A dynamic ordered logit model with fixed effects

Chris Muris, Pedro Raposo, and Sotiris Vandoros

April 12, 2023

Abstract

We study a fixed-T panel data logit model for ordered outcomes that accommodates fixed effects and state dependence. We provide identification results for the autoregressive parameter, regression coefficients, and the threshold parameters in this model. Our results require only four observations on the outcome variable. We provide conditions under which a composite conditional maximum likelihood estimator is consistent and asymptotically normal. We use our estimator to explore the determinants of self-reported health in a panel of European countries over the period 2003–2016 and find evidence for state dependence in self-reported health.[1]

EconLit Code: C23.

[1]Muris: Department of Economics, McMaster University. Contact: muerisc@mcmaster.ca. Raposo: Católica Lisbon School of Business and Economics. Contact: pedro.raposo@ucp.pt. Vandoros: King's College London and Harvard University. Contact: s.vandoros@kcl.ac.uk. The data were made available by Eurostat (Contract RPP 132-2018-EU-SILC). We are grateful to the editor and three anonymous referees for extensive feedback; to Pedro Pita Barros, Irene Botosaru, Geraldo Cerqueiro, Petros Migiakis, Krishna Pendakur, Pedro Portugal and Paulo Rodrigues for helpful discussions. This research was supported by the Social Sciences and Humanities Research Council of Canada (grant number IG 435-2021-0778).

1 Introduction

Certain individual-level conditions may tend to persist over time. A common example of this in the literature is self-reported health status. Health status often depends on its value in the previous period, as health conditions may persist over time, given that recovery can take long, and that an illness may even have permanent effects. For example, Table 1 presents a transition matrix for self-reported health status in the United Kingdom for the period 2003-2016.[2]

For individuals who report a value of current health in a given year (rows, on a 5-point scale with 5 being the highest), it shows the relative proportion of those who report a certain value in the subsequent year (columns). A striking feature of this transition matrix is that much of the mass is on or near the main diagonal. This feature is found across all countries in our analysis. In other words, health status is persistent: individuals tend to remain at the same level of health.

There are at least two explanations for this observed persistence (Heckman, 1981; Honoré and Kyriazidou, 2000): unobserved heterogeneity and state dependence. Consider first *unobserved heterogeneity.* It refers to unobservable characteristics that affect the propensity to report permanently higher or lower health status. These permanent unobserved (and observed) characteristics of individuals that affect their health outcome have been extensively addressed in the health literature (see for example Jones and Wildman, 2008; Lorgelly and Lindley, 2008). As a result, it is important to take into account the role of unobserved heterogeneity when analyzing health data. The appropriate econometric approach is to allow for fixed effects.

Consider now the second explanation for the observed persistence in health outcomes: *state dependence*. This refers to the possibility that past health status may be related to current health status even after conditioning on unobserved heterogeneity. State dependence arises if health shocks are persistent, in the sense that a shock on health can have a long-lasting effect (a typical example is injury leading to disability). Using a random effects approach, Contoyannis et al. (2004) found evidence for such persistence in respondents of the British Household Panel survey.

[2]More information about the data and source is in Section 7, and Section C of the Supplemental Appendix, where we analyze these data using the methodology proposed herein.

We propose and analyze a panel data ordered logit model that includes both fixed effects and a lagged dependent variable. This allows a researcher faced with panel data and an ordinal outcome variable to disentangle unobserved heterogeneity from state dependence, and to quantify state dependence. Specifically, we study the *dynamic ordered logit model with fixed effects,* or *DOLFE* from now on,

$$Y_{i,t}^* = \alpha_i + X_{i,t}\beta + \rho 1\left\{Y_{i,t-1} \geq k\right\} - U_{i,t}, t = 1,2,3, \tag{1}$$

$$Y_{i,t} = \begin{cases} 1 & \text{if } Y_{i,t}^* < \gamma_2, \\ 2 & \text{if } \gamma_2 \leq Y_{i,t}^* < \gamma_3, \\ \vdots & \\ J & \text{if } Y_{i,t}^* \geq \gamma_J, \end{cases} \tag{2}$$

$$U_{i,t} | \left(\alpha_i, X_i, Y_{i,<t}\right) \sim LOG(0,1), t = 1,2,3, \tag{3}$$

where k, $2 \leq k \leq J$, is a fixed and known cutoff for the lagged dependent variable. The person-specific parameter α_i captures unobserved heterogeneity, which we allow to be correlated with the other quantities in the model in an unrestricted way (fixed effects). The time-varying covariates $X_{i,t} \in \mathbb{R}^{K_x}$ are collected across time periods in $X_i = (X_{i,1}, X_{i,2}, X_{i,3})$, and the lagged dependent variables for period t are collected in $Y_{i,<t} = (Y_{i,0}, \cdots, Y_{i,t-1})$. The autoregressive parameter ρ is the regression coefficient on the lagged dependent variable $1\left\{Y_{i,t-1} \geq k\right\}$; β is the regression coefficient on the contemporaneous covariates; and the threshold parameters γ_j map the underlying latent variable $Y_{i,t}^*$ into the observed ordered outcome $Y_{i,t}$. Equation (3) restricts the error terms $U_{i,t}$ to be i.i.d. logistic,[3] and is a strict exogeneity assumption on the regressors and past outcomes.

This model combines a number of noteworthy features. First, it is a model for discrete ordered outcomes, and therefore a *nonlinear* model. Second, it is *dynamic*, in the sense that the current

[3]Section B in the Supplemental Appendix discusses an extension that drops the logistic assumption.

outcome depends directly on the outcome in the previous period. This feature, called *state dependence*, is governed by the autoregressive parameter ρ. Third, it allows for *unobserved heterogeneity* in an unrestricted way, i.e. it is a *fixed effects* model. Fourth, the model is specified for only a *small number of time periods*, $T = 3$. Period 0 is unmodelled, but an observation on the outcome variable in time 0 is required for identification.

We believe that we are the first to provide identification and estimation results for all common parameters in a dynamic ordered logit model with fixed effects and a fixed number of time periods. Using four time periods of data on the ordinal outcome variable, we identify the autoregressive coefficients on the lagged dependent variable, and the regression coefficients on the exogenous regressors. We also identify the threshold parameters, which provides an additional interpretation of the magnitude of the estimated coefficients. Our identification result suggests a composite conditional maximum likelihood estimator for the parameters in our model. We establish conditions under which that estimator is consistent and asymptotically normal.

We use our estimator to investigate the determinants of self-reported health, focusing on the link between income and health in a panel of European countries over the period 2003-2016. We obtain two main findings. First, even after controlling for unobserved heterogeneity, persistence plays a positive and significant role in one's self-reported health. In other words, one's health is dependent on the health in the previous period. This is to be expected, as health problems may extend over a number of periods, or become permanent; and healthy people may be more likely to remain healthy compared to those having existing health problems. Second, we find that when controlling for unobserved heterogeneity, the link between income and health becomes statistically insignificant, suggesting that other factors might explain the association between the two. This is in line with studies that report a smaller or insignificant association when using fixed effects (Gunasekara et al, 2011; Larrimore, 2011).

2 Related literature in econometrics

We believe that our paper is the first to provide identification and estimation results for a panel data model with (i) ordered outcomes; (ii) a lagged dependent variable; (iii) fixed effects; and (iv) a fixed number of time periods. Our econometric contribution is related to several strands of literature, each of which features a subset of these features.

Most closely related is the literature on binary and multinomial choice models with fixed effects and lagged dependent variables, which features all but (i). The seminal work by Honoré and Kyriazidou (2000) builds on Chamberlain (1985) to estimate the parameters in dynamic binary choice logit models with fixed effects and time-varying regressors. Hahn (2001) discusses the information bound for this model. Bartolucci and Nigro (2010, 2012) propose and analyze a quadratic exponential approximation. Kitazawa (2022) and Honoré and Weidner (2020) construct moment conditions that identify the finite-dimensional parameters. Honoré and Kyriazidou (2019) discuss identification of some closely related models. Honoré and Tamer (2006), Aristodemou (2021), and Khan et al. (2021, 2022) obtain results for models that do not have logistic errors. For the multinomial model, Chamberlain (1980) studies the static logit case; Magnac (2000) studies the dynamic version with logistic errors. For the multinomial model without logistic errors, see Shi et al. (2018) for results on the static model. Khan et al. (2021) study both static and dynamic models. We supplement these results by showing that in an *ordered* choice model, the thresholds in the latent variable model can be identified along with the regression coefficients and the autoregressive parameter.

The literature on static ordered logit models with fixed effects features all but (ii). This model was analyzed by Das and van Soest (1999), Johnson (2004), Baetschmann et al. (2015), Muris (2017), Botosaru and Muris (2017), and Botosaru et al. (2022). Bartolucci et al. (2022) start from the static model to test for the presence of state dependence. Our result differs from the results in those papers by providing identification and estimation results for a *dynamic* version of the ordered logit model.

The literature on random effects dynamic ordered choice models features all but (iii). Random

effects dynamic ordered choice models have been studied and applied extensively (Contoyannis et al., 2004; Albarran et al., 2019). Such approaches require strong restrictions on the relationship between the unobserved heterogeneity and the exogeneous variables in the model. Such restrictions are usually unappealing to economists, as evidenced by the fact that they are rarely used in linear models. Our approach does not impose random effects restrictions and is the first to provide results for a fixed effects version of this model.

Note that our approach is fixed-T consistent. The difficulty of allowing for fixed effects and state dependence is alleviated when one can assume that $T \to \infty$, referred to as "large-T". Large-T fixed effects dynamic ordered choice models have been studied by Carro and Traferri (2014) and Fernández-Val et al. (2017), see also Carro (2007) for the binary outcome case. In the large-T case, one can use techniques that correct for the bias that derives from including fixed effects in the nonlinear panel model. This approach does not feature (iv). These techniques are not appropriate for our empirical application, which is a rotating panel with $T = 4$.

One limitation of our approach, discussed in more detail in Section 3, is that we restrict the way in which the lagged dependent variable enters the model. The random effects and large-T approaches can accommodate a richer dynamic specification. In subsequent work, Honoré et al. (2021), building on Honoré and Weidner (2020), find moment conditions for a fixed-T, fixed effects ordered logit model with arbitrary dynamics. For that approach, the logit assumption on the error terms is essential, whereas our approach extends to non-logit models with minor modifications (see Section B in the Supplemental Appendix). Our estimation approach differs from that in Honoré et al. (2021). Our estimator is likelihood-based, whereas the estimator in Honoré et al. (2021) is based on conditional moments. An advantage of our likelihood approach is that it is computationally very straightforward and requires only one step, whereas the GMM objective function in Honoré et al. (2021) requires a preliminary step to estimate the weight matrix and to turn the conditional moments into unconditional ones. A disadvantage of our approach is that it is based on information from a subpopulation $X_{i,2} = X_{i,3}$ and therefore requires the choice of a bandwidth parameter is a continuous regressor is present. This is not the case for the approach in

Honoré et al. (2021), nor for the approaches in Kitagawa (2021) and Khan et al. (2022).

3 Binary dynamics

Our approach allows for "binary dynamics": the lagged dependent variable only shows up through the term $\rho 1\left\{Y_{i,t-1} \geq k\right\}$, where k is known and ρ is estimated. With binary dynamics, there is a ceteris paribus effect on the future outcome $Y_{i,t+1}$ if we exogenously change the value of $Y_{i,t}$ from a group of low values ($Y_{i,t} < k$) to the group of high values ($Y_{i,t} \geq k$), but there is no effect for a change in $Y_{i,t}$ that leaves the group unchanged.

The purpose of this section is to give a few examples where such a structure is plausible. We start with an example from health economics, motivating the empirical application in Section 7. We then give two more examples, from financial economics and sports economics. These examples have a common structure: we identify groups of high and low values of Y_{it} and argue that the ceteris paribus effect on future values of $Y_{i,t+1}$ of between-group changes in $Y_{i,t}$ is much more important than within-group changes in $Y_{i,t}$.

Our first example is our empirical application on self-reported health status. While there are multiple levels of self-reported health, being above or below a threshold at which medical care is demanded in period $t-1$ is essential in determining health status in period t. Galama and Kapteyn (2011) formalized this important issue by building on the seminal papers by Grossman (1972a; 1972b).

Second, government bonds are ordered by rating (20 tiers), which are grouped into two categories: Investment Grade (IG) and Non-Investment Grade (NIG). IG Governments can easily raise credit, NIG Governments cannot. This affects economic indicators and returns expectations, raising a country's economic profile and next period's credit rating (Rigobon, 2002).

Third, finishing in the top four positions of the soccer domestic league qualifies a team for the European Champions League competition the following season. This leads to considerable financial benefits (Peeters and Szymanski, 2014), enabling the club to sign better soccer players,

thus increasing the chance of performing better in next year's league (Ferri et al., 2017).

4 Identification

We normalize $\gamma_k = 0$, where k is as in equation (1). This normalization is without loss of generality because α_i is unrestricted. Our model implies that the binary variable $D_{i,t}(k) = 1\{Y_{i,t} \geq k\}$ follows the dynamic binary choice logit model in Honoré and Kyriazidou (2000), HK hereinafter. Specifically, equation (3) in HK applies to the transformed model

$$D_{i,t}(k) = 1\left\{X_{i,t}\beta + \rho D_{i,t-1}(k) + \alpha_i - U_{i,t} \geq 0\right\}, \tag{4}$$

i.e. the transformed model follows a dynamic binary choice logit model with fixed effects. The implied conditional probabilities relevant for our analysis are

$$P\left(D_{i,0}(k) = 1 \middle| X_i, \alpha_i\right) \equiv p_0(X_i, \alpha_i), \tag{5}$$

and, for $t = 1, 2, 3$,

$$P\left(D_{i,t}(k) = 1 | X_i, \alpha_i, D_{i,<t}(k)\right) = \Lambda\left(\alpha_i + X_{i,t}\beta + \rho D_{i,t-1}(k)\right), \tag{6}$$

where the logistic CDF is denoted by $\Lambda(u) = \exp(u)/(1+\exp(u))$, and we let $D_{i,<t}(k) = (D_{i,0}(k), \cdots, D_{i,t-1}(k))$ HK provide conditions that guarantee identification of β and ρ by constructing a conditional probability that features (β, ρ) but that is free of α_i.

If $Y_{i,t}$ has at least three points of support, there is information in $Y_{i,t}$ beyond $D_{i,t}(k)$. In the remainder of this section we show that this information can be used to identify the threshold parameters

$$\gamma \equiv (\gamma_2, \gamma_3, \cdots, \gamma_{k-1}, \gamma_{k+1}, \cdots, \gamma_J).$$

Knowledge of γ helps with the interpretation of all model parameters, see Section 4.1.

We now construct a conditional probability that features (β, ρ, γ) but not the incidental parameters α_i. To this end, extend the definition

$$D_{i,t}(j) = 1\{Y_{i,t} \geq j\}, 2 \leq j \leq J,$$

to thresholds $j \neq k$, and abbreviate $D_{i,t} \equiv D_{i,t}(k)$. Define the events $\left(A_{i,jl,d_0,d_3}, B_{i,jl,d_0,d_3}, C_{i,jl,d_0,d_3}\right)$, with $2 \leq j \leq k \leq l \leq J$, as

$$\begin{aligned} A_{i,jl,d_0d_3} &= \left\{D_{i,0} = d_0, D_{i,1} = 0, D_{i,2}(l) = 1, D_{i,3}(j) = d_3\right\}, \\ B_{i,jl,d_0d_3} &= \left\{D_{i,0} = d_0, D_{i,1} = 1, D_{i,2}(j) = 0, D_{i,3}(l) = d_3\right\}, \\ C_{i,jl,d_0d_3} &= A_{i,jl,d_0d_3} \cup B_{i,jl,d_0d_3}. \end{aligned} \tag{7}$$

The event A_{i,jl,d_0d_3} corresponds to moving up during the middle periods $t = 1,2$, starting below k to moving up to at least $l \geq k$. The event B_{i,jl,d_0d_3} corresponds to moving down from at least k to below $j \leq k$.

If $j = k = l$, the event C_{i,kk,d_0d_3} corresponds to switchers (observations with $D_{i,1} + D_{i,2} = 1$), as in HK, see (4). But in the ordered model, it is possible to vary the cutoffs in the periods $t = 2,3$, if the dependent variable has more than two points of support. Varying the cutoffs in this way is what distinguishes our conditioning event from HK's. This allows us to identify the threshold parameters.

The following result shows that different choices of (j,l) reveal different combinations of thresholds in certain conditional probabilities that do not depend on the incidental parameters α_i. In what follows, the change in the regressors from period 1 to 2 is given by $\Delta X_i = X_{i,2} - X_{i,1}$.

Theorem 1 (Sufficiency)**.** *For DOLFE (the dynamic ordered logit model with fixed effects), for*

any (j,l) *such that* $2 \leq j \leq k \leq l \leq J$*, and for any* $d_0, d_3 \in \{0,1\} \times \{0,1\}$,

$$P\left(A_{i,jl,d_0d_3} \middle| X_i, C_{i,jl,d_0d_3}, X_{i,2} = X_{i,3}\right) = 1 - \Lambda\left(\Delta X_i\beta + \rho\left(d_0 - d_3\right) + \left(1 - d_3\right)\gamma_l + d_3\gamma_j\right) \quad (8)$$

$$P\left(B_{i,jl,d_0d_3} \middle| X_i, C_{i,jl,d_0d_3}, X_{i,2} = X_{i,3}\right) = \Lambda\left(\Delta X_i\beta + \rho\left(d_0 - d_3\right) + \left(1 - d_3\right)\gamma_l + d_3\gamma_j\right). \quad (9)$$

Identification of all model parameters is achieved by considering all possible combinations of cutoffs. It is clear from Theorem 1 that different choices for (j,k,l,d_0,d_3) reveal information about distinct linear combinations of (β,ρ,γ). If the following assumption holds, the joint information across all choices of (j,k,l,d_0,d_3) is sufficient to identify all model parameters.

Assumption 1. *For all* (j,l) *such that* $2 \leq j \leq k \leq l$*, and for all* $d_0, d_3 \in \{0,1\}$

$$Var\left(\Delta X_i | X_{i,2} = X_{i,3}, C_{i,jl,d_0,d_3}\right)$$

is invertible.

This assumption guarantees that for each choice of (j,l), there is sufficient variation in ΔX_i in the subpopulation of stayers $X_{i,2} = X_{i,3}$ to identify the regression coefficient. This assumption can be weakened: we need sufficient variation for only *some* (j,l). However, if it fails for sufficiently many (j,l), not all threshold parameters may be identified.

Denote by $Y_i = (Y_{i,0}, Y_{i,1}, Y_{i,2}, Y_{i,3})$ the time series of dependent variables for a given individual.

Theorem 2 (Identification)**.** *If Assumption 1 holds, then* (β,ρ,γ) *can be identified from the joint distribution of the vector* (X_i, Y_i) *generated by DOLFE (the dynamic ordered logit model with fixed effects).*

4.1 Interpretation of the parameters

A key difference between ordered choice models with $J > 2$ and binary choice models with $J = 2$ is the presence of the threshold parameters γ_j. These parameters measure the width of each category on the scale of the latent variable $Y^*_{i,t}$. They may be used for interpretations not available for binary

choice models with fixed effects, which feature only one threshold. Here, we discuss such an interpretation by building on the approach in Muris (2017, III.c) for static ordered choice models with fixed effects.[4]

To do this, let

$$Y_{i,t}^{*}(x) = \alpha_i + x\beta + \rho 1\left\{Y_{i,t-1} \geq k\right\} - U_{i,t}, \tag{10}$$

be the latent variable for an individual with regressors set to a counterfactual value x, while keeping everything else equal. Note that $Y_{i,t}^{*} = Y_{i,t}^{*}(X_{i,t})$. Define the counterfactual for the ordered outcome as

$$Y_{i,t}(x) \geq j \Leftrightarrow Y_{i,t}^{*}(x) \geq \gamma_j. \tag{11}$$

Consider an individual who in time period t is observed to be in an intermediate category, i.e. $Y_{i,t} = j \in \{2, \cdots, J-1\}$. The latent variable model tells us that

$$Y_{i,t}^{*} = \alpha_i + X_{i,t}\beta + \rho 1\left\{Y_{i,t-1} \geq k\right\} - U_{i,t} \in \left[\gamma_j, \gamma_{j+1}\right).$$

Let

$$\delta_m^j = \frac{\gamma_{j+1} - \gamma_j}{\beta_m}, \tag{12}$$

where β_m is the regression coefficient associated with regressor m, and pick a $\Delta \geq \delta_m^j$. Then

$$\begin{aligned} Y_{i,t}^{*} + \Delta \times \beta_m &\geq \gamma_j + \delta_m^j \times \beta_m \\ &= \gamma_j + \left(\gamma_{j+1} - \gamma_j\right) \\ &= \gamma_{j+1}, \end{aligned}$$

[4]Alternatively, we could extend the results on partial effects for static nonlinear panels in Botosaru and Muris (2017) to the dynamic case.

so that by (11), we have that

$$P\left(Y_{i,t}\left(X_{i,t}+\Delta\times e_m\right)\geq j+1\middle|Y_{i,t}=j\right)=1,$$

where e_m is a vector with K_x elements that are all zero, except for the mth element that is equal to 1. In other words, a change of at least δ_m^j in the value of regressor m moves *any* individual who is observed to be in category $j=2,\cdots,J-1$ to at least category $j+1$, all else equal.[5]

5 Estimation

Theorem 1 suggests that one can use a conditional maximum likelihood estimator (CMLE) to estimate a linear combination of the model parameters for each choice of $2\leq j\leq k\leq l$. Theorem

[5]A similar interpretation can be constructed for the autoregressive coefficient ρ. Define the counterfactual latent variable as a function of the lagged outcome holding the regressors at $X_{i,t}$, i.e.

$$Y_{i,t}^*\left(y_{t-1}\right)=\alpha_i+X_{i,t}\beta+\rho 1\left\{y_{t-1}\geq k\right\}-U_{i,t}, \tag{13}$$

and define $Y_{i,t}\left(y_{t-1}\right)$ analogous to (11). Let

$$\delta_\rho^j=\frac{\gamma_{j+1}-\gamma_j}{\rho}, \tag{14}$$

and following the same reasoning as above, if $\delta_\rho^j<1$, then

$$P\left(Y_{i,t}\left(1\right)\geq j+1\middle|Y_{i,t}=j,Y_{i,t-1}=0\right)=1.$$

In other words, if $\delta_\rho^j<1$, then all individuals with $Y_{i,t-1}<k$ and $Y_{i,t}=j\in\{2,\cdots,J-1\}$ would move to at least category $j+1$ if we change their $Y_{i,t-1}$ to at least k, all else equal. Thus, we may think of δ_ρ^j as a measure of state dependence. Even if $\delta_\rho^j\geq 1$, and the exact reasoning above does not apply, we may still interpret it as a measure of state dependence.

2 suggests that a *composite* CMLE (CCMLE), based on the combination of conditional likelihoods across all choices of (j,l), may be used to estimate all model parameters (β,ρ,γ). In this section we define the CCMLE and study its large sample behavior by extending HK's results for dynamic binary choice models to our setting.

The basic idea is to construct an estimator based on the conditional likelihood in Theorem 1. Which subset of parameters is involved will depend on the choice of (j,l). The conditional likelihood for a given (j,l) features the subvector θ_{jl} of θ of length K_{jl},

$$\theta_{jl} = \begin{cases} (\beta,\rho) & \text{if } j=k=l, \\ (\beta,\rho,\gamma_j,\gamma_l) & \text{if } j<k<l, \\ (\beta,\rho,\gamma_l) & \text{if } j=k<l, \\ (\beta,\rho,\gamma_j) & \text{if } j<k=l, \end{cases}$$

as the coefficients on augmented regressors $Z_{i,jl}$,

$$Z_{i,jl} = \begin{cases} (\Delta X_i, D_{i,0}-D_{i,3,jl}) & \text{if } j=k=l, \\ (\Delta X_i, D_{i,0}-D_{i,3,jl}, D_{i,3,jl}, 1-D_{i,3,jl}) & \text{if } j<k<l, \\ (\Delta X_i, D_{i,0}-D_{i,3,jl}, 1-D_{i,3,jl}) & \text{if } j=k<l, \\ (\Delta X_i, D_{i,0}-D_{i,3,jl}, D_{i,3,jl}) & \text{if } j<k=l, \end{cases}$$

where

$$D_{i,3,jl} = \begin{cases} D_{i,3}(j) & \text{if } D_{i,1}=0, \\ D_{i,3}(l) & \text{if } D_{i,1}=1. \end{cases}$$

We similarly define

$$D_{i,2,jl} = \begin{cases} D_{i,2}(l) & \text{if } D_{i,1} = 0, \\ D_{i,2}(j) & \text{if } D_{i,1} = 1, \end{cases}$$

and define

$$h_{i,jl}\left(\theta_{jl}\right) = 1\left\{D_{i,1} \neq D_{i,2,jl}\right\} \times \ln\left[\frac{\exp\left(Z_{i,jl}\theta_{jl}\right)^{D_{i1}}}{1+\exp\left(Z_{i,jl}\theta_{jl}\right)}\right]. \tag{15}$$

Our notation mirrors that in HK: Compare our (15) to $h(\cdot)$ on p. 848 of HK. The only differences are capitalization, our jl-subscript, and that we have D instead of y due to binarization. Thus, for every (j,l), our $h_{i,jl}$ takes the form of HK's unweighted objective function.

In view of the conditioning in Theorem 1, we must consider a version of (15) local to $X_{i,2} = X_{i,3}$. With discrete covariates, we could use

$$l^{D}_{i,jl}\left(\theta_{jl}\right) = 1\left\{X_{i,2} = X_{i,3}\right\} \times h_{i,jl}\left(\theta_{jl}\right).$$

The limiting distribution of an estimator based on $l^{D}_{i,jl}$ can be obtained using standard methods. Our empirical application has both continuous and discrete covariates. To accommodate that case, partition

$$X_{i,t} = \begin{bmatrix} X_{i,t,c} & X_{i,t,d} \end{bmatrix}$$

into a subvector of $K_{x,c}$ continuous covariates $X_{i,t,c}$, and a subvector $X_{i,t,d}$ with $K_{x,d}$ discrete covariates, where $K_{x,c}, K_{x,d} \geq 1$ and $K_{x,c} + K_{x,d} = K_x$.

Define

$$l_{i,jl}\left(\theta_{jl}\right) = K\left(\frac{X_{i,2,c} - X_{i,3,c}}{\sigma_n}\right) \times 1\left\{X_{i,2,d} - X_{i,3,d} = 0\right\} \times h_{i,jl}\left(\theta_{jl}\right), \tag{16}$$

where K is a kernel function and σ_n is a bandwidth satisfying conditions described below. Our estimator maximizes the sample average of (16) summed over (j,l):

$$\widehat{\theta}_n = \text{argmax}_{\theta\in\Theta} \sum_{2\leq j\leq k\leq l\leq J} \sum_{i=1}^{n} l_{i,jl}\left(\theta_{jl}\right). \tag{17}$$

Theorem 3. *If Assumption 1 holds, and if the following assumptions hold:*

(C1) $\{(Y_i, X_i)\}_{i=1}^n$ *is a random sample of n observations from DOLFE;*

(C2) $\theta_0 \in \Theta$ *and* Θ *is compact;*

(C3) Conditional on $X_{i,2,d} - X_{i,3,d} = 0$*, the random vector* $X_{i,2,c} - X_{i,3,c}$ *is absolutely continuously distributed with conditional density denoted* $f_{c|d}(\cdot|0)$ *that is bounded from above on its support, and strictly positive and continuous in a neighborhood of zero. Furthermore,*

$$q_0 := P\left(X_{i,2,d} - X_{i,3,d} = 0\right) > 0;$$

(C4) The function $E\left[\|\Delta X_i\|^2 \middle| X_{i,2,c} - X_{i,3,c} = \cdot, X_{i,2,d} - X_{i,2,d} = 0\right]$ *is bounded on its support;*

(C5) For each j,l*, the function* $E\left[h_{i,jl}\left(\theta_{jl}\right) \middle| X_{i,2,c} - X_{i,3,c} = \cdot, X_{i,2,d} - X_{i,2,d} = 0\right]$ *is continuous in a neighborhood of zero for all* θ_{jl}*;*

(C6) The function $E\left[\Delta X_i' \Delta X_i \middle| X_{i,2,c} - X_{i,3,c} = \cdot, X_{i,2,d} - X_{i,3,d} = 0\right]$ *has full column rank* K_x *in a neighbourhood of zero;*

(C7) $K : \mathbb{R}^{K_{x,c}} \to \mathbb{R}$ *is a function of bounded variation that satisfies: (i)* $\sup_{v\in\mathbb{R}^{K_{x,c}}} |K(v)| < \infty$*, (ii)* $\int |K(v)|\, dv < \infty$*, and* $\int K(v)\, dv = 1$*;*

(C8) σ_n *is a sequence of positive numbers such that* $\sigma_n \to 0$ *as* $n \to \infty$*,*

then $\widehat{\theta}_n \xrightarrow{p} \theta_0$ *as* $n \to \infty$*.*

Proof. See Appendix A.3. Our proof extends HK's proof of their Theorem 1. □

To describe the limiting distribution of the CCMLE (17), let θ_{jl0} denote the true value of θ_{jl} (the corresponding subvector of θ_0); let Θ_{jl} denote the associated parameter space; and let $h_{jl}^{(1)}(\cdot)$ and $h_{jl}^{(2)}(\cdot)$ denote the first and second derivatives of h_{jl} with respect to θ.

Theorem 4. *Let the assumptions of Theorem 3 hold, and additionally assume:*

(N1) $\theta_0 \in int(\Theta)$*;*

(N2) $f_{c|d}(\cdot|0)$ (defined in Theorem 3, C3) is s times differentiable on its support and has bounded derivatives;

(N3) For each j,l, the function $E\left[h_{i,jl}^{(1)}\left(\theta_{jl0}\right)\middle|X_{i,2,c}-X_{i,3,c}=\cdot,X_{i,2,d}-X_{i,3,d}=0\right]$ is s times differentiable on its support and has bounded derivatives;

(N4) For each j,l, the function $E\left[h_{i,jl}^{(2)}\left(\theta_{jl}\right)\middle|X_{i,2,c}-X_{i,3,c}=\cdot,X_{i,2,d}-X_{i,3,d}=0\right]$ is continuous in a neighborhood of zero for all $\theta_{jl}\in\Theta_{jl}$;

(N5) The function $E\left[\|\Delta X_i\|^6\middle|X_{i,2,c}-X_{i,3,c}=\cdot,X_{i,2,d}-X_{i,3,d}=0\right]$ is bounded on its support;

(N6) The function $E\left[h_{i,jl}^{(1)}\left(\theta_{jl0}\right)h_{i,jl}^{(1)}\left(\theta_{jl0}\right)'\middle|X_{i,2,c}-X_{i,3,c}=\cdot,X_{i,2,d}-X_{i,3,d}=0\right]$ is continuous in a neighbourhood of zero;

(N7) $K:\mathbb{R}^{K_{x,c}}\to\mathbb{R}$ is an s'th order bias-reducing kernel that satisfies conditions (N7i) and (N7ii) in HK's Theorem 2;

Then, as $n\to\infty$ and $\sqrt{n\sigma_n^{K_{x,c}}}\sigma_n^s\to 0$,

$$\sqrt{n\sigma_n^{K_{x,c}}}\left(\widehat{\theta}_n-\theta_0\right)\xrightarrow{d}\mathcal{N}\left(0,J^{-1}VJ^{-1}\right),$$

where

$$J=\sum_{j,l}J_{j,l},\ J_{j,l}=-f_{c|d}\left(0|0\right)q_0E\left[h_{jl}^{(2)}\left(\theta_{jl,0}\right)\middle|X_{i,2}-X_{i,3}=0\right],$$

and V is the asymptotic variance of $\sum_{j,l}Z_{n,jl}\left(\theta_0\right)$,

$$\begin{aligned}Z_{n,jl}\left(\theta_0\right)=&\frac{1}{\sqrt{n\sigma_n^{K_{x,c}}}}\sum_{i=1}^{n}K\left(\frac{X_{i,2,c}-X_{i,3,c}}{\sigma_n}\right)\times 1\left\{X_{i,2,d}-X_{i,3,d}=0\right\}\times h_{i,jl}^{(1)}\left(\theta_{jl,0}\right)\\&-\frac{1}{\sqrt{n\sigma_n^{K_{x,c}}}}\sum_{i=1}^{n}E\left[K\left(\frac{X_{i,2,c}-X_{i,3,c}}{\sigma_n}\right)\times 1\left\{X_{i,2,d}-X_{i,3,d}=0\right\}\times h_{i,jl}^{(1)}\left(\theta_{jl,0}\right)\right].\end{aligned}$$

Proof. See Appendix A.4. □

We recommend estimating standard errors using the bootstrap, and forming confidence intervals based on a normal approximation using these bootstrap standard errors. We investigate the performance

of this procedure in a simulation study in Section 6, and use it in the empirical application in Section 7.

6 Simulation study

We now describe the results of a simulation study on the finite sample properties of the CCMLE introduced in Section 5. The simulation study was designed with the empirical application in Section 7 in mind: we use the regressor values and parameter estimates from Section 7 to generate data from DOLFE.[6] The goal of this section is also to inform our choice of bandwidth for the empirical application, and to investigate the reliability of bootstrap-based confidence intervals.

First, we document the behavior of the CCMLE for the autoregressive parameter and threshold parameters for various values of the sample size n and various values of the bandwidth parameter σ_n. For each of the $S = 100$ simulation runs, we use the regressor values of the first n observations, generate $\alpha_i \sim \mathcal{N}(0,1)$ and then compute $Y_{i,t}$ from DOLFE. We repeat this for sample sizes $n \in \{500, 1000, 5000, 10000, 50000, 100000, 150000, 200000, 260601\}$, where $n = 260601$ is the sample size in the empirical application. We compute the CCMLE for bandwidth parameter values $\sigma_n \in \{0.01, 0.05, 0.1, 0.25, 0.5, 1.0, 2.0, 5.0, 10.0\}$, and use the Gaussian kernel.

Table 2 reports results for the autoregressive parameter, with true value $\rho = 0.733$. Table 3 presents additional results for the estimators of the threshold parameters $\gamma_2 = -3.275$ and $\gamma_4 = 3.326$. Table 2 reveals significant finite sample bias at $n = 500$ across all choices of σ_n. Some of the bias is still present at $n = 1000$. Once the sample size reaches $n = 5000$, the bias is negligible for a sufficiently large value of the bandwidth. For sample sizes above $n = 150000$, the bias is small across all bandwidths. For the preferred bandwidth in our empirical application ($\sigma_n = 1$) the bias at $n = 10000$ and above is $0.732 - 0.733 = -0.001$. The standard deviation of the estimator declines with n.

[6]The parameter estimates we use are in Table 6, column (a). Summary statistics for the regressors are in Supplementary Appendix Section C.

The results for γ_2 and γ_4 in Table 3 also show a small finite sample bias for $n = 50000$ and above. For sample sizes up to 200000, the choice of bandwidth matters for the bias of the threshold parameter estimators. However, for our preferred bandwidth $\sigma_n = 1$, a sample size of 50000 seems to be sufficient for a negligible bias (-0.002 and 0.000, respectively).

Second, we investigate the performance of bootstrap inference for the CCMLE. We focus on the regression coefficient $\beta_1 = 0.049$ on the continuous covariate $\log income_{it}$. Table 4 presents coverage rates of 95% confidence intervals constructed from a normal approximation, the CCMLE point estimate, and bootstrap standard errors from 100 replications.

The main takeaways are: (i) for $n \geq 1000$, the 95% confidence interval that we propose has close to nominal coverage for our preferred bandwidth $\sigma_n = 1$; (ii) at small sample sizes, the confidence intervals are conservative; (iii) confidence intervals shrink with the sample size; (iv) confidence intervals size almost does not decrease in σ_n beyond $\sigma_n = 1$.

Third, we study what happens if the true model has arbitrary dynamics, in the sense that the true DGP has outcome equation

$$\begin{aligned} Y_{i,t}^* = \alpha_i + \rho_2 1\left\{Y_{i,t-1} = 2\right\} + \rho_3 1\left\{Y_{i,t-1} = 3\right\} + \rho_4 1\left\{Y_{i,t-1} = 4\right\} + \\ + \beta_1 \log income_{it} + \cdots + \beta_6 other_{it} - U_{it}, \end{aligned}$$

which coincides with the model underlying our CCMLE only if $\rho_2 = 0$ and $\rho_3 = \rho_4 = 0.733$. We fix $n = 10000$ and $\sigma_n = 1$ and consider results for specifications with

$$(\rho_2 = \delta, \rho_3 = 0.733, \rho_4 = 0.733 + \delta),$$

varying $\delta \in \{0.0, 0.1, 0.2, 0.3, 0.4\}$. The lowest value of $\delta = 0$ corresponds to our model with binary dynamics. A value of $\delta = 0.40$ is close to a model with dynamics that are linear in $Y_{i,t-1}$, which is a severe misspecification relative to $\delta = 0$. Table 5 presents a summary of our results.

The bottom line is that the CCMLE is robust against (severe) misspecification. The mean of

the CCMLE has the right sign and order of magnitude for all coefficients except for ρ.[7] The results for $\beta_6 = -0.370$ are illustrative: under the correct specification, the bias is negligible (0.004 on -0.370). At a medium level of misspecification, the bias increases, but only to about .01.

These three simulation studies inform our estimation strategy for the empirical application below. In particular, they inform our choice of bandwidth (preferred bandwidth $\sigma_n = 0.1$, present results for $\sigma_n \in (0.1, 1, 10)$) and they suggest that our bootstrap-based confidence intervals have the right coverage.

7 Persistence in health status

Our analysis uses panel data for the period 2003-2016 from the European Union Statistics on Income and Living Conditions (EU-SILC), see Eurostat (2017) for detailed documentation.[8] EU-SILC provides a set of indicators on income and poverty, social inclusion, living conditions and, importantly, health status. For each country in the European Union, plus Iceland, Norway, and Switzerland, EU-SILC contains data on a representative sample of the population of those with 18 years and older.

EU-SILC is a rotating panel. Every individual is followed over a period of two to four years. The total number of individual-years for the period 2003-2016 is 1,273,877. Our identification result demands four observations per individual, so we restrict attention to individuals who report valid information on their health status for 4 consecutive years. There are 260,601 individuals (1,042,404 individual-years) who satisfy this restriction and have no missing outcomes or explanatory variables. The proportion of incomplete samples differs across countries. As a result, the sample we work with may not be representative of the EU-SILC's population. For example, of the 27 countries that contribute to our sample, the largest contributors are Italy (with 43,385 individuals), Spain (25,634), and Poland (22,628); the smallest are Portugal (12), Iceland (1,496), and Slovakia

[7]The results for ρ are hard to interpret, as the interpretation of the coefficient estimated by CCMLE changes with δ.

[8]The data were made available to us by Eurostat under Contract RPP 132-2018-EU-SILC.

(1,982).

The outcome variable in our analysis is self-reported health status: self-perceived physical health, elicited during EU-SILC interviews. The person answers the question about how her physical health is in general, at the date of the survey, by classifying it as one of the following: (1) bad and very bad (12% in our sample); (2) fair (26%); (3) good (44%); (4) very good (19%).[9] Of $260,601 \times 3 = 781,803$ health transitions that we observe, most often there is no change in health status (65.6%). Decreases by one unit (16%) are slightly more frequent than increases by one unit (15%). Two-unit increases (1.2%) and decreases (1.5%) are infrequent, and three-unit increases (0.08%) and decreases (0.12%) are rare.

A number of studies have found a positive association between income and health (Carrieri and Jones, 2017; Ettner, 1996; Frijters et al., 2005; Mackenbach et al., 2005), but empirical evidence of a strong effect is sometimes limited (Larrimore, 2011; Gunasekara et al, 2011; Johnston et al., 2009). Therefore, it is important to control for income when examining the determinants of health status.

In Section C of the Supplemental Appendix, we provide detailed descriptive statistics on all the variables used in the application. In our fixed effects results below, we do not further control for variables that do not change over the sample period for a given individual. However, we include a set of time-invariant explanatory variables when we obtain results for non-fixed effects estimators. Coefficient estimates for these variables are omitted from the main text, and reported in Table 3 in Section C of the Supplemental Appendix.

We estimate the parameters in the dynamic ordered logit model with fixed effects, with latent variable outcome equation

$$\begin{aligned} SRH^*_{i,t} = \alpha_i + \rho 1\left\{SRH_{i,t-1} \geq 3\right\} + \beta_1 \log income_{it} + \beta_2 child_{it} + \beta_3 married_{it} + \\ + \beta_4 unemp_{it} + \beta_5 retired_{it} + \beta_6 other_{it} - U_{it}, \end{aligned} \tag{18}$$

[9]We have merged the categories "bad" and "very bad" in the original reported variable, because there is only a small fraction of observations with "very bad" health status.

and present the results in Table 6. The first four columns (a-d, "DOLFE") present the results for (18) using the estimator described in Section 5. Different values of σ_n refer to different values of the bandwidth parameter in Theorems 3 and 4.[10] Column (d) omits the employment variables, to check whether the relationship between employment status and income matters for estimation of the effect of income on health.

The application of our estimator requires a mass of cross-section units whose change in regressors from period 2 to 3 is in a small neighborhood around 0, see assumption (C3) in Theorem 3. In our sample there are 218,254 individuals with no change in their discrete regressors. Among those, there are 213,743 with a change of less than 1 in the continuous regressor; 91,310 with a change of less than 0.1; and 9,014 with a change of less than 0.01. This suggests that the assumption is reasonable for our application.

We also present estimation results for several other estimators. Results for the static ordered logit model with fixed effects, i.e. setting $\rho = 0$ in (18), are obtained using the estimator in Muris (2017), and presented in column (e) ("FEOL"). Column (f) ("DOL") estimates a dynamic ordered logit model without fixed effects, i.e. (18) with $\alpha_i = \alpha$. Column (g) ("OL") presents results for cross-sectional ordered logit estimator that does not take into account fixed effects or dynamics (i.e. $\alpha_i = \alpha$ and $\rho = 0$ in (18)). We also present results for a static linear model with (h, "FELM") and without (i, "LM") fixed effects. The standard errors for all estimators are obtained using the bootstrap (500 replications). For the estimators that are not of the fixed effects type, we additionally control for education, gender, education level, and the level of urbanization.

Income. Our main explanatory variable of interest is income (log income, coefficient β_1). Across almost all specifications we find a positive association between income and health. The only exception is column (b), where the point estimate is negative, and about the same magnitude as the standard error.

[10]Our preferred choice for the bandwidth parameter is $\sigma_n = 1.0$. See Section 6 for a simulation study that investigates the role of this bandwidth parameter, and that provides some support for our preference.

Controlling for unobserved heterogeneity leads to a very strong reduction in the magnitude of the association. For example, a comparison of columns (e) and (g) indicates that for the static case, controlling for unobserved heterogeneity reduces the coefficient on income by more than a factor of 20. For this comparison, note that the threshold differences increase, suggesting that the scale increases; compare also the coefficients on the other variables, with an unchanged order of magnitude. We are not the first to observe a limited association between income and health. In a review of the literature, Gunasekara et al. (2011) found a small positive link between income and self-reported health, which is diminished when controlling for unmeasured confounders.

The estimated effect of income also changes when we control for state dependence. Comparing columns (f) and (g), we see that controlling for state dependence in a model without unobserved heterogeneity reduces the association between income and health. So, individually controlling for unobserved heterogeneity or for dynamics diminishes the magnitude of the association between health and income.

To get a sense of the magnitude of the coefficient, we obtain the sample analog of $\delta^{j}_{\log(income)}$, the parameter introduced in Section 4.1. For our main specification in column (a), we find that

$$\widehat{\delta}^{3}_{\log(income)} = \frac{\widehat{\gamma}_4 - \widehat{\gamma}_3}{\widehat{\beta}_1} \approx \frac{3.326}{0.049} \approx 70,$$

meaning that an increase in log income of 70 would be needed to move all of those in category 3 to category 4. The result for category 2 is similar in magnitude.

Finally, a comparison between columns (a) and (d) shows that the estimate for income association is not robust to controlling for employment status. This seems to indicate that employment status was driving (part of) the association between income and self-reported health.

State dependence. We estimate an autoregressive parameter of around 0.75, with threshold differences of about 3. The estimated ratio of ρ to the thresholds (which measure the distance from category 3) are much lower than for column (f). This confirms the importance of controlling for unobserved heterogeneity, which reduces the estimated magnitude of state dependence by a

factor of 3. Nevertheless, even when controlling for unobserved (and observed) heterogeneity, we find strong evidence for large, positive state dependence in health.

In terms of δ_ρ^j, we find that

$$\widehat{\delta}_\rho^3 = \frac{\widehat{\gamma}_4 - \widehat{\gamma}_3}{\widehat{\rho}} \approx \frac{3.326}{0.733} \approx 4.5,$$

so that changing from a $SRH_{i,t-1} < 3$ to a $SRH_{i,t-1} \geq 3$ leads to a change in the latent variable that is approximately $1/4.5$ of the amount required to move all individuals with $SRH_{i,t} = 3$ to $SRH_{i,t}(1) = 4$.

Another way to gain a sense of the magnitude of persistence, also available for binary choice methods, is to compare estimates of ρ to estimates of regression coefficients. For example, in our preferred specification in column (a), a health shock that lifts a person from any category below 3, to category 3 or 4, has an impact on future health that is almost 4 times that of becoming unemployed. The impact is more than 5 times that of marrying.

Other time-varying covariates.[11] The literature so far has been inconclusive on how retirement is associated with health. On one hand, retiring allows more time for health-promoting activities, and reduces work-related stress. On the other hand, people may lose traction and motivation and may become less active. Therefore, while Coe and Zamarro (2011) find that retirement improves health, Behncke (2012) finds an increase in the likelihood of disease following retirement. In our DOLFE model, the coefficient is statistically insignificant. This suggests that the association between retirement and health may not be as strong as previously thought. Compared to the FEOL model, the association between retirement and health disappears when controlling for state dependence. The results on retirement should be treated with caution. People are more likely to retire at an advanced age, so the retirement variable may be picking up some of the age effect on

[11] It is worth mentioning that as covariates are restricted to having identical values in two periods, the empirical analysis focuses on a sub-group of individuals, and any interpretation of the results should be based on the understanding that these conditions must hold.

health.

The extensive literature on the link between unemployment and health in particular, and economic conditions and health more generally, is broadly inconclusive. Some studies have suggested a protective role of unemployment on health (Ruhm, 2000), while others suggest that unemployment is detrimental for health (McInerney and Mellor, 2012). Our results appear to be more in line with the findings of Ruhm (2015) and Böckerman and Ilmakunnas (2009). In DOLFE, the coefficient of being unemployed is negative and statistically significant. Controlling for state dependence does not change things compared to the FEOL model.

Having children is insignificant in our DOLFE model, while previous studies have provided mixed findings on this issue (Mckenzie and Carter, 2013; Evenson and Simon, 2005). This is also insignificant in the FEOL model, suggesting that previous findings on having children might have been driven by unobserved heterogeneity.

Being married is generally considered a protective factor for health (Kaplan and Kronick, 2006; Molloy et al., 2009). In our model, however, it is statistically insignificant – as opposed to the FEOL model in which it was positive and significant. Thus, controlling for state dependence appears to be important for this variable.

8 Conclusion

We study a fixed$-T$ dynamic ordered logit model with fixed effects. We demonstrate identification of the autoregressive coefficients on the lagged dependent variable, the regression coefficients on the exogenous regressors, as well as differences of the threshold parameters. Our results require only four time periods of data on the ordinal outcome variable. We propose estimators for all these parameters, study their asymptotic behavior, and use them to study the relationship between income and self-reported health while controlling for state dependence and unobserved heterogeneity.

References

[1] Albarran, P., R. Carrasco, and J.M. Carro, 2019. Estimation of Dynamic Nonlinear Random Effects Models with Unbalanced Panels. Oxford Bulletin of Economics and Statistics 81 (6), 1424–1441.

[2] Aristodemou, E., 2021. Semiparametric Identification in Panel Data Discrete Response Models. Journal of Econometrics 220 (2), 253–71.

[3] Baetschmann, G., K.E. Staub, and R. Winkelmann, 2015. Consistent Estimation of the Fixed Effects Ordered Logit Model. Journal of the Royal Statistical Society A, 178 (3), 685–703.

[4] Bartolucci, F., and V. Nigro, 2010. A Dynamic Model for Binary Panel Data With Unobserved Heterogeneity Admitting a $\sqrt{n}$-Consistent Conditional Estimator. Econometrica, 78 (2), 719–33.

[5] Bartolucci, F. and V. Nigro, 2012. Pseudo Conditional Maximum Likelihood Estimation of the Dynamic Logit Model for Binary Panel Data. Journal of Econometrics, 170 (1), 102–116.

[6] Bartolucci, F., C. Pigini, and F. Valentini, 2022. Testing for state dependence in the fixed-effects ordered logit model. MPRA Working paper 113890.

[7] Behncke, S., 2012. Does retirement trigger ill health? Health Economics, 21 (3), 282–300.

[8] Böckerman, P., and P. Ilmakunnas, 2009. Unemployment and self-assessed health: evidence from panel data. Health Economics, 18 (2), 161–179.

[9] Botosaru, I. and C. Muris, 2017. Binarization for panel models with fixed effects. CeMMAP working paper CWP31/17.

[10] Botosaru, I., C. Muris, and K. Pendakur, 2022. Identification of time-varying transformation models with fixed effects, with an application to unobserved heterogeneity in resource shares. Journal of Econometrics, forthcoming.

[11] Carrieri, V., and A.M. Jones, 2017. The income–health relationship 'beyond the mean': New evidence from biomarkers. Health Economics, 26 (7), 937–956.

[12] Carro, J.M., 2007. Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects. Journal of Econometrics, 140 (2), 503–528.

[13] Carro, J.M., and A. Traferri, 2014. State dependence and heterogeneity in health using a bias-corrected fixed-effects estimator. Journal of Applied Econometrics, 29 (2), 181–207.

[14] Chamberlain, G., 1980. Analysis of Covariance with Qualitative Data. The Review of Economic Studies, 47 (1), 225–238.

[15] Chamberlain, G., 1985. Heterogeneity, omitted variable bias, and duration dependence. In: "Longitudinal Analysis of Labor Market Data", edited by J. J. Heckman and B.S. Singer, 3–38. Econometric Society Monographs. Cambridge University Press.

[16] Coe, N.B., and G. Zamarro, 2011. Retirement effects on health in Europe. Journal of Health Economics, 30 (1), 77–86.

[17] Contoyannis, P., A.M. Jones, and N. Rice, 2004. The dynamics of health in the British Household Panel Survey. Journal of Applied Econometrics, 19 (4), 473–503.

[18] Das, M., and A. van Soest, 1999. A panel data model for subjective information on household income growth. Journal of Economic Behavior and Organization, 40 (4), 409–426.

[19] Ettner, S.L., 1996. New evidence on the relationship between income and health. Journal of Health Economics, 15 (1), 67–85.

[20] Eurostat (2017). Methodological Guidelines and Description of EU-SILC Target Variables. 2016 operation.

[21] Evenson, R.J., and R.W. Simon, 2005. Clarifying the relationship between parenthood and depression. Journal of Health and Social Behavior, 46 (4), 341–358.

[22] Fernández-Val, I., Y. Savchenko, and F. Vella, 2017. Evaluating the Role of Income, State Dependence and Individual Specific Heterogeneity in the Determination of Subjective Health Assessments. Economics and Human Biology, 25, 85–98.

[23] Ferri, L., R. Macchioni, M. Maffei, and A. Zampella, 2017. Financial versus sports performance: The missing link. International Journal of Business and Management, 12 (3), 36–48.

[24] Frijters, P., J.P. Haisken-DeNew, and M.A. Shields, 2005. The causal effect of income on health: Evidence from German reunification. Journal of Health Economics, 24 (5), 997–1017.

[25] Galama, T. and A. Kapteyn, 2011. Grossman's missing health threshold. Journal of Health Economics, 30 (5), 1044-1056.

[26] Grossman, M., 1972a. On the concept of health capital and the demand for health. Journal of Political Economy, 80 (2), 223-255.

[27] Grossman, M. 1972b. The Demand for Health: A Theoretical and Empirical Investigation. Columbia University Press.

[28] Gunasekara, F., K. Carter, and T. Blakely, 2011. Change in income and change in self-rated health: Systematic review of studies using repeated measures to control for confounding bias. Social Science and Medicine, 72 (2), 193–201.

[29] Hahn, J., 2001. The information bound of a dynamic panel logit model with fixed effects. Econometric Theory 17 (5), 913–932.

[30] Heckman, J.J., 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process. In: Structural Analysis of Discrete Data with Econometric Applications, C.F. Manski, D. McFadden (eds). MIT Press: Cambridge, MA, 179–195.

[31] Honoré, B. E., and E. Kyriazidou, 2000. Panel Data Discrete Choice Models with Lagged Dependent Variables. Econometrica, 68 (4), 839–874.

[32] Honoré, B. E., and E. Kyriazidou, 2019. Identification in Binary Response Panel Data Models: Is Point-Identification More Common Than We Thought? Annals of Economics and Statistics, 134, 207–226.

[33] Honoré, B. E., C. Muris, and M. Weidner, 2021. Dynamic Ordered Panel Logit Models. arXiv:2107.03253.

[34] Honoré, B.E., and E. Tamer, 2006. Bounds on Parameters in Panel Dynamic Discrete Choice Models. Econometrica, 74 (3), 611-629.

[35] Honoré, B. E., and M. Weidner, 2020. Moment Conditions for Dynamic Panel Logit Models with Fixed Effects. arXiv:2005.05942.

[36] Johnson, E. G., 2004. Panel data models with discrete dependent variables. Doctoral Dissertation, Stanford University.

[37] Johnston, D.W., C. Propper, and M.A. Shields, 2009. Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. Journal of Health Economics, 28 (3), 540–552.

[38] Jones, A.M., and J. Wildman, 2008. Health, income and relative deprivation: Evidence from the BHPS. Journal of Health Economics, 27 (2), 308-324.

[39] Kaplan, R.M., and R.G. Kronick, 2006. Marital status and longevity in the United States population. Journal of Epidemiology and Community Health, 60 (9), 760–765.

[40] Khan, S., F. Ouyang, and E. Tamer, 2021. Inference on semiparametric multinomial response models. Quantitative Economics, 12 (3), 743–777.

[41] Khan, S., M. Ponomareva, and E. Tamer, 2022. Identification of Dynamic Binary Response Models. Working paper.

[42] Kitazawa, Y, 2022. Transformations and moment conditions for dynamic fixed effects logit models. Journal of Econometrics, 229 (2), 350–662.

[43] Larrimore, J., 2011. Does a higher income have positive health effects? Using the earned income tax credit to explore the income-health gradient. The Milbank Quarterly, 89 (4), 694–727.

[44] Lorgelly, P.K., and J. Lindley, 2008. What is the relationship between income inequality and health? Evidence from the BHPS. Health Economics, 17 (2), 249-265.

[45] Mackenbach, J.P., P. Martikainen, C.W. Looman, J.A. Dalstra, A.E. Kunst, and E. Lahelma, 2005. The shape of the relationship between income and self-assessed health: an international study. International Journal of Epidemiology, 34 (2), 286–293.

[46] Magnac, T., 2000. Subsidised training and youth employment: Distinguishing unobserved heterogeneity from state dependence in labour market histories. The Economic Journal, 110 (466), 805-837.

[47] McInerney, M., and J.M. Mellor, 2012. Recessions and seniors' health, health behaviors, and healthcare use: Analysis of the Medicare Current Beneficiary Survey. Journal of Health Economics, 31 (5), 744–751.

[48] Mckenzie, S.K., and K. Carter, 2013. Does transition into parenthood lead to changes in mental health? Findings from three waves of a population based panel study. Journal of Epidemiological Community Health, 67 (4), 339–345.

[49] Molloy, G.J., E. Stamatakis, G. Randall, and M. Hamer, 2009. Marital status, gender and cardiovascular mortality: behavioural, psychological distress and metabolic explanations. Social Science and Medicine, 69 (2), 223–228.

[50] Muris, C., 2017. Estimation in the fixed-effects ordered logit model. Review of Economics and Statistics, 99 (3), 465–477.

[51] Newey, W.K., and D. McFadden, 1994. Large Sample Estimation and Hypothesis Testing. In: Engle, R.F. and McFadden, D., Eds., Handbook of Econometrics, Vol. 4, North Holland, Amsterdam.

[52] Peeters, T., and S. Szymanski, 2014. Financial fair play in European football. Economic Policy, 29 (78), 343-390.

[53] Rigobon, R., 2002. The curse of non-investment grade countries. Journal of Development Economics, 69 (2), 423-449.

[54] Ruhm, C.J., 2000. Are recessions good for your health? The Quarterly Journal of Economics, 115 (2), 617–650.

[55] Ruhm, C.J., 2015. Recessions, healthy no more? Journal of Health Economics, 42, 17–28.

[56] Shi, X., M. Shum, and W. Song, 2018. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. Econometrica, 86 (2), 737–761.

		$P\left(Y_{i,t+1}=y' \middle\vert Y_{i,t}=y\right)$				
	y/y'	1	2	3	4	5
$P\left(Y_{i,t}=y\right)$	1	30.12	47.59	15.06	6.63	0.60
	2	9.61	47.03	34.89	7.21	1.26
	3	1.14	9.73	54.88	29.03	5.22
	4	0.17	1.13	13.79	60.43	24.48
	5	0.02	0.31	3.85	31.95	63.87

Table 1: Current and future self-reported health, United Kingdom.

	σ_n								
	0.01	0.05	0.1	0.25	0.5	1.0	2.0	5.0	10.0
$n = 500$	1.539	0.832	0.824	0.832	0.832	0.831	0.830	0.830	0.830
	(3.241)	(0.493)	(0.368)	(0.329)	(0.319)	(0.317)	(0.317)	(0.317)	(0.317)
$n = 1000$	1.508	0.807	0.785	0.787	0.784	0.784	0.784	0.784	0.783
	(2.475)	(0.435)	(0.309)	(0.248)	(0.233)	(0.231)	(0.231)	(0.231)	(0.231)
$n = 5000$	0.829	0.767	0.751	0.739	0.736	0.735	0.735	0.736	0.736
	(0.352)	(0.149)	(0.106)	(0.092)	(0.088)	(0.086)	(0.085)	(0.085)	(0.085)
$n = 10000$	0.733	0.736	0.732	0.733	0.735	0.735	0.734	0.734	0.734
	(0.259)	(0.108)	(0.079)	(0.058)	(0.053)	(0.052)	(0.052)	(0.052)	(0.052)
$n = 50000$	0.745	0.737	0.737	0.736	0.735	0.736	0.736	0.736	0.736
	(0.102	(0.047)	(0.039)	(0.035)	(0.033)	(0.031)	(0.031)	(0.031)	(0.031)
$n = 100000$	0.725	0.728	0.730	0.732	0.732	0.732	0.732	0.732	0.732
	(0.079	(0.029)	(0.021)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)
$n = 150000$	0.740	0.734	0.733	0.732	0.732	0.732	0.733	0.733	0.733
	(0.056)	(0.026)	(0.022)	(0.020)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
$n = 200000$	0.732	0.732	0.732	0.732	0.732	0.732	0.732	0.732	0.732
	(0.058)	(0.027)	(0.021)	(0.017)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
$n = 260601$	0.735	0.731	0.732	0.732	0.732	0.732	0.732	0.732	0.732
	(0.045)	(0.021)	(0.017)	(0.014)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)

Table 2: Simulation results for autoregressive parameter $\rho = 0.733$. Averages of estimator value across $S = 100$ replications in regular font. Standard deviations are below them, in parentheses.

	σ_n								
	0.01	0.05	0.1	0.25	0.5	1.0	2.0	5.0	10.0
Threshold parameter: $\gamma_2 = -3.275$									
$n = 500$	-15.292	-9.344	-7.853	-7.000	-6.790	-6.695	-6.713	-6.767	-6.744
$n = 1000$	-14.707	-6.704	-5.178	-4.375	-4.295	-4.273	-4.273	-4.268	-4.268
$n = 5000$	-4.276	-3.329	-3.295	-3.280	-3.279	-3.283	-3.286	-3.287	-3.288
$n = 10000$	-3.951	-3.369	-3.356	-3.332	-3.319	-3.313	-3.313	-3.313	-3.313
$n = 50000$	-3.325	-3.287	-3.282	-3.275	-3.275	-3.275	-3.275	-3.275	-3.275
$n = 100000$	-3.314	-3.270	-3.268	-3.268	-3.270	-3.271	-3.271	-3.272	-3.272
$n = 150000$	-3.303	-3.292	-3.289	-3.286	-3.283	-3.281	-3.281	-3.280	-3.280
$n = 200000$	-3.286	-3.275	-3.274	-3.272	-3.272	-3.272	-3.272	-3.272	-3.272
$n = 260601$	-3.283	-3.280	-3.278	-3.277	-3.277	-3.277	-3.277	-3.277	-3.277
Threshold parameter: $\gamma_4 = 3.326$									
$n = 500$	13.420	9.837	8.536	7.501	7.426	7.347	7.298	7.284	7.282
$n = 1000$	12.091	5.524	4.964	4.865	4.767	4.763	4.738	4.740	4.740
$n = 5000$	5.477	3.454	3.375	3.363	3.371	3.373	3.373	3.372	3.372
$n = 10000$	4.456	3.407	3.358	3.340	3.330	3.324	3.321	3.320	3.320
$n = 50000$	3.411	3.336	3.333	3.331	3.330	3.330	3.330	3.330	3.330
$n = 100000$	3.324	3.341	3.342	3.340	3.339	3.340	3.340	3.340	3.340
$n = 150000$	3.375	3.339	3.333	3.332	3.331	3.331	3.331	3.331	3.331
$n = 200000$	3.338	3.329	3.330	3.329	3.329	3.329	3.329	3.329	3.329
$n = 260601$	3.320	3.327	3.327	3.327	3.326	3.326	3.326	3.326	3.326

Table 3: Simulation results for CCMLE of threshold parameters $\gamma_2 = -3.275$ and $\gamma_4 = 3.326$. Reported values are averages over $S = 100$ simulation replications.

	σ_n								
	0.01	0.05	0.1	0.25	0.5	1.0	2.0	5.0	10.0
$n = 500$	0.98	0.98	0.95	0.97	0.99	0.99	0.98	0.99	0.99
	23.0	6.5	4.1	3.1	2.7	2.6	2.5	2.5	2.5
$n = 1000$	0.95	0.96	0.97	0.95	0.95	0.96	0.95	0.96	0.96
	14.2	3.6	2.3	1.5	1.3	1.2	1.2	1.2	1.2
$n = 5000$	0.99	0.99	0.96	0.92	0.93	0.93	0.94	0.94	0.94
	4.6	1.5	1.1	0.8	0.7	0.7	0.6	0.6	0.6
$n = 10000$	0.92	0.92	0.95	0.96	0.94	0.94	0.93	0.93	0.92
	2.4	0.9	0.6	0.5	0.4	0.4	0.4	0.4	0.4

Table 4: Simulation results for the 95% confidence interval for β_1 based on a normal approximation with bootstrap standard errors from 100 bootstrap replications. Results are based on $S = 100$ simulation replications. Coverage probabilities are in regular font. Confidence interval width is in small font below.

				σ_n		
	(ρ_2, ρ_3, ρ_4)	β_1	β_6	ρ	γ_2	γ_4
		0.049	-0.370		-3.275	3.326
$\delta = 0$	$(0, 0.733, 0.733)$	0.004	-0.374	0.738	-3.274	3.346
		(0.098	(0.265)	(0.063)	(0.180)	(0.196)
$\delta = 0.1$	$(0.1, 0.733, 0.833)$	0.004	-0.376	0.678	-3.325	3.377
		(0.100)	(0.254)	(0.064)	(0.177)	(0.201)
$\delta = 0.2$	$(0.2, 0.733, 0.933)$	0.007	-0.362	0.624	-3.352	3.441
		(0.092)	(0.254)	(0.061)	(0.176)	(0.200)
$\delta = 0.3$	$(0.3, 0.733, 1.033)$	0.014	-0.328	0.569	-3.389	3.479
		(0.094)	(0.254)	(0.060)	(0.179)	(0.207)
$\delta = 0.4$	$(0.4, 0.733, 1.133)$	0.018	-0.322	0.511	-3.409	3.540
		(0.093)	(0.250)	(0.062)	(0.177)	(0.207)

Table 5: Simulation results for our sensitivity analysis for a model with general dynamics, with $n = 10000$ and $\sigma_n = 1$.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
	DOLFE	DOLFE	DOLFE	DOLFE	FEOL	DOL	OL	FELM	LM
	$\sigma_n = 1$	$\sigma_n = 0.1$	$\sigma_n = 10$	$\sigma_n = 1$					
log(income)	0.049	-0.047	0.059	0.061	0.020	0.340	0.492	0.003	0.194
	(0.033)	(0.056)	(0.029)	(0.029)	(0.019)	(0.004)	(0.004)	(0.003)	(0.002)
child	-0.030	0.006	-0.031	-0.026	0.021	0.060	0.089	0.002	0.033
	(0.051)	(0.069)	(0.050)	(0.049)	(0.032)	(0.005)	(0.005)	(0.005)	(0.002)
married	0.139	-0.041	0.157	0.130	0.164	0.073	0.141	0.029	0.062
	(0.087)	(0.119)	(0.086)	(0.088)	(0.053)	(0.007)	(0.008)	(0.009)	(0.003)
unemp	-0.188	-0.230	-0.178		-0.196	-0.242	-0.308	-0.033	-0.127
	(0.070)	(0.110)	(0.068)		(0.038)	(0.014)	(0.015)	(0.007)	(0.006)
retired	-0.132	-0.043	-0.139		-0.154	-0.050	-0.097	-0.027	-0.047
	(0.082)	(0.119)	(0.080)		(0.041)	(0.010)	(0.011)	(0.007)	(0.004)
other	-0.370	-0.207	-0.369		-0.473	-0.771	-1.087	-0.082	-0.460
	(0.061)	(0.087)	(0.061)		(0.040)	(0.010)	(0.012)	(0.007)	(0.005)
ρ	0.733	0.723	0.733	0.734		1.987			
	(0.020)	(0.025)	(0.020)	(0.017)		(0.023)			
γ_2	-3.275	-3.260	-3.272	-3.211	-3.487	-2.506	-1.992		
	(0.054)	(0.068)	(0.053)	(0.048)	(0.015)	(0.007)	(0.006)		
γ_4	3.326	3.356	3.329	3.321	3.997	3.089	2.603		
	(0.055)	(0.076)	(0.055)	(0.054)	(0.024)	(0.006)	(0.006)		

Table 6: Main results.