_{cci}d_{aaci} + d_{aai}d_{ccci} - 2d_{aci}d_{acci}}{2(d_{aai}d_{cci} - d_{aci}^2)}\,O_p(T^{-1/2})$$
$$- \frac{d_{aai}\mathbf{d}_{\gamma cci} + d_{cci}\mathbf{d}_{\gamma aai} - 2d_{aci}\mathbf{d}_{\gamma aci}}{2(d_{aai}d_{cci} - d_{aci}^2)}$$

Evaluating at γ_0, using the fact that $\eta_i(\gamma) = \eta_{i0} + O_p(T^{-1/2})$, adding $1/T^2$ in numerators and denominators and replacing by expectations:

$$B = -\frac{1}{2}\frac{1}{\left(E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2\right)^2} \tag{A24}$$
$$\left\{\left[E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}\,d_{aci0}\right) - E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)\right]\right.$$
$$\left[E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{acci0}\right) + E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,d_{aaai0}\right) - 2E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,d_{aaci0}\right)\right]$$
$$+\left[E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ai0}\right)E\left(\frac{1}{T}\,d_{aci0}\right) - E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma ci0}\right)\right]$$
$$\left.\left[E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,d_{aaci0}\right) + E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{ccci0}\right) - 2E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,d_{acci0}\right)\right]\right\}$$
$$-\frac{1}{2}\frac{1}{E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,d_{cci0}\right) - E\left(\frac{1}{T}\,d_{aci0}\right)^2}$$
$$\left[E\left(\frac{1}{T}\,d_{aai0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma cci0}\right) + E\left(\frac{1}{T}\,d_{cci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma aai0}\right) - 2E\left(\frac{1}{T}\,d_{aci0}\right)E\left(\frac{1}{T}\,\mathbf{d}_{\gamma aci0}\right)\right]$$
$$+\,O_p(T^{-1/2}) \tag{A25}$$

Finally, taking the expected value of this expression will not change anything, except that the remainder would be $O(T^{-1})$ instead of $O_p(T^{-1/2})$.

Part C. To analyze C, we need the following result:

$$\frac{\partial}{\partial a_i}E\left(\mathbf{d}_{\gamma ci}\right) = E\left(\mathbf{d}_{\gamma aci}\right) + E\left(\mathbf{d}_{\gamma ci}d_{ai}\right) \tag{A26}$$

This works with other derivatives of expectations as well.

We are interested in the following derivative, which we will call C^a:

$$\begin{aligned}
C^a &= -\frac{\partial}{\partial a_i}\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{aci}) - E(d_{cci})E(\mathbf{d}_{\gamma ai})}{E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2}\right) \\
&= -\frac{\frac{\partial}{\partial a_i}\left(E(\mathbf{d}_{\gamma ci})E(d_{aci}) - E(d_{cci})E(\mathbf{d}_{\gamma ai})\right)}{E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2} \\
&+ \frac{\left(E(\mathbf{d}_{\gamma ci})E(d_{aci}) - E(d_{cci})E(\mathbf{d}_{\gamma ai})\right)\frac{\partial}{\partial a_i}\left(E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2\right)}{\left(E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2\right)^2}
\end{aligned}$$

Working with the derivative and using the above rule, we get:

$$\begin{aligned}
C^a = &-\frac{1}{E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2} \\
&\{E(\mathbf{d}_{\gamma ci})\left[E(d_{aaci}) + E(d_{aci}d_{ai})\right] + E(d_{aci})\left[E(\mathbf{d}_{\gamma aci}) + E(\mathbf{d}_{\gamma ci}d_{ai})\right] \\
&-E(d_{cci})\left[E(\mathbf{d}_{\gamma aai}) + E(\mathbf{d}_{\gamma ai}d_{ai})\right] - E(\mathbf{d}_{\gamma ai})\left[E(d_{acci}) + E(d_{cci}d_{ai})\right]\} \\
&+ \frac{E(\mathbf{d}_{\gamma ci})E(d_{aci}) - E(d_{cci})E(\mathbf{d}_{\gamma ai})}{\left(E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2\right)^2} \\
&\{E(d_{aai})\left[E(d_{acci}) + E(d_{cci}d_{ai})\right] + E(d_{cci})\left[E(d_{aaai}) + E(d_{aai}d_{ai})\right] \\
&-2E(d_{aci})\left[E(d_{aaci}) + E(d_{aci}d_{ai})\right]\}
\end{aligned}$$

Likewise, for C^c we have:

$$\begin{aligned}
C^c = &-\frac{1}{E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2} \\
&\{E(\mathbf{d}_{\gamma ai})\left[E(d_{acci}) + E(d_{aci}d_{ci})\right] + E(d_{aci})\left[E(\mathbf{d}_{\gamma aci}) + E(\mathbf{d}_{\gamma ai}d_{ci})\right] \\
&-E(d_{aai})\left[E(\mathbf{d}_{\gamma cci}) + E(\mathbf{d}_{\gamma ci}d_{ci})\right] - E(\mathbf{d}_{\gamma ci})\left[E(d_{aaci}) + E(d_{aai}d_{ci})\right]\} \\
&+ \frac{E(\mathbf{d}_{\gamma ai})E(d_{aci}) - E(d_{aai})E(\mathbf{d}_{\gamma ci})}{\left(E(d_{aai})E(d_{cci}) - [E(d_{aci})]^2\right)^2} \\
&\{E(d_{cci})\left[E(d_{aaci}) + E(d_{aai}d_{ci})\right] + E(d_{aai})\left[E(d_{ccci}) + E(d_{cci}d_{ci})\right] \\
&-2E(d_{aci})\left[E(d_{acci}) + E(d_{aci}d_{ci})\right]\}
\end{aligned}$$

We then evaluate at γ_0 and take the expected value of these expressions.

Putting everything together. If we, finally, add all the terms of B and C from before, which is equal to $\mathbf{d}_{\gamma Mi}(\gamma) - \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) = B + C$, we get exactly minus (A15). Therefore, the modified score equal the standard score minus the first order term of the bias, because we are subtracting it with the modification $B+C$. The reminder of this expansion for $\mathbf{d}_{\gamma Mi}(\gamma)$ is $O(T^{-1})$, as opposed to $O(1)$ that is the order of magnitude of the bias of $\mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma))$. This shows that MMLE reduced the order of the bias of the MLE.