JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **29**: 181–207 (2014)
Published online 2 September 2012 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2301

STATE DEPENDENCE AND HETEROGENEITY IN HEALTH USING A BIAS-CORRECTED FIXED-EFFECTS ESTIMATOR

JESUS M. CARRO[a*] AND ALEJANDRA TRAFERRI[b]

[a] *Department of Economics, Universidad Carlos III de Madrid, Gatafe, Spain*
[b] *Institute of Economics, Pontificia Universidad Católica de Chile, Santiago, Chile*

SUMMARY

This paper estimates a dynamic ordered probit model of self-assessed health with two fixed effects: one in the linear index equation and one in the cut-points. This robustly controls for heterogeneity in unobserved health status and in reporting behavior, although we cannot separate both sources of heterogeneity. We find important state dependence effects, and small but significant effects of income and other socioeconomic variables. Having dynamics and flexibly accounting for unobserved heterogeneity matters for those estimates. We also contribute to the bias correction literature in nonlinear panel models by comparing and applying two of the existing proposals to our model. Copyright © 2012 John Wiley & Sons, Ltd.

Received 14 July 2010; Revised 08 July 2012

Supporting information may be found in the online version of this article.

1. INTRODUCTION

We use data from the British Household Panel Survey (BHPS) to study the determinants of self-assessed health status (SAH), and we estimate a dynamic ordered probit model, controlling for state dependence and two sources of heterogeneity: heterogeneity in unobserved factors affecting health and heterogeneity in reporting behavior. Our model and estimation strategy allow us to answer two important questions in this literature: (i) What are the relative contributions of state dependence and heterogeneity in explaining the observed persistence in SAH? (ii) What are the effects of some socioeconomic variables, such as income and marital status, on SAH?

Many socioeconomic studies use self-assessed health status as a proxy for true overall individual health status. Moreover, self-assessed health status has been shown to be a good predictor of mortality and demand for medical care (see, for example, van Doorslaer *et al*., 2004). Self-assessed health status, like other health outcome variables, exhibits a high degree of persistence. Knowing the true magnitude of the state dependence effect as a source of that persistence is important because state dependence determines the long-run effect of a policy that affects current health status. Moreover, only a flexible account of permanent unobserved heterogeneous factors that determine the self-assessed health level reported by each individual will allow us to obtain good estimates of the state dependence effect. Additionally, a proper modeling of the relationship between unobserved factors and socioeconomic variables is required to make correct inferences about the effect of those variables on self-assessed health status.

Contoyannis *et al*. (2004) have estimated a random-effects dynamic ordered probit, controlling for unobserved heterogeneity only in the level equation. Halliday (2008) has studied the relative contribution of state dependence and unobserved heterogeneity in SAH, reducing the model to a binary outcome model and taking a different random-effects approach. Halliday's approach is potentially more flexible and less

* Correspondence to: Jesus M. Carro, Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Gatafe, Spain. E-mail: jcarro@eco.uc3m.es

parametric than that of Contoyannis and co-authors. Halliday included only age as a covariate because he focused on studying the evolution of SAH over the life cycle, finding evidence of a substantial amount of unobserved heterogeneity in health. Furthermore, as Halliday has commented, a larger amount of heterogeneity than he allows for might exist.[1] In addition, both of these papers have had to deal with the 'initial conditions problem' that arises when taking a random-effects approach in dynamic models.

In a situation like this, we would prefer to take a fixed-effects approach to avoid imposing arbitrary restrictions on the distribution of the unobserved heterogeneity and its correlation with the observable variables, and to avoid the initial conditions problem. Despite these advantages, there have been few applications of nonlinear panel models with fixed effects in health economics, as noted in Jones (2009). A notable recent exception is Jones and Schurer (2011), which uses Chamberlain's conditional fixed-effects logit to study the gradient in health satisfaction with respect to income. Jones and Schurer (2011) conclude that the underlying assumptions of the statistical model of the unobserved heterogeneity matter for assessing the link between income and health. Therefore, this conclusion confirms the importance of estimating a model that makes no assumptions about the distribution of the heterogeneity. The main shortcomings of Chamberlain's conditional logit estimator are its loss of information, its difficulties in calculating marginal effects, and the fact that it does not include state dependence. In contrast, state dependence is an important policy parameter of interest in our paper. Also, ignoring state dependence may affect estimation of the effect of the observable socioeconomic variables, such as income.

Contoyannis *et al*. (2004) and Halliday (2008) used random effects, and Jones and Schurer (2011) used a special fixed-effects estimator (rather than, for example, an ordered probit model with fixed effects) because of the known problems in estimating general nonlinear panel data models with fixed effects using the panel datasets available. This estimation problem is known as the incidental parameters problem, and it results in large finite-sample biases of the maximum likelihood estimation (MLE) when using panel data where T is not very large. This problem is more severe in a model like ours that is dynamic and contains more than one fixed effect. Implementing a solution to this problem in the estimation of dynamic ordered-choice models with fixed effects is one of the main contributions of our paper.

An important area of research in microeconometrics has been concerned with solving the incidental parameters problem by developing bias reduction methods. Some examples are Hahn and Newey (2004), Hahn and Kuersteiner (2011), Arellano and Hahn (2006), Carro (2007), Fernandez-Val (2009) and Bester and Hansen (2009).[2] This fast-growing literature offers several bias correction methods that are potentially useful in estimating our model. Bester and Hansen (2009) included an application of their penalty function approach to a dynamic ordered probit model with two fixed effects. Thus the penalty function approach is directly applicable to our problem, whereas other estimators require some transformation to adapt them to our model. However, previous studies in this literature, including some of those mentioned in this paragraph, show simulation results for other models, such as a dynamic logit, which indicate that the penalty function approach proposed by Bester and Hansen (2009) is not the best estimator in terms of finite-sample performance. Our own simulations show that for our sample size the remaining bias is still significant when using Bester and Hansen's penalty function to estimate our model. Thus we must consider applying other proposed methods.

[1] Computational difficulties in Halliday's more flexible approach limit the amount of heterogeneity for which he can allow (see section 5.1.2 in Halliday, 2008).

[2] 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.

DOI: 10.1002/jae

In this paper, we derive explicit formulas of the modified MLE (MMLE) used in Carro (2007) for the dynamic ordered probit model under consideration. We evaluate its finite sample performance and compare it with the Bester and Hansen's penalty estimator.[3] The MMLE has better finite-sample properties for all sample sizes considered in the simulation and negligible biases for the sample size of our data. These Monte Carlo experiments are another contribution of this paper because, as Arellano and Hahn (2007) note, more research is needed to know 'how well each of the methods recently proposed work for other specific models and data sets of interest in applied econometrics'. Also, Greene and Henshen (2008) comment on the lack of studies concerning the applicability of the recent proposals for bias reduction estimators in binary-choice models to ordered-choice models.

The rest of the paper proceeds as follows. Section 2 presents our model of self-assessed health status and the data we use, and further explains the relation of this paper to a previous study that used random effects to analyze self-assessed health status in the same dataset. Section 3 presents the estimation problem and the method we propose. We also comment on other possible solutions from the nonlinear bias correction literature and use simulations to evaluate the finite-sample performance of different alternatives. These simulations justify the selection of MMLE as our estimator. Section 4 presents the estimation results. The estimates of our model and comparison with random-effects estimates show that there are important state dependence effects, and a statistically significant effect of income and other socioeconomic variables. The results also show that flexibly accounting for permanent unobserved heterogeneity matters. Our conclusions are provided in Section 5.

2. MODEL AND DATA

2.1. Empirical Model of Self-Assessed Health

We consider the following dynamic panel data ordered probit with fixed effects as a reduced-form model of self-assessed health status (SAH):

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

where x_{it} is a set of exogenous variables that influence SAH, ε_{it} is a time and individually varying error term that is assumed to be $\varepsilon_{it} \underset{\text{i.i.d.}}{\sim} N(0,1)$, and h^*_{it} is latent health. The reported SAH (h_{it}), which is what we observe, is determined according to the following thresholds:

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

where $h_{it} = -1$ 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; these account for permanent unobserved heterogeneity, both in unobserved factors affecting health (index shifts) and in reporting behavior (cut-point shifts) in an unrestricted way. An example of index shifts are genetic traits. Cut-point shifts occur if individuals use different thresholds to assess their health and report different values of SAH even though they have the same level of true health.[4] Note that in addition to the usual scale normalization in discrete-choice models (i.e. restricting the variance of ε_{it} to equal one), we are also normalizing one of the two cut-points to zero. The somewhat more conventional normalization of setting the intercept in the linear index equal

[3] The MMLE is obtained from modifying the score of the MLE so that the order of the bias in T is reduced.
[4] See Lindeboom and van Doorslaer (2004) for a test that shows evidence of existence of these two different types of shifts.

DOI: 10.1002/jae

to zero is not available to us because the distribution of the intercept, including its mean, is unrestricted in the fixed-effects approach. An alternative normalization would be to have the two fixed effects in the two cut-points and leave the linear index equation without an intercept.

As this discussion on normalization shows, it is not possible to separately identify individual effects that impact only h_{it}^* (index shifts) from those that impact the cut-points. Therefore, although we control for the two mentioned sources of unobserved heterogeneity, we cannot separate them. Additionally, having only the fixed effect in the linear index (α_i) would also account for heterogeneity in the cut-points, but in a very restrictive way. Specifically, by introducing only one individual effect (α_i), we would be assuming that both sources of unobserved heterogeneity must have effects in opposite directions in $Pr(h_{it}=1)$ and $Pr(h_{it}=-1)$; furthermore, we would be restricting how these two effects differ in magnitude for all individuals. We do not have evidence in favor of these assumptions. Moreover, given the different sources of the unobserved heterogeneity and the potential relations among them and the observable variables, it is likely that these assumptions are too restrictive and lead to incorrect inferences. In contrast, by having two fixed effects in (2) we do not impose any restrictions on the cut-point shifts, nor on the index shift. This constitutes an important divergence from previous studies, such as Contoyannis *et al.* (2004).

In addition to the parameters that capture the effect of heterogeneity, β captures the effect of exogenous variables, and ρ_1 and ρ_{-1} are the parameters that allow for state dependence in this model. Determining the relative importance of state dependence versus permanent unobserved heterogeneity as alternative sources of persistence is crucial because each has a very different implication. State dependence may arise for structural reasons, such as differing abilities to deal with new health shocks depending on a previous health status or willingness to invest in health, which changes as the health status evolves. For example, people may be less prone to invest in their health after a health shock that lowers their returns to that investment. In any case, as in labor force participation, regardless of the underlying source, state dependence gives the long-run effect of a policy affecting health status today. This is why it is so useful to know the magnitude of the state dependence.

2.2. Data and x Variables

This study uses the British Household Panel Survey (BHPS), a longitudinal survey of private households in Great Britain. It was designed as an annual survey of each adult (16+) member of a representative sample of more than 5000 households, with approximately 10,000 individual interviews. The same individuals are re-interviewed in successive waves; if they leave their original households, they are re-interviewed along with all adult members of their new households. Similarly, new adult members joining the sample households and children who have reached the age of 16 become eligible for the interview process. We use 16 waves of data (years 1991–2006) and include only individuals who gave a full interview. An unbalanced panel of individuals who were interviewed in at least eight subsequent waves is used. Our sample consists of 76,128 observations from 6375 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?’ In wave 9, the SAH question and categories were reworded. This makes comparison with other waves difficult and wave 9 is not used in our empirical analysis.

The original five SAH categories are collapsed to a three-category variable, creating a new SAH variable that is 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 variables x that we use are: three dummy variables representing marital status (Married, Widowed, Divorced/Separated), with Single as the reference category, the size of the

Table I. 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	85.91	11.84	2.25	100
in	Fair	43.22	45.18	11.59	100
t-1	Poor or very poor	17.66	31.60	50.74	100
	Proportion	72.80	19.67	7.53	100

household (the number of people living in the same household), the number of children in the household, household income, year dummies (excluding the necessary number to avoid prefect collinearity), and a quadratic function of age. The question about SAH that we use to construct our dependent variable asks respondents to compare their health with people their own age. However, SAH becomes worse over time in the raw sample data, perhaps indicating that the age effect over health is not totally discounted by respondents. This can be seen in Table A.2 in the online Appendix.[5] For this reason, we include age as an explanatory variable. The income variable is the logarithm of equivalized real income, adjusted using the retail price index, and equivalized by McClement's scale to adjust for household size and composition. This income consists of the sum of non-labor and labor income in the reference year.

Variables that are time-constant and specific for individuals, such as the level of education or gender, are not included in the set of explanatory variables because they cannot be separately identified from permanent unobserved heterogeneity.[6] Fixed effects account for these variables as well as for unobserved characteristics, and we cannot separate their effects. This is sometimes seen as a drawback of the fixed-effects approach. However, the random-effects approach separately identifies the effect of these variables only because of the unrealistic assumption that unobserved characteristics are independent from them (for example, that unobserved healthy lifestyle is independent of education). Even with a correlated random-effects approach, if correlation is allowed in a Mundlak (1978) and Chamberlain (1984) style and initial conditions are controlled for following 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 without making further assumptions. Contoyannis *et al*. (2004) follow Wooldridge's (2005) proposal, and they comment about this impossibility of separating the effect of variables, such as education, from the effects of the unobservable variables that are correlated with them.

The tables in section A of the online Appendix (supporting information) and Table 1 here contain some descriptive statistics of 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. Supporting information Table A.2 presents the variation of SAH across different characteristics and health variables. For example, married or single people respond in the 'excellent' or 'good' health category more frequently than widows or divorced people. There is high persistence in SAH reported as seen in Table 1, which shows the transition probabilities. In this table, the largest numbers are on the diagonal for all three values of SAH_{t-1}.

2.3. Relation to Contoyannis *et al*. (2004)

There is a clear connection between this paper and Contoyannis *et al*. (2004): both papers use the British Household Panel Survey to study the dynamics of SAH. Nevertheless, there are several aspects

[5] See Contoyannis *et al*. (2004) for further discussion on this point.

[6] They are, however, included in the random-effects estimation we make for comparison.

DOI: 10.1002/jae

considered in Contoyannis *et al*. (2004) that are not studied here. In particular, their paper contains a more detailed data description than in our paper, and further discussion of the estimated model; it also addresses other issues, such as sample attrition, that are not considered here.[7] However, our paper complements and adds to Contoyannis *et al*. (2004) in various ways.

First, we use more periods from the BHPS than they do. They only use the first eight waves because the ninth contains a different question and categorization of SAH. While we also drop the 9th wave, we incorporate waves after wave 9 in our estimation. Because 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 is the case for the first wave. We treat the 10th wave like an initial observation and condition it out in our likelihood, leaving the probability for that observation totally unrestricted. In this model that has covariates X, Contoyannis *et al*. (2004) cannot do this because of their method of solving the initial conditions problem and their use of random effects.

Second, Contoyannis *et al*. (2004) impose homogeneous cut-points, whereas we have two individual specific effects: one in the linear index and one in the cut-points. Although we cannot separately identify both sources of unobserved heterogeneity, our approach is robust to heterogeneous cut-points freely correlated with any determinant of SAH.

Finally, we use fixed effects instead of a random-effects approach. The advantages of this are that no arbitrary restriction is imposed on the correlation between permanent unobserved heterogeneity and the observable variables, and there is no initial conditions problem.

To make an assessment of the contributions of this paper with respect to the previous literature, we also estimate our models using the same type of specification and estimation method as Contoyannis *et al*. (2004). Thus we also estimate (2) using a correlated random-effects specification with only an individual effect in the linear index equation (the α_i parameter in (1)), but with homogeneous cut-points. Therefore, in this correlated random effects specification:

$$h_{it} = \begin{cases} -1 & \text{if} \quad h_{it}^* < c_1 \\ 0 & \text{if} \quad c_1 < h_{it}^* \leq c_2 \\ 1 & \text{if} \quad h_{it}^* > c_2 \end{cases} \tag{3}$$

where c_1 and c_2 are (homogeneous) parameters to be estimated, h_{it}^* is defined in (1), and α_i in (1) is assumed to be

$$\alpha_i = \gamma_0 + \gamma'_1 h_{i0} + \gamma'_2 \bar{x}_i + u_i \tag{4}$$

where $\bar{x}_i$ is the average over the sample period of the exogenous variables, and $u_i \underset{\text{i.i.d.}}{\sim} N\left(0, \sigma_u^2\right)$ independently of everything else. h_{i0} is in (4) to address the initial condition problem following Wooldridge (2005). In this specification, it is not possible to separately identify an intercept in the linear index and the cut-points. As a result of that, we have adopted the conventional normalization that sets the intercept in the linear index equal to zero.

[7] An unbalanced panel (with random attrition) in a dynamic panel model does not pose any complications to a fixed-effect estimator (as opposed to a random-effects estimator), as long as it does not imply many individuals with a very small number of periods; and in our sample all observations have at least eight periods. However, the assumption of attrition at random seems unrealistic. Contoyannis *et al*. (2004) made a test and found evidence of non-random attrition, but they also found that the bias this may be causing to the estimates is negligible. Given this result based on the same dataset as ours, to avoid distraction from the main theme of this paper we do not consider non-random attrition here.

3. ESTIMATION METHOD

3.1. Estimation Problem and Possible Solutions

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

$$\Pr(h_{it}=-1|x_{it},h_{it-1},c_i,\alpha_i)=1-\Phi(c_i+\mu_{it}) \tag{5}$$

$$\Pr(h_{it}=0|x_{it},h_{it-1},c_i,\alpha_i)=\Phi(c_i+\mu_{it})-\Phi(\mu_{it}) \tag{6}$$

$$\Pr(h_{it}=1|x_{it},h_{it-1},c_i,\alpha_i)=1-\Pr(h_{it}=-1|.)-\Pr(h_{it}=0|.)=\Phi(\mu_{it}) \tag{7}$$

where

$$\mu_{it}=\alpha_i+\rho_1 1\left(h_{i,t-1}=1\right)+\rho_{-1}1\left(h_{i,t-1}=-1\right)+x'_{it}\beta \tag{8}$$

Conditioning on the first observation h_{i0}, and taking into account that, as explained in Section 2.3, we do not observe SAH at the 9th wave ($t=8$), the log-likelihood is

$$\begin{aligned} l(\rho_1,\rho_{-1},\beta,\alpha,\mathbf{c}) = &\sum_{i=1}^{N}\Big(\sum_{t=1}^{7}\left[\sum_{d=-1}^{1}1\{h_{it}=d\}\log\Pr(h_{it}=d|x_{it},h_{it-1},c_i,\alpha_i)\right] \\ &+\sum_{t=10}^{15}\left[\sum_{d=-1}^{1}1\{h_{it}=d\}\log\Pr(h_{it}=d|x_{it},h_{it-1},c_i,\alpha_i)\right]\Big) \end{aligned} \tag{9}$$

where $Pr(h_{it}=d|x_{it},h_{it-1},c_i,\alpha_i)$ is defined in equations (5)–(8) for $d=-1,0,1$.

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

These methods can be grouped into three approaches, based on the object corrected. The first approach constructs an analytical or numerical bias correction in a fixed-effect estimator. Fernandez-Val (2009), among others, takes this approach and applies his analytical bias correction to dynamic binary-choice models. The second approach is to correct the bias in moment equations. An example of this is Carro (2007), who uses an estimator of this type to correct the bias in dynamic binary-choice models. The third group is that which corrects 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. The HS penalty estimator studied in Bester and Hansen (2009) is the first option we consider because our model is also a dynamic ordered probit and because alternative approaches require transformations or derivations. This estimator also has the advantage of being easier to compute than the MMLE in Carro (2007) and the bias correction in Fernandez-Val (2009) because, unlike the other two, the HS does not require the calculation of expectations. This advantage is more relevant in our case because it has two fixed effects.

DOI: 10.1002/jae

Arellano and Hahn (2007) show how the different approaches are related. Asymptotically, all the approaches always reduce the order of the bias of the MLE from the standard $O(T^{-1})$ to $O(T^{-2})$ in the general classes of models for which 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) as well as in Bester and Hansen (2009) illustrates this point. Consider the model in which $y_{it} \underset{\text{i.i.d.}}{\sim} N\left(\eta_i, \sigma_0^2\right)$. The ML estimator of σ_0^2 is $\widehat{\sigma}^2_{\mathbf{MLE}} = \frac{1}{NT}\sum_i\sum_t \left(y_{it} - \widehat{\eta}_i\right)^2$. It is well known that $\widehat{\sigma}^2_{\mathbf{MLE}}$ is not a consistent estimator of σ_0^2 when $N \to \infty$ with fixed T because it converges to $\frac{T-1}{T}\sigma_0^2$. In this case the problem is easily remedied. $\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \widehat{\eta}_i\right)^2$ is the fixed T consistent estimator of σ_0^2. The MMLE from Carro (2007) produces exactly this estimator, correcting not only the $O(T^{-1})$ term of the bias but also all the asymptotic bias in this simple example. The HS removes the $O(T^{-1})$ term of the bias, but it does not attain the fixed-T consistent estimator. Fernandez-Val's (2009) one-step bias correction to the ML estimator does not produce a fixed-T consistent estimator either, but its iterated form does. Thus differences may appear between these different approaches when they are applied to specific models.

On the other hand, the incidental parameters problem can be seen as a finite-sample bias problem in the context of panel data. The problem is not important when T is large relative to N. However, because our panel does not have a large number of periods it is reasonable to doubt that the good asymptotic properties of the MLE when T goes to infinity (sufficiently fast) are a good approximation to our finite sample. Simulations show that we would need panels with many more time periods than are usually found in practice. The relevant implication is that we have to examine the finite-sample performance of the estimators for our model and sample size. In the methods considered here, this 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 require some derivation to obtain the specific correction for this case. However, they do this type of comparison using binary choice (probit and logit) models. Additionally, Carro (2007) and Fernandez-Val (2009) conduct 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 those models. From these comparisons, we conclude that the HS penalty approach is not the best choice, and significant biases can be found for sample sizes with T smaller than 13. Given this result, we consider another of the proposed methods to estimate our ordered probit and evaluate its finite sample properties. Fernandez-Val's (2009) and Carro's (2007) corrections are interesting candidates because they are equally superior to other alternatives in terms of finite-sample performance in the relevant existing comparisons. In the following subsections, 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.

3.2. MMLE for a Dynamic Ordered Probit with Two Fixed Effects

The model we want to estimate is defined in (1) and (2), and its log-likelihood is (9). Let $\gamma = (\beta, \rho_1, \rho_{-1})$ and $\eta_i = (\alpha_i, c_i)$. Partial derivatives are denoted by the letter d, so the first-order conditions are $\mathbf{d}_{\eta i}(\gamma, \eta_i) \equiv \frac{\partial l_i(\gamma,\eta_i)}{\partial \eta_i}$ 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 $\mathbf{d}_{\eta i}(\gamma, \eta_i) = 0$. The MLE of γ is obtained by maximizing the concentrated log-likelihood $\left(\sum_{i=1}^{N} l_i(\gamma, \eta_i(\gamma))\right)$, i.e. by solving the following first-order condition:

DOI: 10.1002/jae

$$\frac{1}{TN}\sum_{i=1}^{N}\mathbf{d}_{\gamma i}(\gamma,\eta_i(\gamma))=0 \tag{10}$$

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

To reduce the bias of the estimation, we follow Carro (2007) in modifying the score of the concentrated log-likelihood by adding a term that removes the first-order term of the asymptotic bias in T. By doing so, the MMLE of the γ parameters of model (2) is the value that solves the following score equation:

$$\begin{aligned}\mathbf{d}_{\gamma Mi}(\gamma)=\mathbf{d}_{\gamma i}(\gamma,\eta_i(\gamma))-\frac{1}{2}\frac{1}{d_{\alpha\alpha i}d_{cci}-d_{\alpha ci}{}^2}\Bigg[&d_{\alpha\alpha i}\left(\mathbf{d}_{\gamma cci}+d_{\alpha cci}\frac{\partial\widehat{\alpha}_i}{\partial\gamma}+d_{ccci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)\\&+d_{cci}\left(\mathbf{d}_{\gamma\alpha\alpha i}+d_{\alpha\alpha\alpha i}\frac{\partial\widehat{\alpha}_i}{\partial\gamma}+d_{\alpha\alpha ci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)-2d_{\alpha ci}\left(\mathbf{d}_{\gamma\alpha ci}+d_{\alpha\alpha ci}\frac{\partial\widehat{\alpha}_i}{\partial\gamma}+d_{\alpha cci}\frac{\partial\hat{c}_i}{\partial\gamma}\right)\Bigg]\\&-\frac{\partial}{\partial\alpha_i}\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{\alpha ci})-E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})}{E(d_{\alpha\alpha i})E(d_{cci})-[E(d_{\alpha ci})]^2}\right)\Big|_{\eta_i=\eta_i(\gamma)}\\&-\frac{\partial}{\partial c_i}\left(\frac{E(\mathbf{d}_{\gamma\alpha i})E(d_{\alpha ci})-E(d_{\alpha\alpha i})E(\mathbf{d}_{\gamma ci})}{E(d_{\alpha\alpha i})E(d_{cci})-[E(d_{\alpha ci})]^2}\right)\Big|_{\eta_i=\eta_i(\gamma)}=0\end{aligned} \tag{11}$$

where $\mathbf{d}_{\gamma i}(\gamma,\eta_i(\gamma))$ is the standard first-order condition from the concentrated log-likelihood, as in (10). $\mathbf{d}_{\gamma ci}=\frac{\partial^2 l_i}{\partial\gamma\partial c_i}$, $d_{\alpha\alpha i}=\frac{\partial^2 l_i}{\partial\alpha_i^2}$, $\mathbf{d}_{\gamma\alpha ci}=\frac{\partial^3 l_i}{\partial\gamma\partial c_i\partial\alpha_i}$, and so on. From the first order conditions of α_i and c_i we obtain $\hat{\alpha}_i(\gamma)$ and $\hat{c}_i(\gamma)$ in order to concentrate the log-likelihood. All expectations are conditional on the same set of information as the likelihood. These expectations can be computed by recursively conditioning, as we do to write the conditional likelihood. The parametric model (equations (1), (2) and the assumption about ε_{it}) from which we write the likelihood also gives the parametric form of the expectations we need to calculate.[8]

In the Appendix, we show how this modification on the score of the concentrated log-likelihood in (11) is a first-order adjustment on the asymptotic bias of the ML score; thus the first-order condition is more nearly unbiased and the order of the bias of the estimator is reduced from $O(T^{-1})$ to $O(T^{-2})$. Furthermore, the bias is corrected without changing the asymptotic variance of the MLE.

3.3. Simulations

3.3.1. *First DGP: Performance for Different T* and Degrees of Persistence

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

$$\alpha_i=\frac{1}{2}\sum_{t=1}^{4}x_{it}+u_i,\quad \text{where } u_i\sim N(x_{i0},1) \tag{12}$$

$$c_i=|z_i|,\quad \text{where } z_i\sim N(x_{i0},1) \tag{13}$$

so that they are correlated with the explanatory variables. When the unobserved heterogeneity and the covariates are correlated the problem becomes more severe than when they are independent. We study

[8] Section B in the online Appendix gives an explanation for computing the MMLE.

DOI: 10.1002/jae

Table II. Monte Carlo results: dynamic ordered 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	−0.217	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

Note: See a detailed description of the simulated model and other characteristics of the DGP in Section 3.3.1.

the performance of estimators under this condition because we consider it to be more realistic.[9] x_{it} follows a Gaussian AR(1) with autoregressive parameter equal to 0.5. Initial conditions are $x_{i0} \sim N(0, 1)$ and $h_{i0}^* = \alpha_i + \beta_0 x_{i0} + \varepsilon_{i0}$. We perform 1000 replications, with a population of $N=250$ individuals. For each simulation we estimate the MLE, the MMLE given by equation (11) and the HS estimator defined in Bester and Hansen (2009). The HS estimator is the value of the parameters that maximize the following penalized objective function:

$$\sum_{i=1}^{N} lk_i(\beta, \rho_1, \rho_{-1}, \alpha_i, c_i) - \sum_{i=1}^{N} \frac{1}{2} \text{trace}\left(\widehat{I}_{\alpha c_i}^{-1} \widehat{V}_{\alpha c_i}\right) - \frac{k}{2} \tag{14}$$

where lk_i is the log-likelihood of i, $\hat{I}_{\alpha c_i}$ is the sample information matrix for $e_i = (\alpha_i, c_i)'$, $\widehat{V}_{\alpha c_i}$ is a robust estimator of $\text{var}\left(\frac{1}{\sqrt{T}} \frac{\partial l_i}{\partial e_i}\right)$, and $k = dim(e_i)$. This penalty term is easier to calculate than the modification of the score in (11) because the penalty term does not involve any expectation.

Results from this experiment for different T are reported in Table 2, which shows the mean bias and the root mean squared error (RMSE). We find that for all T the MMLE performs better than the other two estimators. Comparing it with the HS, the differences are greater for $T=4$ and $T=8$, where the HS

[9] In the simulations of an ordered probit in Bester and Hansen (2009), the fixed effects are independent of the covariates. We have simulated and compared MMLE and HS in this case as well. The bias is smaller for all T values, but the conclusions from the comparison between MMLE and HS are the same as in the dependency case. Because the latter is more relevant in practice, we do not report the independency case.

is closer to the MLE than to the MMLE. When using the MMLE, the bias is smaller than 10% of the true values with $T=10$ for all but one of the ρ parameters. With $T=12$, the MMLE has negligible biases for all the parameters, whereas the HS contains biases and RMSEs larger than the MMLE with $T=10$. Even with $T=16$, the HS exhibits mean biases greater than the MMLE with $T=10$. It is not until $T=20$ that the HS has small biases and small RMSEs. Therefore, the HS requires more periods (at least more than 16) to have small finite-sample biases. Given this and the fact that the sample sizes we have in our empirical analysis are smaller than $T=14$, we use the MMLE.

The reasons for the MMLE's better performance is the use of the specific structure of the model when calculating the modification term. This structure translates into the expectations in the modification term. The likelihood includes the fact that we know the distribution of one of the explanatory variables: the lag of the dependent variable. Therefore, we write the likelihood for each period (conditional on the previous period) up to the likelihood of the initial condition, in a recursive manner. This is used in the modification, so it includes expectations, using the known distribution of h_{it-1} conditional on h_{it-2}. The HS is generally written so that it does not make any intensive use of a specific likelihood and thus it does not include such expectations. Therefore, the HS does not exploit all the information that our specification provides and it requires more periods to attain the same performance as the MMLE, thus confirming the idea expressed in Bester and Hansen (2009) that the simplicity of the HS (because it does not calculate expectations) might come at a cost, leading to a poorer performance than the other approaches.

Quality of inference. We consider the quality of inference on finite samples, based on these estimators. Table 3 presents the coverage of 95% confidence intervals, and the average estimated asymptotic standard errors divided by the standard deviation calculated from the Monte Carlo simulations. The latter ratio is very close to 1 in all cases for the MMLE and in most cases for the other estimators, which indicates that we have good estimates for the variance and the problem lies in the bias. This corresponds with the fact that we are correcting a bias without altering the asymptotic variance.

Table III. Monte Carlo results: inference over dynamic order probit parameters: conference intervals coverage and estimation of the standard errors

Parameter:	β		ρ_1		ρ_{-1}	
True value:	1		0.5		−0.5	
Estimator	% Coverage CI 95%	SE/SD	% Coverage CI 95%	SE/SD	% Coverage CI 95%	SE/SD
$T=8$						
MLE	0%	0.85	47%	0.87	48%	0.90
HS	0%	0.86	74%	0.91	73%	0.94
MMLE	64%	1.02	87%	0.93	85%	0.96
$T=10$						
MLE	0%	0.81	54%	0.91	53%	0.91
HS	3.5%	0.83	82%	0.96	78%	0.95
MMLE	74%	0.94	90%	0.96	89%	0.96
$T=12$						
MLE	0%	0.89	58%	0.91	62%	0.93
HS	8.8%	0.92	85%	0.96	83%	0.98
MMLE	81%	1.00	92%	0.95	92%	0.97
$T=20$						
MLE	2%	0.90	77%	0.96	73%	0.94
HS	48%	0.93	91%	1	88%	0.98
MMLE	90%	0.97	95%	0.98	93%	0.95

Note: This is for the simulation experiment in Table 2. We have used the inverse of the Hessian as estimator of variance. SE/SD is the average estimated asymptotic standard error divided by standard deviation calculated from the Monte Carlo simulations.

DOI: 10.1002/jae

In terms of inference, the coverage of the confidence intervals is extremely poor for the MLE, specifically for β. Even with $T=20$, the coverage for β is smaller than 3%. The HS estimator improves inference with respect to the MLE, but it remains far from the theoretical coverage of 95%; the coverage for β is particularly bad even with $T=20$. Therefore, also in terms of inference the MMLE is clearly the best estimator of these three for doing inference, for all periods and parameters.

Performance for different degrees of persistence. To check whether results are maintained under different scenarios of state dependence, we present simulations for different values of ρ_1 and ρ_{-1}, with $T=10$ in the online Appendix. The DGP is the same as that of Table 2 except for the values of ρ_1 and ρ_{-1}.Here the state dependence changes from very negative to very positive, including the case with no state dependence. In terms of bias and RMSE, we find that the MMLE performs better than the other methods for all cases. In principle, having a negative state dependence may improve all the estimators because it induces higher variance in y_{it}. This is the case for the estimation of β, where the three estimation methods improve, but it is not the case for the estimation of ρ_1 and ρ_{-1}, where the MMLE improves but the MLE and HS perform worse than with positive state dependence.

*3.3.2. **Simulations Based on Real Data***

Finally, we perform a simulation based on the real data used in this paper. This will provide further evidence about the finite-sample performance of the MMLE and will provide increased robustness to our choice of estimator. The DGP takes the estimates obtained by MMLE and reported in Table 4 as the true model. It takes the real data for all the individuals used in that estimation and all the significant x variables, leaving out the time dummies. Therefore, in this DGP, x_{it} is a vector containing

Table IV. Estimates

	1	2	3
Variable	Pooled	Correlated random effects	MMLE
Health in t-1: good	0.6527*** (0.0185)	0.5028*** (0.0234)	0.4875*** (0.0186)
Health at t-1: poor	−0.4417*** (0.0233)	−0.3259*** (0.0343)	−0.4375*** (0.0242)
Age	0.0011 (0.0032)	0.0200 (0.0210)	0.0205 (0.0222)
Age squared	−0.0000 (0.0000)	−0.0007*** (0.0001)	−0.0005*** (0.0001)
Married	0.0344 (0.0286)	0.1722 (0.0752)	0.0749 (0.0606)
Separated/divorced	−0.0580 (0.0358)	0.0475 (0.1028)	0.0375 (0.0729)
Widowed	−0.0243 (0.0408)	0.3668** (0.1329)	0.0542 (0.0918)
Household size	−0.0782*** (0.0138)	−0.0112 (0.0189)	−0.0388** (0.0177)
Number of children	0.0647*** (0.0155)	0.0423 (0.0189)	0.0472** (0.0188)
Household income	0.0816*** (0.0122)	0.0188 (0.0191)	0.0396*** (0.0147)
Male	−0.0095 (0.0175)	0.0116 (0.0265)	
Non-white	−0.0890* (0.0467)	−0.1277* (0.0709)	
Higher/1st degree	0.1540*** (0.0345)	0.1563*** (0.0466)	
HND/A level	0.0810*** (0.0250)	0.0696* (0.1862)	
CSE/O level	0.0860*** (0.0225)	0.0923*** (0.0327)	
Cut-point 1	0.0192 (0.1233)	−0.0277*** (0.2265)	
Cut-point 2	1.0698*** (0.1235)	1.0528*** (0.2267)	
σ_u^2		0.0686	
Mean c_i			1.1323
Variance c_i			0.3277
Mean α_i			−0.0743
Variance α_i			0.6311
Correlation (α_i, c_i)			−0.3326
Akaike information criterion	38544.0	37334.3	37275.2

Note: Standard errors are reported in parentheses. Number of individuals used in estimation of all models is 1739. Estimates of year dummies in all models and within means of variables in random effects are not reported. Asterisks indicate significance at *10%; **5%; ***1%.

DOI: 10.1002/jae

observations of the following variables: age, squared age, household size, number of children, and income. The true values of the parameters are: $\rho_1 = 0.4875$, $\rho_{-1} = -0.4375$, $\beta' = (0.0205, -0.0005, -0.0388, 0.0472, 0.0396)$. $N = 1739$, T is the same as in our data (i.e. between 8 and 14 periods), and $\varepsilon_{it} \sim N(0, 1)$.

α_i and c_i are the estimates of these parameters by MMLE. The distributions of these two parameters are found in Figure 1. The distribution of α_i is not normal and is correlated with c_i (correlation coefficient between α_i and c_i is -0.33). Thus the distribution of unobserved heterogeneity is not an arbitrary and statistically convenient distribution, but an empirically founded distribution that captures both real correlations with the covariates and correlations between fixed effects. These correlations and distributions of α_i and c_i are richer than those in the previous simulation experiments. Furthermore, this is the relevant DGP to compare the proposed strategy for dealing with unobserved heterogeneity with the random-effects approach previously used in the literature. Making this comparison with an arbitrarily chosen DGP may imply a too favorable assumption to the random effects, as in our first DGP, or a too arbitrarily unfavorable one. However, this case is the relevant case for our empirical analysis.

For the reasons discussed above, we evaluate the finite-sample performance of the random-effects approach (CRE) described at the end of Section 2.3, in addition to the MLE, HS and MMLE. To make the comparison as close as possible with the estimators used with real data, we include the following constant variables as covariates when estimating by random effects: gender, race, and education indicators. These are implicitly included in the DGP through the estimated α_i and c_i, since in the fixed effects these variables cannot be separately identified from the fixed effects.

The results of this simulation are presented in the online Appendix. The MMLE is decidedly the best of all estimators in terms of RMSE. More specifically, the bias and RMSE for the CRE are twice the bias and RMSE of the MMLE for some parameters, such as ρ_1 and the β for household size. As in the previous simulation experiments with similar number of periods, the MMLE exhibit small biases.

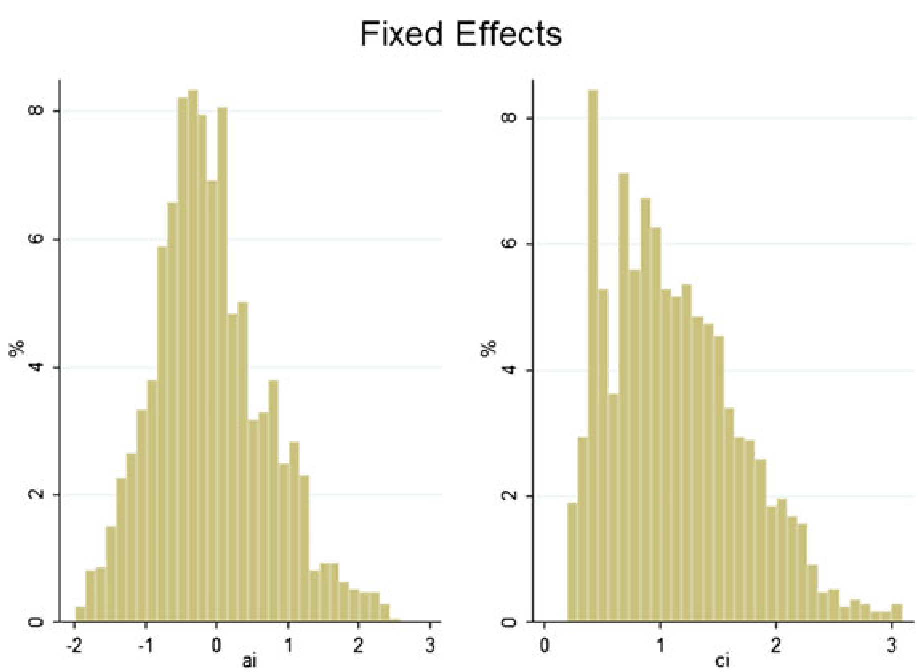

Figure 1. Density estimate (histogram) of the fixed effects from MMLE

DOI: 10.1002/jae

Table V. Average marginal effects on probability of reporting good and poor health for significant variables

(a) Good

	1		2		3	
			Correlated random			
	Pooled	SE	Effects	SE	MMLE	SE
Health in t-1: good	0.2528	0.0071	0.1883	0.0114	0.1653	0.0080
Health in t-1: poor	−0.1550	0.0078	−0.1149	0.0139	−0.1403	0.0520
Age	−0.0005	0.0003	−0.0170	0.0089	−0.0089	0.0064
Household size	−0.0282	0.0050	−0.0040	0.0112	−0.0120	0.0054
Number of children	0.0233	0.0056	0.0150	0.0141	0.0145	0.0058
Household income	0.0294	0.0044	0.0067	0.0094	0.0122	0.0045

(b) Poor

	1		2		3	
			Correlated random			
	Pooled	SE	Effects	SE	MMLE	SE
Health in t-1: good	−0.1399	0.0046	−0.1057	0.0206	−0.0984	0.1153
Health in t-1: poor	0.1477	0.0081	0.0968	0.0164	0.1268	0.0947
Age	0.0003	0.0002	0.0105	0.0060	0.0058	0.0117
Household size	0.0173	0.0031	0.0024	0.0069	0.0081	0.0086
Number of children	−0.0143	0.0034	−0.0090	0.0084	−0.0095	0.0102
Household income	−0.0181	0.0027	−0.0040	0.0058	−0.0081	0.0082

4. ESTIMATION RESULTS

Table 4 presents the coefficient estimates for our model based on three different estimators. This includes different specifications of the heterogeneity. The first estimated model (column 1) is a pooled version of the model in (1) and (2), without individual specific effects. The second estimated model (column 2) is the correlated random-effects model described in equations (3) and (4). It is similar to models estimated in Contoyannis *et al*. (2004). It has homogeneous cut-points and uses a random-effects approach to control for the individual specific intercept in the linear index. The last specification (column 3) is described in previous sections; it is the model in (1) and (2) treating α_i and c_i as fixed effects, and it is estimated by MMLE.

To compare magnitudes of the effects across variables and estimates we look at the relative effects (i.e. ratio of coefficients), and the average marginal effects reported in Table 5 for the variables with a coefficient significantly different from zero.[10,11]

The pooled model exacerbates the state dependence effect due to the lack of permanent unobserved heterogeneity. Although it is not reported, we also estimated the model in (1) and (2) by MLE.

[10] These marginal effects are also called partial effects. The marginal effects are averaged across the first eight waves of the panel, as well as across the values of the covariates for each individual. Thus we first calculate the marginal effect for each individual in the sample at the observed values of the regressors, and then we calculate their average, rather than calculating the marginal effect at the average value of the covariates. We do this to obtain summary measures of the marginal effects that are representative of the population's situation (see Chamberlain, 1984, p. 1273). Moreover, a measure that substitutes the values of the covariates, and especially the individual specific effect α_i, with their means (or any other fixed value) ignores any possible correlation between them.

[11] An alternative way to identify and estimate the marginal effects is the approach taken in Chernozhukov *et al*. (2010). They show that in a model like ours, with fixed effects, when T is fixed the (average and quantile) marginal effects are not point identified. However, they are set identified, and they propose a way to estimate bounds on the partial effect. These nonparametric bounds tighten as T grows. The main advantage is that the bounds analysis applies to any T, whereas our bias correction method depends on T not being very small. However, the bounds analysis is only available with discrete covariates for the moment. In contrast, bias correction methods work well in many examples, including with continuous covariates, and they consistently point estimate the identified average effect.

DOI: 10.1002/jae

As seen in the simulations, it is severely biased, estimating much lower state dependence effects and a higher effect for the other explanatory variables.

Of more interest is the comparison between the correlated random-effects model and the fixed-effects model estimated by MMLE. These estimates are in columns 2 and 3 of Tables 4 and 5, respectively. The first difference is in the variables that are statistically significant. Table 4 shows that in the MMLE household size, number of children, and household income have an impact that is statistically different from zero. However, none of them has a significant effect in the random-effect estimates. The average marginal effect of those variables correspondingly increases in absolute value in the MMLE case with respect to the random-effects model, especially for household income. Regarding the state dependence effect (effect of h_{it-1}), there are also changes. The effect of $h_{it-1}=$ good decreases in absolute value when estimating by MMLE, and the effect of $h_{it-1}=$ poor increases. Comparing coefficients in Table 4, we can also see that the effect of $h_{it-1}=$ poor increases proportionally less than the effect of the other relevant explanatory variables. In the random-effects specification the ratio of the coefficient of $1(h_{i,t-1}=\text{poor})$ to the coefficient of 'Household income' is approximately 17, whereas in the MMLE that ratio is 11. In any case, this partial increase in the effect of state dependence and of the effect of the explanatory variables is remarkable because the model in column 3 allows for more permanent unobserved heterogeneity and more flexibility than the model in column 2.[12]

Moreover, many of the differences in the estimated effects of the explanatory variables between the correlated random-effects model and the fixed-effects model estimated by MMLE are statistically significant. If the restrictions imposed by the correlated random-effects model are correct, its estimates are more precise (i.e. efficient) than the estimates of the fixed-effects model (even after the modification of the MLE), although both are consistent. Given this, we have used a Hausman-type test to determine whether those important differences are only because of the less precise estimates given in column 3. We have made the test over the average marginal effects instead of the parameters in Table 4 for two reasons. First, marginal effects (including their average), and not the parameters in equations (1) and (2), are usually the parameters of interest in nonlinear models. Second, the average marginal effects do not suffer the different scales problem that would prevent magnitudes in columns 2 and 3 of Table 4 from being directly comparable or directly interpretable. The average marginal effects of both models are well defined within the same scale, as any other marginal effect over choice probabilities, and their magnitude has the same clear interpretation. If we were primarily interested in a single average marginal effect, such as the effect of $h_{i,t-1}=$ good over the probability of $h_{i,t}=$ good, we could use a t-statistic that ignores the other effects. Doing this for all the average marginal effects, we reject at 5% the null hypothesis that both estimates are the same for four variables. Doing a joint test, we also reject the null hypothesis that the correlated random-effects estimates and the fixed-effects MMLE estimates are the same, thus rejecting the restrictions imposed in the correlated random-effects model.[13]

The previous two paragraphs are a clear indication that ignoring the added dimension of heterogeneity and flexibility in the distribution of the fixed effects matters for the estimation of both the model's parameter and the marginal effects of variables. It is not only the amount of heterogeneity, but also

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

[13] In the Hausman-type test we have used the variance–covariance matrix of the fixed-effects estimates only, instead of subtracting from it the variance of the random effects. We do this to avoid having the difference be a non-positive definite matrix because of the use of different estimates of the variance of the errors. Under correct specification, this represents a lower bound for this test and a rejection here will also be a rejection when using the well-defined difference in the variance–covariance matrices. A different solution to the non-positive definiteness problem is to use a score test that is asymptotically equivalent to the Hausman test. Doing such a score test we also reject the null hypothesis at any reasonable level of significance. See White (1982) and Ruud (1984) for further information on this score test.

Table VI. Marginal effects of income on probability of reporting good health by age and gender

	Age					
	<31	31–40	41–50	51–60	61–70	>70
Marginal effects at the average						
Female	0.0158***	0.0158***	0.0158***	0.0158***	0.0157**	0.0153
Male	0.0158***	0.0157***	0.0158***	0.0158***	0.0157**	0.0157**
Average marginal effects						
Female	0.0128***	0.0125***	0.0118***	0.0118***	0.0121***	0.0118***
Male	0.0130***	0.0120***	0.0122***	0.0120***	0.0117***	0.0116**

Note: Asterisks indicate significance at
*10%; **5%; ***1%.

the other restrictions being imposed on the model in column 2 that matters for estimation of the parameters of interest.

Aside from the formal test of random effects versus fixed effects, we look at the unobserved heterogeneity 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, and both exhibit large variation. The average for α_i is -0.074 and for c_i is 1.13. The standard deviations of these distributions are 0.79 and 0.57, respectively. In the random-effects specification, α_i is the compound equation (4) that includes a linear relation to some observables and an additive unobserved term that is assumed to follow a normal distribution. Given the estimates of the parameters of equation (4), the estimated average for α_i in the random-effects model is 1.41, and its standard deviation is 0.9626. With respect to the heterogeneity on the cut-points, the average of $-c_i$, the first cut-point, is -1.13 and the estimate of the first cut-point in the random effects specification is -0.03. Additionally, as shown in the right-hand panel of Figure 1, there is large variation in c_i among individuals that is ignored by the estimated random-effects model. Moreover, the normality of the distribution of α_i is rejected at 1%.[14] Finally, the correlation between α_i and c_i is -0.33; therefore, there are rich interactions between both fixed effects forming a joint distribution that is not the simple combination of their marginal distributions.

Focusing on the MML estimates, we find evidence of strong positive state dependence. With respect to socioeconomic variables, we find that aging and household size have a small but significant negative effect on SAH.[15] Number of children has a positive and significant average marginal effect, and it is the largest average effect in absolute value of all the x variables. Household income has the second largest effect among the x variables, and it is also a positive and significant average effect. Jones and Schurer (2011) focused on the gradient of health satisfaction with respect to income and did not include state dependence. We account for state dependence and it is interesting to determine whether that is also affecting the estimates of the effect of income. In Table 6 we show, for each age–gender group, the marginal effect of income at the average values of the explanatory variables and unobserved heterogeneity. That effect at the average is the marginal effect calculated in Jones and Schurer (2011). The age pattern is similar to that found in Figure 4 of Jones and Schurer (2011): it is positive, decreasing slightly with age, and significant for all age groups except the oldest group.

[14] c_i cannot be normal by definition because it is restricted to be positive.

[15] Halliday (2008) found, based on Akaike information criterion (AIC), that a quadratic function of age was only weakly preferred to the linear model and that no significant losses were found by using a linear model in age. We have estimated model 3 in Table 4 excluding age^2 as an explanatory variable, and in our case the fit is significantly worse because the effect of age increases more than linearly at older ages. Additionally, when introducing the quadratic term, the AIC changes to a greater degree than in Halliday (2008). Here, in the linear model AIC is 37373.4 and in the quadratic model it is 37275.2—nearly 100 points smaller.

Table VII. Sample versus predicted proportions of SAH (%)

Panel A: Total proportions

	Poor or very poor	Fair	Excellent or good
Sample	16	31	53
Predicted MMLE	15	32	53
Predicted CRE	12	31	57
Predicted pooled	13	28	60

Panel B: Proportions of people reporting excellent or good SAH

		Predicted		
	Sample	MMLE	CRE	Pooled
By number of children				
0	52	53	57	59
1	55	54	56	61
2	58	56	57	63
3+	50	51	54	60
By income quartiles				
1st quartile	47	50	53	56
2nd quartile	51	52	56	59
3rd quartile	56	55	58	62
4th quartile	58	57	59	62

However, significant differences arise when comparing the magnitude of the effect. While Jones and Schurer's estimated effect for men takes values between 2.5 and 5 percentage points for all but the oldest age group, our estimated effect at the average is never greater than 1.6 percentage points.[16] These much smaller marginal effects are in accordance with the intuition that the magnitude of the positive effect of a variable will probably tend to be overestimated when not accounting for relevant state dependence. Finally, we also calculate the average marginal effect for the same age–gender groups. We find a similar pattern, but even smaller point estimates. The difference between the marginal effect at the average and the average marginal effect comes mainly from the former, ignoring the correlation between explanatory variables and the unobserved heterogeneity.

With respect to how the models fit the observed data, in addition to the information criteria (AIC) reported in Table 4, some predictions of the estimated models and their sample counterparts are provided in Table 7. Overall, the MMLE model fits the data better because its predictions are closer to the actually observed proportions in the sample. In a similar manner, the MMLE predictions are better at capturing the inverted-U shape of the proportion of individuals reporting excellent or good health as we look at people with higher number of children. They are also better at capturing the slope in the increasing pattern when looking at people with higher incomes.[17]

In addition to considering the average marginal effects reported in Table 5, 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 significant (at 10%) marginal effects over the probability of reporting good and bad health, for the same variables as in Table 5. Notice that, although the average marginal effects are significant, there is a great deal of

[16] Caution should be taken with this comparison: the datasets are different; Jones and Schurer's specification differs in the way covariates enter; they use health satisfaction and not self-assessed health; and they have more categories than us in the outcome variable. In any case, given their categorization and ours, the closer comparison of our marginal effects estimates are with those in their Figure 4, and the economic interpretation of their numbers in Figure 4 and our numbers in Table 6 is meant to be the same.

[17] Note that we are not controlling for any other observable characteristics. Thus there may be other differences between people with a different number of children (or different incomes) that can reinforce or cancel the effect of number of children (or income) on average. Therefore these numbers cannot be interpreted as the effect of the number of children (nor the effect of income).

DOI: 10.1002/jae

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

	Proportion	
Variable	Good	Poor
Health in t-1: good	60.44%	12.25%
Health in t-1: poor	55.43%	34.50%
Age	22.71%	2.53%
Household size	37.21%	11.44%
Number of children	41.81%	12.65%
Household income	44.85%	15.35%

heterogeneity; for around half the population, the marginal effects over the probability of reporting good health is not significantly different from zero for many of these variables.

5. CONCLUSION

In this paper, we have considered the estimation of a dynamic ordered probit of self-assessed health status with two fixed effects: one in the linear index equation and one in the cut-points. The inclusion of two fixed effects, instead of only one as usual, is motivated by the potential existence of two sources of heterogeneity: unobserved health status and reporting behavior. Although we cannot separately identify these two sources of heterogeneity, we robustly control for them by using two fixed effects. Based on our best estimates, the two fixed effects exhibit important variation, and it is relevant to account for both when estimating the effect of other variables. Our estimates also show that state dependence is large and significant even after controlling for unobserved heterogeneity. By a comparison with previously used random-effects estimates, we show that flexibly accounting for more permanent unobserved heterogeneity is important.

The recent literature in bias-adjusted methods of estimation of nonlinear panel data models with fixed effects has produced several potentially equivalent estimators. We find that the a priori most directly applicable correction to our model, the HS estimator proposed in Bester and Hansen (2009), still has significant biases in our sample size. This finding led us to consider the modified MLE proposed in Carro (2007). We derive the expression of the MMLE for our model, conduct 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. The Monte Carlo experiments contribute to the literature on bias-adjusted methods for estimating nonlinear panel data models by showing how well two of the proposed methods work for a specific model and sample size. This information will be useful for other applications when choosing among several correction methods existing in the literature.

ACKNOWLEDGEMENTS

This is a revised version of a paper previously circulated under the title ’Correcting the bias in the estimation of a dynamic ordered probit with fixed effects of self-assessed health status’. We thank the editor and three anonymous referees for a number of comments that helped to improve the paper. We also thank Raquel Carrasco, Matilde Machado and seminar participants at Universidad Carlos III, Boston University, and the Lincoln College Applied Microeconometrics Conference for useful remarks 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, by grants ECO2009-11165 and SEJ2006-05710 from the Spanish Minister of Education, MCINN (Consolider: Ingenio2010) and Consejería de Educación de la Comunidad de Madrid (Excelecon project). Also, the first author would like to thank the Department of Economics at the MIT, and the second author would like to thank the Institute for Economic Development at Boston University, at which they conducted part of this research as visiting scholars.

REFERENCES

Arellano M, Hahn J. 2006. A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. Working paper, CEMFI, Madrid.

Arellano M, Hahn J. 2007. Understanding bias in nonlinear panel models: some recent developments. In Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Vol. **3**, Blundell R, Newey W, Persson T (eds). Cambridge University Press: Cambridge, UK; 381–409.

Bester CA, Hansen C. 2009. A penalty function approach to bias reduction in non-linear panel models with fixed effects. *Journal of Business and Economic Statistic* **27**(2): 131–148.

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

Chamberlain G. 1984. Panel data. In Handbook of Econometrics, Vol. **2**, Griliches Z, Intriligator MD (eds). Elsevier Science: Amsterdam; 1247–1318.

Chernozhukov V, Fernandez-Val I, Hahn J, Newey W. 2010. Average and quantile effects in nonseparable panel models. Mimeo, MIT Department of Economics, Cambridge, MA.

Contoyannis P, Jones AM, Rice N. 2004. The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics* **19**: 473–503.

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

Greene WH, Henshen DA. 2008. Modeling ordered choices: a primer and recent developments. Available: http://ssrn.com/abstract=1213093 [14 August 2012].

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

Hahn J, Newey W.2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* **72**(4): 1295–1319.

Halliday TJ. 2008. Heterogeneity, state dependence and health. *The Econometrics Journal* **11**: 499–516.

Jones AM. 2009. Panel data methods and applications to health economics. In The Palgrave Handbook of Econometrics: Applied Econometrics, Vol. **II**, Mills TC, Patterson K (eds). Palgrave MacMillan: Basingstoke 557–631.

Jones AM, Schurer S. 2011. How does heterogeneity shape the socioeconomic gradient in health satisfaction? *Journal of Applied Econometrics* **26**(4): 549–579.

Lindeboom M, van Doorslaer E. 2004. Cut-point shift and index shift in self-reported health. *Journal of Health Economics* **23**: 1083–1099.

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

Rilstone P, Srivastava VK, Ullah A. 1996. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* **75**: 369–395.

Ruud PA. 1984. Tests of specification in econometrics. *Econometrics Reviews* **3**(2): 211–242.

van Doorslaer E, Jones AM, Koolman X. 2004. Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics* **13**: 629–647.

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

Wooldridge J. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* **20**: 39–54.

DOI: 10.1002/jae

APPENDIX A: REDUCTION OF THE ORDER OF THE BIAS

In this Appendix we show that the modified score presented above corrects the first-order asymptotic bias of the original score. The algebra is somewhat tedious because of the many terms, but the idea is clear. We first expand the score of the MLE around the true value of the fixed effects and make some calculations and substitutions on it to obtain the leading term of the bias of the MLE's score. We then show that the modification in the MMLE's score, equation (11), is subtracting that leading bias term from the score. This follows Carro (2007), and is adapted to our model with two fixed effects.

The notation used is the same as in section 3.2: $\gamma=(\beta,\rho_1,\rho_{-1})$ and $\eta_i=(\alpha_i,c_i)$; we denote partial derivatives by the letter d; bold letters are used to denote vectors; $\mathbf{d}_{\eta i}\equiv\frac{\partial l_i(\gamma,\eta_i)}{\partial\eta_i}$, $\mathbf{d}_{\gamma i}\equiv\frac{\partial l_i(\gamma,\eta_i)}{\partial\gamma}$, $\mathbf{d}_{\gamma ci}=\frac{\partial^2 l_i}{\partial\gamma\partial c_i}$, $d_{\alpha\alpha i}=\frac{\partial^2 l_i}{\partial\alpha_i^2}$, $\mathbf{d}_{\gamma\alpha ci}=\frac{\partial^3 l_i}{\partial\gamma\partial c_i\partial\alpha_i}$, and so on; the derivatives evaluated at the true values of the parameters are represented by including a 0 in the sub-index (e.g. $d_{\eta i0}=d_{\eta i}(\gamma_0,\eta_{i0})$).

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

$$\begin{aligned}\mathbf{d}_{\gamma i}(\gamma_0,\eta_i(\gamma_0)) = {} & \mathbf{d}_{\gamma i0}+d_{\gamma\alpha i0}(\hat{\alpha}_i(\gamma_0)-\alpha_{i0})\\ & +\mathbf{d}_{\gamma ci0}(\hat{c}_i(\gamma_0)-c_{i0})\\ & +\frac{1}{2}\mathbf{d}_{\gamma\alpha\alpha i0}(\hat{\alpha}_i(\gamma_0)-\alpha_{i0})^2+\frac{1}{2}\mathbf{d}_{\gamma cci0}(\hat{c}_i(\gamma_0)-c_{i0})^2\\ & +\mathbf{d}_{\gamma aci0}(\hat{\alpha}_i(\gamma_0)-\alpha_{i0})(\hat{c}_i(\gamma_0)-c_{i0})+O_p\left(T^{-1/2}\right)+\dots\end{aligned} \tag{15}$$

This equation clearly shows that the score evaluated at the true value γ_0 differs from the value of the score we want to obtain, $\mathbf{d}_{\gamma i0}=\mathbf{d}_{\gamma i}(\gamma_0,\eta_{i0})$, as much as $\hat{\alpha}_i(\gamma_0)$ and $\hat{c}_i(\gamma_0)$ differ from α_{i0} and c_{i0}. This is the source of the incidental parameters problem.

Now we need expressions for $(\hat{\alpha}_i(\gamma_0)-\alpha_{i0})$ and $(\hat{c}_i(\gamma_0)-c_{i0})$, for which we do asymptotic expansions, following Rilstone *et al*. (1996):

$$(\hat{\alpha}_i(\gamma_0)-\alpha_{i0})=b^{\alpha}_{-1/2}+b^{\alpha}_{-1}+O_p\left(T^{-3/2}\right) \tag{16}$$

$$(\hat{c}_i(\gamma_0)-c_{i0})=b^{c}_{-1/2}+b^{c}_{-1}+O_p\left(T^{-3/2}\right) \tag{17}$$

where $b^{\alpha}_{-1/2}$ and $b^{c}_{-1/2}$ are the elements of the vector $\mathbf{b}_{-1/2}$, and b^{α}_{-1} and b^{c}_{-1} are the elements of the vector $\mathbf{b}_{-1}$, which are determined as follows:

$$\begin{aligned}\mathbf{b}_{-1/2} &= -Q^{-1}R\\ \mathbf{b}_{-1} &= -Q^{-1}S\mathbf{b}_{-1/2}-\frac{1}{2}Q^{-1}U\left(\mathbf{b}_{-1/2}\otimes\mathbf{b}_{-1/2}\right)\\ R &= \frac{1}{T}\begin{pmatrix} d_{\alpha i0}\\ d_{ci0}\end{pmatrix}\\ Q &= E(\nabla R)\\ S &= \nabla R-Q\\ U &= E(\nabla^2 Q)\end{aligned}$$

From the above expressions we obtain

DOI: 10.1002/jae

$$b^{\alpha}_{-1/2} = \frac{\frac{1}{T} d_{ci0} E\left(\frac{1}{T} d_{c\alpha i0}\right) - \frac{1}{T} d_{ai0} E\left(\frac{1}{T} d_{cci0}\right)}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} \tag{18}$$

$$b^{c}_{-1/2} = \frac{\frac{1}{T} d_{ai0} E\left(\frac{1}{T} d_{c\alpha i0}\right) - \frac{1}{T} d_{ci0} E\left(\frac{1}{T} d_{\alpha\alpha i0}\right)}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} \tag{19}$$

It is also useful to obtain

$$(\hat{\alpha}_i(\gamma_0) - \alpha_{i0})^2 = \left(b^{a}_{-1/2}\right)^2 + O_p\left(T^{-3/2}\right) \tag{20}$$

$$(\hat{c}_i(\gamma_0) - c_{i0})^2 = \left(b^{c}_{-1/2}\right)^2 + O_p\left(T^{-3/2}\right) \tag{21}$$

$$(\hat{\alpha}_i(\gamma_0) - \alpha_{i0})(\hat{c}_i(\gamma_0) - c_{i0}) = b^{a}_{-1/2} b^{c}_{-1/2} + O_p\left(T^{-3/2}\right) \tag{22}$$

With respect to the squares of $b^{\alpha}_{-1/2}$ and $b^{c}_{-1/2}$, we get

$$\left(b^{\alpha}_{-1/2}\right)^2 = \frac{\left(\frac{1}{T} d_{ai0}\right)^2 E\left(\frac{1}{T} d_{cci0}\right)^2 + \left(\frac{1}{T} d_{ci0}\right)^2 E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 - 2\frac{1}{T} d_{ai0} \frac{1}{T} d_{ci0} E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right)}{\left(E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2\right)^2}$$

$$\left(b^{c}_{-1/2}\right)^2 = \frac{\left(\frac{1}{T} d_{ci0}\right)^2 E\left(\frac{1}{T} d_{\alpha\alpha i0}\right)^2 + \left(\frac{1}{T} d_{ai0}\right)^2 E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 - 2\frac{1}{T} d_{ai0} \frac{1}{T} d_{ci0} E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{c\alpha i0}\right)}{\left(E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2\right)^2}$$

Substituting by expectations, and using the information matrix identity ($E(d_{c\alpha i}) = -E(d_{ai} d_{ci})$), we get

$$\left(b^{\alpha}_{-1/2}\right)^2 = -\frac{1}{T} \frac{E\left(\frac{1}{T} d_{cci0}\right)}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} + O_p\left(T^{-3/2}\right) \tag{23}$$

$$\left(b^{c}_{-1/2}\right)^2 = -\frac{1}{T} \frac{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right)}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} + O_p\left(T^{-3/2}\right) \tag{24}$$

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

$$b^{\alpha}_{-1/2} b^{c}_{-1/2} = \frac{1}{T} \frac{E\left(\frac{1}{T} d_{c\alpha i0}\right)}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} + O_p\left(T^{-3/2}\right) \tag{25}$$

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

DOI: 10.1002/jae

$$
\begin{aligned}
b_{-1}^{\alpha} = {} & \frac{1}{2T} \frac{1}{\left(E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 \right)^2} \\
& \Bigg\{ 2E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 \left(E\left(\frac{1}{T} d_{\alpha cci0}\right) + E\left(\frac{1}{T} d_{ai0} d_{cci0}\right) + E\left(\frac{1}{T} d_{ci0} d_{c\alpha i0}\right) \right) \\
& + E\left(\frac{1}{T} d_{cci0}\right)^2 \left[E\left(\frac{1}{T} d_{a\alpha\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{\alpha\alpha i0}\right) \right] \\
& + E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[E\left(\frac{1}{T} d_{\alpha cci0}\right) + 2E\left(\frac{1}{T} d_{ci0} d_{c\alpha i0}\right) \right] \\
& - E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) \left[E\left(\frac{1}{T} d_{ccci0}\right) + 2E\left(\frac{1}{T} d_{ci0} d_{cci0}\right) \right] \\
& - E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[3E\left(\frac{1}{T} d_{\alpha\alpha ci0}\right) + 4E\left(\frac{1}{T} d_{ai0} d_{c\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ci0} d_{\alpha\alpha i0}\right) \right] \Bigg\} \\
& + O_p\left(T^{-3/2}\right)
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
b_{-1}^{c} = {} & \frac{1}{2T} \frac{1}{\left(E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 \right)^2} \\
& \Bigg\{ 2E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 \left[E\left(\frac{1}{T} d_{\alpha\alpha ci0}\right) + E\left(\frac{1}{T} d_{ci0} d_{\alpha\alpha i0}\right) + E\left(\frac{1}{T} d_{ai0} d_{c\alpha i0}\right) \right] \\
& + E\left(\frac{1}{T} d_{\alpha\alpha i0}\right)^2 \left[E\left(\frac{1}{T} d_{ccci0}\right) + 2E\left(\frac{1}{T} d_{ci0} d_{cci0}\right) \right] \\
& + E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[E\left(\frac{1}{T} d_{\alpha\alpha ci0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{c\alpha i0}\right) \right] \\
& - E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[E\left(\frac{1}{T} d_{a\alpha\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{\alpha\alpha i0}\right) \right] \\
& - E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) \left[3E\left(\frac{1}{T} d_{\alpha cci0}\right) + 4E\left(\frac{1}{T} d_{ci0} d_{c\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{cci0}\right) \right] \Bigg\} \\
& + O_p\left(T^{-3/2}\right)
\end{aligned}
\tag{27}
$$

Introducing all these expressions in (15), and taking expectations, we get

$$
\begin{aligned}
E\left(d_{\gamma i}(\gamma_0, \hat{\eta}_i(\gamma_0))\right) = {} & \frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha i0} d_{ci0}\right) E\left(\frac{1}{T} d_{c\alpha i0}\right) - E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha i0} d_{ai0}\right) E\left(\frac{1}{T} d_{cci0}\right)}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} \\
& + \frac{1}{2} \frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ai0}\right)}{\left(E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2\right)^2} \left\{ 2E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 \left(E\left(\frac{1}{T} d_{\alpha cci0}\right) \right.\right. \\
& \left. + E\left(\frac{1}{T} d_{ai0} d_{cci0}\right) + E\left(\frac{1}{T} d_{ci0} d_{c\alpha i0}\right) \right) \\
& + E\left(\frac{1}{T} d_{cci0}\right)^2 \left[E\left(\frac{1}{T} d_{\alpha\alpha\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{\alpha\alpha i0}\right) \right] \\
& + E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[E\left(\frac{1}{T} d_{\alpha cci0}\right) + 2E\left(\frac{1}{T} d_{ci0} d_{c\alpha i0}\right) \right] \\
& - E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{\alpha\alpha i0}\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_{c\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[3E\left(\frac{1}{T} d_{\alpha\alpha ci0}\right) + 4E\left(\frac{1}{T} d_{ai0} d_{c\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ci0} d_{\alpha\alpha i0}\right) \right] \right\} \\
& + \frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0} d_{ai0}\right) E\left(\frac{1}{T} d_{c\alpha i0}\right) - E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0} d_{ci0}\right) E\left(\frac{1}{T} d_{\alpha\alpha i0}\right)}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} \\
& + \frac{1}{2} \frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)}{\left(E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2\right)^2} \\
& \left\{ 2E\left(\frac{1}{T} d_{c\alpha i0}\right)^2 \left[E\left(\frac{1}{T} d_{\alpha\alpha ci0}\right) + E\left(\frac{1}{T} d_{ci0} d_{\alpha\alpha i0}\right) + E\left(\frac{1}{T} d_{ai0} d_{c\alpha i0}\right) \right] \right. \\
& + E\left(\frac{1}{T} d_{\alpha\alpha i0}\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_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[E\left(\frac{1}{T} d_{\alpha\alpha ci0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{c\alpha i0}\right) \right] \\
& - E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) \left[E\left(\frac{1}{T} d_{\alpha\alpha\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{\alpha\alpha i0}\right) \right] \\
& \left. - E\left(\frac{1}{T} d_{c\alpha i0}\right) E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) \left[3E\left(\frac{1}{T} d_{\alpha cci0}\right) + 4E\left(\frac{1}{T} d_{ci0} d_{c\alpha i0}\right) + 2E\left(\frac{1}{T} d_{ai0} d_{cci0}\right) \right] \right\} \\
& + \frac{1}{E\left(\frac{1}{T} d_{\alpha\alpha i0}\right) E\left(\frac{1}{T} d_{cci0}\right) - E\left(\frac{1}{T} d_{c\alpha i0}\right)^2} \left[E\left(\frac{1}{T}\mathbf{d}_{\gamma aci0}\right) E\left(\frac{1}{T} d_{c\alpha i0}\right) \right. \\
& \left. - \frac{1}{2} E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha\alpha i0}\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_{\alpha\alpha i0}\right) \right] + O\left(T^{-1}\right)
\end{aligned}
\tag{28}
$$

DOI: 10.1002/jae

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.

Modified Score

The modified score in (11) can be decomposed into three terms, $\mathbf{d}_{\gamma Mi}(\gamma) = A + B + C$, such that

$$A = \mathbf{d}_{\gamma i}(\gamma, \eta_i(\gamma)) \tag{29}$$

$$\begin{aligned} B = -\frac{1}{2}\frac{1}{d_{\alpha\alpha i}d_{cci} - {d_{c\alpha i}}^2} & \\ \Bigg[d_{\alpha\alpha i}\left(\mathbf{d}_{\gamma cci} + d_{\alpha cci}\frac{\partial \hat{\alpha}_i}{\partial \gamma} + d_{ccci}\frac{\partial \hat{c}_i}{\partial \gamma} \right) & \\ + d_{cci}\left(\mathbf{d}_{\gamma\alpha\alpha i} + d_{\alpha\alpha\alpha i}\frac{\partial \hat{\alpha}_i}{\partial \gamma} + d_{\alpha\alpha ci}\frac{\partial \hat{c}_i}{\partial \gamma} \right) & \\ -2d_{c\alpha i}\left(\mathbf{d}_{\gamma aci} + d_{\alpha\alpha ci}\frac{\partial \hat{\alpha}_i}{\partial \gamma} + d_{\alpha cci}\frac{\partial \hat{c}_i}{\partial \gamma} \right) \Bigg] & \end{aligned} \tag{30}$$

$$\begin{aligned} C = & -\frac{\partial}{\partial a_i}\left(\frac{E(\mathbf{d}_{\gamma ci})E(d_{c\alpha i}) - E(d_{cci})E(\mathbf{d}_{\gamma\alpha i})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2} \right)\Bigg|_{\eta_i = \eta_i(\gamma)} \\ & -\frac{\partial}{\partial c_i}\left(\frac{E(\mathbf{d}_{\gamma\alpha i})E(d_{c\alpha i}) - E(d_{\alpha\alpha i})E(\mathbf{d}_{\gamma ci})}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2} \right)\Bigg|_{\eta_i = \eta_i(\gamma)} \end{aligned} \tag{31}$$

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

Part *B*. We first want to derive an expression for $\partial\hat{\alpha}_i/\partial\gamma$ 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 $\partial\hat{\alpha}_i/\partial\gamma$ and $\partial\hat{c}_i/\partial\gamma$ we get

$$\frac{\partial \hat{\alpha}_i(\gamma)}{\partial \gamma} = \frac{\mathbf{d}_{\gamma ci}d_{c\alpha i} - d_{cci}\mathbf{d}_{\gamma\alpha i}}{d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2} \tag{32}$$

$$\frac{\partial \hat{c}_i(\gamma)}{\partial \gamma} = \frac{\mathbf{d}_{\gamma\alpha i}d_{c\alpha i} - d_{\alpha\alpha i}\mathbf{d}_{\gamma ci}}{d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2} \tag{33}$$

evaluating at γ_0 and replacing by expectations:

$$\frac{\partial \hat{\alpha}_i(\gamma_0)}{\partial \gamma} = \frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}d_{c\alpha i0}\right) - E\left(\frac{1}{T}d_{cci0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha i0}\right)}{E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{cci0}\right) - E\left(\frac{1}{T}d_{c\alpha i0}\right)^2} + O_p\left(T^{-\frac{1}{2}}\right) \tag{34}$$

$$\frac{\partial \hat{c}_i(\gamma_0)}{\partial \gamma} = \frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha i0}\right)E\left(\frac{1}{T}d_{c\alpha i0}\right) - E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)}{E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{cci0}\right) - E\left(\frac{1}{T}d_{c\alpha i0}\right)^2} + O_p\left(T^{-\frac{1}{2}}\right) \tag{35}$$

Introducing in (31) and rearranging terms:

DOI: 10.1002/jae

$$
\begin{aligned}
B = & -\frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}d_{c\alpha i0}\right) - E\left(\frac{1}{T}d_{cci0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha i0}\right)}{E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{cci0}\right) - E\left(\frac{1}{T}d_{c\alpha i0}\right)^2} \\
& \frac{d_{\alpha\alpha i}d_{\alpha cci} + d_{cci}d_{\alpha\alpha\alpha i} - 2d_{c\alpha i}d_{\alpha\alpha ci}}{2(d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2)} \\
& -\frac{d_{\alpha\alpha i}d_{\alpha cci} + d_{cci}d_{\alpha\alpha\alpha i} - 2d_{c\alpha i}d_{\alpha\alpha ci}}{2(d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2)} O_p\left(T^{-1/2}\right) \\
& -\frac{E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha i0}\right)E\left(\frac{1}{T}d_{c\alpha i0}\right) - E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)}{E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{cci0}\right) - E\left(\frac{1}{T}d_{c\alpha i0}\right)^2} \\
& \frac{d_{cci}d_{\alpha\alpha ci} + d_{\alpha\alpha i}d_{ccci} - 2d_{c\alpha i}d_{\alpha cci}}{2(d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2)} \\
& -\frac{d_{cci}d_{\alpha\alpha ci} + d_{\alpha\alpha i}d_{ccci} - 2d_{c\alpha i}d_{\alpha cci}}{2(d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2)} O_p\left(T^{-1/2}\right) \\
& -\frac{d_{\alpha\alpha i}\mathbf{d}_{\gamma cci} + d_{cci}\mathbf{d}_{\gamma\alpha\alpha i} - 2d_{c\alpha i}\mathbf{d}_{\gamma aci}}{2(d_{\alpha\alpha i}d_{cci} - d_{c\alpha i}^2)}
\end{aligned} \tag{36}
$$

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:

$$
\begin{aligned}
B = & -\frac{1}{2}\frac{1}{\left(E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{cci0}\right) - E\left(\frac{1}{T}d_{c\alpha i0}\right)^2\right)^2} \\
& \Bigg\{\left[E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)E\left(\frac{1}{T}d_{c\alpha i0}\right) - E\left(\frac{1}{T}d_{cci0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma ai0}\right)\right] \\
& \left[E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{\alpha cci0}\right) + E\left(\frac{1}{T}d_{cci0}\right)E\left(\frac{1}{T}d_{a\alpha\alpha i0}\right) - 2E\left(\frac{1}{T}d_{c\alpha i0}\right)E\left(\frac{1}{T}d_{\alpha\alpha ci0}\right)\right] \\
& +\left[E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha i0}\right)E\left(\frac{1}{T}d_{c\alpha i0}\right) - E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma ci0}\right)\right] \\
& \left[E\left(\frac{1}{T}d_{cci0}\right)E\left(\frac{1}{T}d_{\alpha\alpha ci0}\right) + E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{ccci0}\right) - 2E\left(\frac{1}{T}d_{c\alpha i0}\right)E\left(\frac{1}{T}d_{\alpha cci0}\right)\right]\Bigg\} \\
& -\frac{1}{2}\frac{1}{E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}d_{cci0}\right) - E\left(\frac{1}{T}d_{c\alpha i0}\right)^2} \\
& \left[E\left(\frac{1}{T}d_{\alpha\alpha i0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma cci0}\right) + E\left(\frac{1}{T}d_{cci0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma\alpha\alpha i0}\right) - 2E\left(\frac{1}{T}d_{c\alpha i0}\right)E\left(\frac{1}{T}\mathbf{d}_{\gamma aci0}\right)\right] \\
& +O_p\left(T^{-1/2}\right)
\end{aligned} \tag{37}
$$

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:

$$\frac{\partial}{\partial \alpha_i} E\left(\mathbf{d}_{\gamma ci}\right) = E\left(\mathbf{d}_{\gamma \alpha ci}\right) + E\left(\mathbf{d}_{\gamma ci} d_{\alpha i}\right) \tag{38}$$

This works with other derivatives of expectations as well.

C is the sum of two derivatives, which we call C^{α} and C^{c} respectively, evaluated at $\eta_i = \eta_i(\gamma)$. C^{α} is equal to

$$\begin{aligned}
C^{\alpha} &= -\frac{\partial}{\partial a_i}\left(\frac{E\left(\mathbf{d}_{\gamma ci}\right)E(d_{c\alpha i}) - E(d_{cci})E\left(\mathbf{d}_{\gamma \alpha i}\right)}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2}\right) \\
&= -\frac{\frac{\partial}{\partial a_i}\left(E\left(\mathbf{d}_{\gamma ci}\right)E(d_{c\alpha i}) - E(d_{cci})E\left(\mathbf{d}_{\gamma \alpha i}\right)\right)}{E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2} \\
&\quad + \frac{\left(E\left(\mathbf{d}_{\gamma ci}\right)E(d_{c\alpha i}) - E(d_{cci})E\left(\mathbf{d}_{\gamma \alpha i}\right)\right)\frac{\partial}{\partial a_i}\left(E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2\right)}{\left(E(d_{\alpha\alpha i})E(d_{cci}) - [E(d_{c\alpha i})]^2\right)^2}
\end{aligned}$$

Working with the derivative and using the above result, we get

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

Likewise, for C^{c} we have

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

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

DOI: 10.1002/jae

Putting everything together. Finally, if we 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 (29). Therefore, the modified score equals 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)*, which 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.

DOI: 10.1002/jae