

# Estimation of Dynamic Nonlinear Random Effects Models with Unbalanced Panels\*

Pedro Albarran<sup>†</sup>      Raquel Carrasco<sup>‡</sup>      Jesus M. Carro<sup>§</sup>

May 2018

## Abstract

This paper presents estimation methods for dynamic non-linear models with correlated random effects (CRE) when having unbalanced panels. Unbalancedness is often encountered in applied work and ignoring it in dynamic non-linear models produces inconsistent estimates even if the unbalancedness process is completely at random. We show that selecting a balanced panel from the sample can produce efficiency losses or even inconsistent estimates of the average marginal effects. We allow the process that determines the unbalancedness structure of the data to be correlated with the permanent unobserved heterogeneity. We discuss how to address the estimation by maximizing the likelihood function for the whole sample and also propose a Minimum Distance approach, which is computationally simpler and asymptotically equivalent to the Maximum Likelihood estimation. Our Monte Carlo experiments and empirical illustration show that the issue is relevant. Our proposed solutions perform better both in terms of bias and RMSE than the approaches that ignore the unbalancedness or that balance the sample.

*JEL classification:* C23, C25

*Keywords:* Unbalanced Panels, correlated random effects, dynamic non-linear models.

---

\*We are grateful to César Alonso-Borrego, Wiji Arulampalam, Ricardo Mora, Jeff Wooldridge, and participants at the 21st International Panel Data Conference, the IAAE 2015 Annual Conference, the XXXIX SAEe and Econometrics Seminar at Oxford University for helpful comments on this work. All remaining errors are our own. The authors gratefully acknowledge research funding from the Spanish Ministry of Education, Grants ECO2012-31358 and ECO2015-65204-P.

<sup>†</sup>Department of Economics (FAE), Universidad de Alicante. E-mail: albarran@ua.es

<sup>‡</sup>Department of Economics, Universidad Carlos III de Madrid. E-mail: rcarras@eco.uc3m.es

<sup>§</sup>Corresponding Author. Department of Economics, Universidad Carlos III de Madrid. E-mail: jcarro@eco.uc3m.es

# 1 Introduction

The purpose of this paper is to present and evaluate estimation methods for dynamic non-linear models with correlated random effects (CRE) when the panel data are unbalanced.<sup>1</sup> Unbalanced panels are often encountered in applied work. For example, in large households panel data sets like the PSID for the U.S or the GSOEP for Germany, some individuals drop out (potentially non-randomly) of the sample. At a firm level, Compustat and Datastream International also have an unbalanced structure. In other cases, like in the so called “rotating panels”, the unbalancedness is generated by the sample design (for instance, in the Monthly Retail Trade Survey for the U.S, or in the Household Budget Continuous Survey for Spain).

It is well-known how to estimate CRE dynamic non-linear models with balanced panels. However, the existing estimation methods cannot be in general directly implemented with unbalanced panels. Ignoring the unbalancedness produces inconsistent estimates, as we will discuss. Obtaining a balanced sub-sample from the unbalanced panel, so that the existing CRE methods for balanced panels could then be used, is also problematic. If we balance the sample by taking a subset of individuals that are observed over the same periods, we are making an endogenous selection of the sample unless the unbalancedness is independent of the individual effects. Another possibility to balance the sample is to take the subset of periods at which all individuals are observed (see Wooldridge, 2005). But this is in some cases infeasible because of the lack of enough number of common periods across individuals and, when feasible, it implies important efficiency losses.

In a dynamic setting under the CRE approach the so-called “initial conditions problem” arises. Heckman (1981) and Wooldridge (2005) propose solutions to deal with it, but these are developed only for balanced panels. Furthermore, the initial conditions problem is exacerbated when the panel is unbalanced because it affects to each first period of observation in the data set. This implies that, as we will show, even assuming that unbalancedness is completely at random is not enough to allow us to ignore it in the

---

<sup>1</sup>The CRE approach has been found useful to estimate nonlinear dynamic models in many cases, because it is not subject to the incidental parameters problem that the fixed-effects (FE) approach suffers and it does not require a large number of periods. Examples of applications using CRE are Hyslop (1999), Contoyannis et. al. (2004), Stewart (2007), and Akee et. al. (2010). Arellano and Honoré (2001) and Arellano (2003) discuss the general literature of non-linear dynamic models with unobserved heterogeneity, using both the FE and the CRE approaches.

estimation.<sup>2</sup>

We propose methods to deal with the unbalancedness structure of the data in the estimation of models with lags of the endogenous variable and other explanatory variables that are strictly exogenous. We consider unbalancedness processes that are independent of the time-varying shocks, but allow them to be correlated with the time-invariant unobserved heterogeneity. Therefore we are not restricted to the case of unbalancedness completely at random. We first discuss how to address the unbalancedness problem by maximizing the likelihood function for the whole sample. This can be computationally cumbersome because specific parameters to each sub-panel need to be estimated jointly with the common parameters of the model. We then propose to estimate the model for each subpanel separately and then to obtain estimates of the common parameters across subpanels by minimum distance (MD). This method allows to use the same estimation routines that we would use if we had a balanced panel, while keeping the good asymptotic properties of the maximum likelihood (ML) estimator for the whole sample.

A simulation study shows that these methods perform well compared to other alternatives both in terms of bias and RMSE. As an empirical illustration, we estimate an export participation equation with dynamic effects using unbalanced data for Spanish manufacturing firms. Our results show that the unbalancedness issue is relevant in practice, and there is evidence of unbalancedness correlated with the unobserved heterogeneity.

To the best of our knowledge only Wooldridge (2010) addresses the issue of estimating CRE models with unbalanced panels, but considering only static models with strictly exogenous variables. He proposes several strategies for allowing the time invariant unobserved heterogeneity to be correlated with the observed covariates and the selection mechanism for unbalanced panels. However, the assumption of lack of dynamic effects is very restrictive, and the solutions in Wooldridge (2010) cannot be directly extended to dynamic models because the unbalancedness also affects how to deal with the initial conditions problem.

Although unbalanced panels could be seen as a particular case of missing data, the problem we address cannot be solved using the existing literature on panel models with

---

<sup>2</sup>This problem also affects RE models assuming that the time invariant unobserved heterogeneity is independent of the time-varying covariates. Examples of papers using the RE approach are Arulampalam and Stewart (2009), Campa (2004), or Bernard and Jensen (2004). The CRE model that we cover in detail contains the RE as a particular case.

missing data. One strand of this literature relies on missingness at random and on using moment conditions that are valid both with complete and with missing data (e.g. Pacini and Windmeijer, 2015). In our case, the sets of moments conditions (the first order conditions of the likelihood) under complete data are not valid. The reason is that the likelihood for complete balanced panels does not account for the different initial conditions, or for the potential relation between the unbalancedness and permanent unobserved heterogeneity, as we will show. Other strand of the literature relies on having some variables upon which you can condition to make the missing process conditionally independent of the main model (e.g. Wooldridge, 2007), or on having additional information and assumptions about the missing process (e.g. Bhattacharya, 2008). In contrast with that, we do not assume anything about the relation of the missing process and observable variables, nor have any additional information related to the missing process. Also, the moment conditions for panel data models considered in this literature are based on the fixed effects approach so they do not deal with the initial conditions problem, which is crucial in the unbalanced case.

The rest of the paper is organized as follows. Section 2 presents the general model and the likelihood functions that account for the unbalancedness. Section 3 formalizes the existing approaches, that is, those ignoring the unbalancedness and making the sample balanced, and discuss the restrictive conditions under which they could work. Section 4 presents the ML and MD estimators for the model that accounts for the unbalancedness. Section 5 describes how the estimation can be implemented and several practical issues. In Section 6 we study the finite sample properties of the different estimators by means of Monte Carlo simulations. In Section 7, as an empirical illustration, we estimate an export market participation equation using firms' level data. Finally, Section 8 concludes.

## 2 General framework

We present a general approach that can be applied to dynamic non-linear panel data models. Let us denote

$$\begin{aligned} Y_i &= (y_{i1}, \dots, y_{iT})', \\ X_i &= (X'_{i1}, \dots, X'_{iT})', \\ S_i &= (s_{i1}, \dots, s_{iT})', \end{aligned}$$

where  $i = 1, \dots, N$  represents cross-sectional units,  $y_{it}$  is the (scalar) outcome, and  $X_{it}$  is a row vector of dimension  $K$  of covariates. The possibility of having an unbalanced panel is captured through a set of selection indicators,  $s_{it}$ , which take the value 1 if unit  $i$  is observed in period  $t$ , that is

$$s_{it} = \begin{cases} 1 & \text{if } y_{it} \text{ and } X_{it} \text{ are observed} \\ 0 & \text{otherwise.} \end{cases}$$

Notice that the balanced situation can be seen as a particular case of this setting, when  $s_{it} = 1$  for all  $i$  and  $t$ . We only consider cases in which either both  $y_{it}$  and  $X_{it}$  are observed or both are not observed. We define  $t_i$  as the first period in which unit  $i$  is observed, i.e.

$$t_i = \{t: s_{it} = 1 \text{ and } s_{ij} = 0 \forall j < t\},$$

and  $T_i$  as the number of periods we observe for unit  $i$ ,

$$T_i = \sum_{t=1}^T s_{it}$$

Another characteristic of the panels considered is that all the observations for unit  $i$  are consecutive.<sup>3</sup> This means that

$$\begin{aligned} s_{it} &= 1 \forall t \in [t_i, t_i + T_i - 1] \\ s_{it} &= 0 \forall t < t_i \text{ or } t > t_i + T_i - 1. \end{aligned}$$

Let  $M_i$  be the  $(T_i \times T)$  matrix that select the set of  $X_i$  that we observe, that is,  $M_i X_i = (X'_{it_i}, \dots, X'_{it_i+T_i-1})'$ . The element  $(j, k)$  of  $M_i$ ,  $m_{i,(jk)}$ , is

$$m_{i,(jk)} = \begin{cases} 1 & \text{if } s_{ik} = 1 \text{ and } j = k - t_i + 1 \\ 0 & \text{otherwise.} \end{cases}$$

If the panel is balanced,  $M_i$  is the identity matrix. Note that  $S_i = \iota'_{T_i} M_i$  where  $\iota_{T_i}$  is a vector of ones with dimension  $T_i$ . We denote by  $J$  the number of different  $S_i$  sequences that we have in the total panel. We refer to the sub-set of units with the same sequence  $S^{(j)}$  as “sub-panel”  $j$ ,  $j = 1, \dots, J$ . In other words, subpanel  $j$  contains all the individuals  $i$  such that  $S_i = S^{(j)}$ . Finally, we consider panels where  $N$  is large and  $T$  and  $J$  are small relative to  $N$ .

---

<sup>3</sup>At the end of this section we discuss how a situation with non-consecutive observations could be addressed.

The kind of models we consider in this paper are as follows. For all  $i$  and  $t$ ,  $y_{it}$  is determined as

$$y_{it} = g(y_{it-1}, X_{it}, \eta_i, \varepsilon_{it}),$$

where  $g(\cdot)$  is a non-linear function, non-additively separable in its latent terms, whose form is known up a vector of parameters that characterized it. For simplicity, we focus on a model with one lag of  $y_{it}$  and contemporaneous values of  $X$ , which is the most frequently model used in the applied literature. However, all our analysis could be extended to higher order chains, or to cases that include in  $X_{it}$  previous values of these strictly exogenous covariates.  $\eta_i$  denotes the vector of permanent unobserved heterogeneous characteristics or "individual effect", and  $\varepsilon_{it}$  are period-specific disturbances that are assumed to be independent and identically distributed across both  $i = 1, \dots, N$  and  $t = 1, \dots, T$  with known distribution. Also  $\varepsilon_{it}$  are independent of  $\eta_i$  and  $X_i$ . This means that we consider models where  $X$  are strictly exogenous covariates with respect to the period-specific unobservables,  $\varepsilon$ , but they can be correlated with the time-constant unobservables,  $\eta_i$ . The function  $g(\cdot)$  together with the distribution of  $\varepsilon$  give the conditional distribution  $F(y_{it} | y_{it-1}, X_i, \eta_i)$  which is our primary object of interest and whose parameters will be estimated. The previous assumptions imply that

$$F(y_{it} | y_i^{t-1}, X_i, \eta_i) = F(y_{it} | y_{it-1}, X_{it}, \eta_i),$$

where  $y_i^{t-1} = (y_{i1}, \dots, y_{it-1})$ .

So far this is a standard model in the (balanced) panel data literature. The new important assumption related to the unbalancedness that we are going to maintain throughout this paper is

**Assumption 1:**

$$\varepsilon_{it} \perp S_i | (\eta_i, X_i) \text{ for all } i \text{ and } t$$

This implies that  $\varepsilon$  is conditionally independent of the sample selection process  $S_i$  that produces the unbalancedness. However, note that this assumption is not restricting the relation between  $S_i$  and  $(\eta_i, X_i)$ . This means that, although we are not going to consider an endogenous selection process (endogenous with respect to the period-specific disturbances), we are going to allow  $S_i$  to be correlated with the unobserved permanent characteristics  $\eta_i$ .

Let  $f(y_{it} \mid y_{it-1}, X_{it}, \eta_i, S_i; \beta)$  be the correctly specified density for the conditional distribution  $F(y_{it} \mid y_i^{t-1}, X_i, \eta_i)$  under assumption 1, and  $h(\eta_i \mid M_i X_i, S_i; \beta_{\eta S_i})$  the correctly specified density of the distribution  $\eta_i \mid M_i X_i, S_i$ .<sup>4</sup> Then, the density of  $(s_{i1}y_{i1}, \dots, s_{iT}y_{iT})$  for a given individual is

$$\begin{aligned} f(s_{i1}y_{i1}, \dots, s_{iT}y_{iT} \mid M_i X_i, S_i) &= \prod_{t=1}^T f(y_{it} \mid s_{it-1}y_{it-1}, M_i X_i, S_i)^{s_{it}s_{it-1}} f(y_{it} \mid M_i X_i, S_i)^{s_{it}(1-s_{it-1})} \\ &= \prod_{t=t_i+1}^{t_i+T_i-1} f(y_{it} \mid y_{it-1}, M_i X_i, S_i) f(y_{it_i} \mid M_i X_i, S_i). \end{aligned} \quad (1)$$

Previous equation can be written as

$$\begin{aligned} f(s_{i1}y_{i1}, \dots, s_{iT}y_{iT} \mid M_i X_i, S_i) &= \\ \int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i-1} f(y_{it} \mid y_{it-1}, M_i X_i, S_i, \eta_i; \beta) f(y_{it_i} \mid M_i X_i, S_i, \eta_i; \lambda_{S_i}) h(\eta_i \mid M_i X_i, S_i; \beta_{\eta S_i}) d\eta_i, \end{aligned} \quad (2)$$

or as

$$\begin{aligned} f(s_{i1}y_{i1}, \dots, s_{iT}y_{iT} \mid M_i X_i, S_i) &= \\ \left[ \int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i-1} f(y_{it} \mid y_{it-1}, M_i X_i, S_i, \eta_i; \beta) h(\eta_i \mid y_{it_i}, M_i X_i, S_i; \pi_{\eta S_i}) d\eta_i \right] f(y_{it_i} \mid M_i X_i, S_i), \end{aligned} \quad (3)$$

depending on whether we integrate out the unobserved effect by specifying the density for the first observation in each sub-panel conditional on the unobserved effect and the density of the unobserved effect, or we specify the density of the unobserved effect conditional on the first observation. Note that in equation (3) we can discard  $f(y_{it_i} \mid M_i X_i, S_i)$  because this term is outside the integral.

Given that previous equations depend on the unobservable  $\eta_i$ , if the first period observed in the sample does not coincide with the beginning of the stochastic process, the first observation will not be independent of  $\eta_i$ . Moreover,  $f(y_{it_i} \mid M_i X_i, S_i, \eta_i; \lambda_{S_i})$  and  $h(\eta_i \mid M_i X_i, S_i; \beta_{\eta S_i})$  in equation (2), or  $h(\eta_i \mid y_{it_i}, M_i X_i, S_i; \pi_{\eta S_i})$  in equation (3) are different for each sub-panel with different observed periods  $S_i$ . Writing an equation for  $f(y_{i1} \mid X_i, \eta_i)$  and  $h(\eta_i \mid X_i)$ , or for  $h(\eta_i \mid y_{i1}, X_i)$ , as Heckman (1981) and Wooldridge (2005) did respectively for the balanced case, is not enough to solve the initial conditions problem for three reasons:

- (i) The conditioning set of covariates is different for each  $S_i$ , as in static models (see Wooldridge, 2010).

---

<sup>4</sup>In our notation for defining functions, the set of parameters of that function appear after a semicolon.

- (ii) The initial observation is different for each  $S_i$ . Thus, even in a model without  $X$  covariates, without further assumptions  $f(y_{it_i}|\eta_i = \eta, S_i) \neq f(y_{rt_r}|\eta_r = \eta, S_r)$  for  $t_i \neq t_r$ .
- (iii) Last but not least, even if the starting period  $t_i$  is the same, there may be correlation between  $S_i$  and the individual characteristics making the distributions of  $\eta_i$  different for each  $S_i$ .

The framework considered so far includes situations in which the selection mechanism  $S_i$  is correlated with the individual effect,  $\eta_i$ . This implies that, if we write the likelihood of the data using expression (2), different distributions of the initial conditions and of the unobserved effects for each sub-panel are required. That is, the densities  $f(y_{it_i}|M_i X_i, S_i, \eta_i; \lambda_{S_i})$  and  $h(\eta_i|M_i X_i, S_i; \beta_{\eta_{S_i}})$  in (2) depend on a vector of parameters different for each sub-panel. Likewise, in equation (3), we need to specify the density of  $\eta_i$  conditional on the initial observation,  $h(\eta_i|y_{it_i}, M_i X_i, S_i; \pi_{S_i})$  and this will have different parameters for each sub-panel.

Also, notice that there are two sets of coefficients in the conditional distribution of interest. One set is  $\beta$ , which represent the parameters that are homogeneous across individuals. The other set is  $\eta_i$  that are heterogeneous coefficients whose distribution is used to integrate them out. The framework considered in this paper applies to any random coefficients model by varying what is included in each set: from the typical case with only the constant term being individual specific, to more general situations where several or even all the coefficients are heterogeneous.

**Unbalancedness independent of the individual effect:** If in addition to **Assumption 1**, we assume,

**Assumption 2:**  $S_i$  is independent of  $\eta_i$  given  $X$ ,

so that  $h(\eta_i|X, S_i) = h(\eta_i|X)$  for any set of periods included in  $X$ . This assumption is relevant, for instance, when having rotating panels. However, even under Assumption 2  $f(y_{it_i}|X_i, \eta_i, S_i)$  is different for each  $S_i$  simply because the process has been running a different number of periods until that first observation, and we are not assuming that the process is on steady state. Likewise, the density of the unobserved effects conditional on the initial conditions,  $h(\eta_i|y_{it_i}, X_i, S_i)$  will be, in general, different for each  $t_i$ . In addition



to that, Assumption 2 is not enough to guarantee that  $h(\eta_i|M_iX_i, S_i)$  is the same for all  $S_i$  because, even if  $S_i$  is independent of  $\eta_i$  given  $X$ ,  $\eta_i$  can still depend on  $X$  and there will be a different conditioning set of observations in  $M_iX_i$  for each  $S_i$ .

**Other unbalancedness patterns:** Some panels present unbalancedness structures that include individuals with non-consecutive observations. In these cases, we could integrate out the holes using the conditional model of  $y_{it}$ . The main limitation of this approach is that it cannot be applied when the model for  $y_{it}$  has  $X$  covariates that are not observed when  $s_{it} = 0$ , as we consider in this paper. We would need to make further assumptions about  $X_{it}$  to be able to either integrate over the missing  $X$  too, or to impute their values for the missing periods so that we can integrate the unobserved  $y_{it}$  using our model  $F(y_{it} | y_{it-1}, X_{it}, \eta_i)$ . In any case, these will require to make assumptions about  $X_{it}$  that would never be considered with balanced panels. This is out of the scope of this paper, which is to estimate the model we would specify if having a balanced sample. Nonetheless, if those further assumptions about  $X$  are made, the approach in these paper could still be adapted.

### 3 Existing approaches

#### 3.1 Ignoring the unbalancedness

In this subsection we study under which assumptions it is possible to ignore the unbalancedness and to treat the data as if they were balanced. That is, we study when it is possible to use the following likelihood that ignores the unbalancedness,

$$\begin{aligned} & \sum_{i=1}^N \log (f (s_{i1}y_{i1}, \dots, s_{iT}y_{iT}|M_iX_i)) \\ &= \sum_{i=1}^N \log \left( \int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i-1} f (y_{it}|y_{it-1}, M_iX_i, \eta_i; \beta) f (y_{it_i}|M_iX_i, \eta_i; \delta) h (\eta_i|M_iX_i; \beta_\eta) d\eta_i \right), \end{aligned} \tag{4}$$

instead of the density given by equation (2).

Note that even though we ignore that each individual belongs to a different sub-panel  $S_i$  and is observed for the first time at a different moment in time, we have to write the likelihood based on the  $M_iX_i$  observations that we have for each individual. The restrictions coming from ignoring the unbalancedness are that the distribution of the

initial observation is the same for all, and that the distribution functions given  $M_i X_i$  are also the same. Given that under Assumption 1 the sample selection process  $S_i$  is strictly exogenous with respect to the idiosyncratic shocks to  $y_{it}$  we have that

$$f(y_{it}|s_{it-1}y_{it-1}, M_i X_i, S_i, \eta_i) = f(y_{it}|y_{it-1}, X_i, \eta_i). \quad (5)$$

In order to have (2) and (4) leading to the equivalent Maximum Likelihood (ML) Estimators of the parameters of the conditional distribution of  $y_{it}|y_{it-1}, M_i X_i, \eta_i$  we need the following conditions:

1. Assumption 2, i.e. for any given  $X$ ,  $h(\eta_i|X, S_i) = h(\eta_i|X)$ , so that  $h(\eta_i|M_i X_i, S_i) = h(\eta_i|M_i X_i)$ .
2.  $h(\eta_i|M_i X_i)$  must be a function common to all  $S_i$ , so that its value changes only as the values of  $X$  at which it is evaluated change (but not as a function of the specific periods at which  $X_i$  is observed).
3. The process is in the steady state (or the initial observations  $y_{t_i}$  come from the same exogenous distribution or rule for all units and  $t_i$ ).
4.  $S_i$  is independent from the shocks to the initial conditions.

Conditions 1 and 2 together imply that  $h(\eta_i|M_i X_i, S_i; \beta_{\eta S_i}) = h(\eta_i|M_i X_i; \beta_{\eta})$  for all  $i$  as (4) imposes, where the common function  $h(\eta_i|M_i X_i; \beta_{\eta})$  must be able to accommodate any differences coming from the different conditioning set ( $M_i X_i$ ) across individuals.

Condition 2 is very restrictive because, for example, in general  $Var(\eta|M_i X_i)$  will be different if the number periods in which  $x_{it}$  are observed is different.<sup>5</sup> A case in which this condition is trivially satisfied is when  $\eta_i$  is independent of  $X_i$ . Condition 2 does not apply when there are no covariates  $X$  in the model.

Conditions 3 and 4 are needed to ensure that all units have the same distribution for the initial condition regardless the period  $t_i$  at which they enter the panel, i.e.

$$f(y_{it_i}|M_i X_i, S_i, \eta_i; \lambda_{S_i}) = f(y_{i1}|M_i X_i, \eta_i; \delta) = \dots = f(y_{i \max t_j}|M_i X_i, \eta_i; \delta) \text{ for all } i,$$

as (4) imposes.

---

<sup>5</sup>The condition is not violated, however, if  $Var(\eta|M_i X_i)$  changes with the number of periods of  $x_{it}$  observed in a deterministic way, e.g.  $Var(\eta|M_i X_i) = \frac{\sigma_{\eta}^2}{\iota_{T_i} M_i \iota_{T_i}}$ .

Unless these four conditions are all satisfied, the estimates of  $\beta$  obtained by ignoring the unbalancedness are inconsistent.

Notice that the assumption that  $S_i$  is independent from the shocks to the initial conditions (condition 4) is not enough to ensure that the conditional densities for each initial observational period coincide. Also notice that unbalancedness completely at random (i.e. independent of everything else) is not enough to allow us to ignore it. For example, suppose that we have two individuals in a process without covariates  $X$ , that  $y_{it}$  starts in  $y_{i0}$  for both  $i$ , and that both follow the same process for  $y_{it}$ . However, we start observing one individual in period  $t_i = 1$  and the other in period  $t_i = 2$ , and this is decided randomly. Therefore, we are in a case in which  $S_i$  is determined completely at random, and in which condition 2 is not needed because there are no covariates. Then,

$$\Pr(y_{i1} | \eta_i, S_i) = \sum_{y_{i0}} \Pr(y_{i1} | y_{i0}, \eta_i, S_i) \cdot \Pr(y_{i0} | \eta_i, S_i) = \sum_{y_{i0}} \Pr(y_{i1} | y_{i0}, \eta_i) \cdot \Pr(y_{i0} | \eta_i) \quad (6)$$

$$\Pr(y_{i2} | \eta_i, S_i) = \sum_{y_{i1}} \Pr(y_{i2} | y_{i1}, \eta_i) \cdot \Pr(y_{i1} | \eta_i). \quad (7)$$

The two probabilities are different unless  $y_{i1}$  is at the steady state (Condition 3).

### 3.2 Using a subset of periods at which all individuals are observed

Wooldridge (2005) points out that a potential solution to the unbalancedness under Assumption 1 is to use the subset of periods constituting a balanced panel. Then, one could apply to that balanced sample the standard solutions to the initial conditions problem. Nonetheless, this approach has two limitations: (i) it discards useful information leading to an efficiency loss, and (ii) the balanced sample may not contain enough number of common periods across individuals, making the estimation infeasible.<sup>6</sup>

Suppose that the correct conditional density of  $s_{i1}y_{i1}, \dots, s_{iT}y_{iT} | M_i X_i, S_i$  is given by (3), excluding the term for the initial observations  $f(y_{it_i} | M_i X_i, S_i)$ . Instead of that, the

---

<sup>6</sup>For example, in a rotating panel with  $T = 5$  with three sub-panels where each sub-panel lasts for three periods (i.e.  $T_i = 3$ ), the first sub-panel starts at  $t_i = 1$ , the second at  $t_i = 2$ , and the third at  $t_i = 3$ , the sub-panels only have one period in common, less than the 3 periods needed for estimation.

following likelihood function is maximized

$$\begin{aligned} & f(y_{it_m}, \dots, y_{iT_m} | M_i X_i) \\ &= \int \prod_{\eta_i, t=t_m}^{T_m} f(y_{it} | y_{it-1}, X_{it_m}^{T_m}, \eta_i) h(\eta_i | y_{i \max t_i}, X_{it_m}^{T_m}) d\eta_i, \end{aligned} \quad (8)$$

where  $t_m \equiv \max_{j \in [1, N]} t_j + 1$ , and  $T_m \equiv \min_{j \in [1, N]} (t_j + T_j - 1)$ . Under Assumption 1  $f(y_{it} | y_{it-1}, X_{it_m}^{T_m}, S_i, \eta_i) = f(y_{it} | y_{it-1}, X_{it_m}^{T_m}, \eta_i)$ . Thus, to have a consistent ML Estimator of the parameters of the conditional distribution of  $y_{it} | y_{it-1}, M_i X_i, \eta_i$  based on (8) we need

$$\begin{aligned} & h(\eta_i | y_{i \max_{j \in [1, N]} t_j}, X_{it_m}^{T_m}) \\ &= \sum_{j=1}^J h(\eta_i | y_{i \max_{j \in [1, N]} t_j}, X_{it_m}^{T_m}, S_i = S^{(j)}) \Pr(S_i = S^{(j)} | y_{i \max_{j \in [1, N]} t_j}, X_{it_m}^{T_m}), \end{aligned} \quad (9)$$

where  $S^{(j)}$  is the  $j$ -th element of the set of  $J$  different  $S_i$  sequences that we have in the panel, and  $X_{it_m}^{T_m} = (X_{it_m}^t, \dots, X_{i, T_m}^t)'$ . So, as long as the  $h(\eta_i | y_{i \max_{j \in [1, N]} t_j}, X_{it_m}^{T_m})$  we specify satisfies this condition and we have enough periods in the balanced sample, the MLE based on (8) will be consistent, though less efficient. However, depending on the nature of  $h(\eta_i | y_{i \max_{j \in [1, N]} t_j}, M_i X_i, S_i)$  (i.e. depending on the nature of the relations between  $\eta_i$  and  $S_i$  and the evolution of the distribution of  $y_{it}$  across periods and sub-panels) approximating  $h(\eta_i | y_{i \max t_i}, X_{i \max t_i + 1}^{\min(t_i + T_i - 1)})$  may require a complex distribution even if  $h(\eta_i | y_{i \max_{j \in [1, N]} t_j}, M_i X_i, S_i)$  is the standard normal distribution.<sup>7</sup>

### 3.3 Using a subset of individuals that are observed the same number of periods

Another possibility to deal with the unbalancedness is to take one single sub-panel from the total sample. This sub-panel is, by definition, a balanced panel where all the individuals are observed from the same first period to the same last period. In many cases, this would be the sub-sample of individuals present in all the waves of the original panel. For example, Contoyannis et. al. (2004) obtain a balanced sub-sample in this way. More generally, one can take the sub-set of individuals observed only in all of some specific consecutive waves.

Although this way of obtaining a balanced sample produces an efficiency loss due to discarding a potentially high proportion of the sample, it avoids the infeasibility of

---

<sup>7</sup>See Section 5.1 for a discussion of the problems with the practical implementation of this approach.

the previous balancing method and may consistently estimate the common parameters of the model. However, this does not allow to identify the average marginal effect of covariates because the conditional distribution of the heterogeneous individual effects will only be valid for this particular sub-group of individuals. Unless the unbalancedness is independent of  $\eta_i$  (Assumption 2) and of  $X$ , the distribution of  $\eta_i$  for this balanced sub-sample is different from the distribution of  $\eta_i$  for the entire sample. And the marginal effects, which are the ultimate parameters of interest, are a function of the distribution of  $\eta_i$ . Therefore, unless  $S_i$  is independent of  $\eta_i$  and of  $X_i$ , the average marginal effects we estimate for this sub-sample are not a consistent estimation of the average marginal effects for the entire population of individuals.

## 4 Estimation with the entire sample

### 4.1 Maximum Likelihood Estimation

The models that accounts for unbalancedness explained in Section 2 can be estimated by Maximum Likelihood (ML). The log-likelihood, if the model specifies the terms in (2), is given by

$$\mathcal{L} = \sum_{i=1}^N \log \int_{\eta_i} \left[ \prod_{t=t_i+1}^{t_i+T_i-1} f(y_{it}|y_{it-1}, M_i X_i, S_i, \eta_i; \beta) f(y_{it_i}|M_i X_i, S_i, \eta_i; \lambda_{S_i}) h(\eta_i|M_i X_i, S_i; \beta_{\eta S_i}) \right] d\eta_i. \quad (10)$$

If the model specifies the terms in (3), then the log-likelihood is given by

$$\mathcal{L} = \sum_{i=1}^N \log \int_{\eta_i} \left[ \prod_{t=t_i+1}^{t_i+T_i-1} f(y_{it}|y_{it-1}, M_i X_i, S_i, \eta_i; \beta) h(\eta_i|y_{it_i}, M_i X_i, S_i; \pi_{\eta S_i}) \right] d\eta_i. \quad (11)$$

These log-likelihood functions will be maximized with respect to the vector of parameters  $\theta = (\beta', \gamma')'$  that can be partitioned into the set of common parameters  $\beta$  and the set of sub-panel specific parameters  $\gamma = (\gamma'_1, \dots, \gamma'_J)'$ . The parameters that are specific to sub-panel  $j$  are, depending on the approach taken to deal with the initial observation,  $\gamma_j = (\lambda'_{S^{(j)}}, \beta'_{\eta S^{(j)}})'$  in (10), or  $\gamma_j = \pi_{\eta S^{(j)}}$  in (11).

The properties of the MLE are well-known, as well as the numerical procedures to obtain it. The problem is that the optimization procedure is cumbersome. Our specific likelihood must be optimized jointly with respect to a high number of parameters, because, due to the unbalancedness, there is a different set of some parameters for each

subpanel. This will typically preclude using standard estimation software and will increase computation time.

**Integrating over the missing observations** A different way to deal with the unbalancedness is to see it as missing observations to be integrated out. In a model without  $X$ , we would specify the initial condition for the earliest starting period in our panel and write the likelihood integrating out the unobserved periods of  $y_{it}$ , using the conditional model  $F(y_{it} | y_{it-1}, \eta_i)$ . This is implicitly assuming that all individuals follow the same process at  $t = 1$  regardless of what is the first period at which we observe them. This could be a realistic assumption when the unbalancedness is at random. However, it is harder to maintain when the unbalancedness can be correlated with individual unobserved characteristics. In a model with  $X$ , this cannot be applied unless you are willing to make further and restrictive assumptions about the stochastic process of  $X$  in a similar way as discussed at the end of section 2.

## 4.2 Minimum Distance Estimation

Here we propose an estimation method that allows us to use the same routine or estimation program as when having a balanced panel, while keeping the good asymptotic properties of the MLE. Also this procedure allows us to reduce the computational burden compared with the MLE.

The proposal has two steps. The first step is to estimate the model for each sub-panel separately. This implies that we can use the same standard software as in balanced panels and, at the same time, very easily accommodate different distributions of  $\eta_i$  for each sub-panel  $S_i$ .

The second step is to obtain estimates of the parameters  $\theta = (\beta', \gamma'_1, \gamma'_2, \dots, \gamma'_J)'$  by Minimum Distance. Let  $\hat{\delta} = (\hat{\delta}'_1, \hat{\delta}'_2, \dots, \hat{\delta}'_J)'$  be the vector of estimated coefficients of the model after the first step. Each  $\hat{\delta}'_j$  contains the estimated coefficients using sub-panel  $j$ , ( $j = 1, 2, \dots, J$ ) of the parameters of the model relevant to this sub-panel, which includes two types of parameters:

$$\hat{\delta}_j = \begin{pmatrix} \hat{\delta}_j^{[c]} \\ \hat{\delta}_j^{[nc]} \end{pmatrix}$$

where  $\hat{\delta}_j^{[c]}$  are the estimates of the parameters  $\beta$  that are common across subpanels, and  $\hat{\delta}_j^{[nc]}$  are the estimates of the non-common parameters  $\gamma_j$  for using sub-panel  $j$  only. We

also have the var-cov matrix of  $\widehat{\delta}$ ,  $Var(\widehat{\delta})$ . It is a block diagonal matrix since different sub-panels have no observations in common:

$$V = Var(\widehat{\delta}) = \begin{pmatrix} Var(\widehat{\delta}_1) & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & Var(\widehat{\delta}_J) \end{pmatrix} \quad (12)$$

where

$$Var(\widehat{\delta}_j) = \begin{bmatrix} Var(\widehat{\delta}_j^{[c]}) & Cov(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]}) \\ Cov(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]}) & Var(\widehat{\delta}_j^{[nc]}) \end{bmatrix}.$$

As described in Chamberlain (1982, 1984), we can impose restrictions on the vector of coefficients  $\delta$  by setting them to be equal to a known function of the structural parameters  $\theta$ :  $\delta = h(\theta)$ . In our case, in order to recover a unique estimate of  $\beta$ , the restriction is that we assume that all the  $\widehat{\delta}_j^{[c]}$  are estimates of the same  $\beta$  parameters that are common across sub-panels. Therefore the restrictions are

$$h(\theta) = \begin{pmatrix} h_1(\theta) \\ \vdots \\ h_J(\theta) \end{pmatrix} = P\theta$$

where  $h_j(\theta) = (\beta', \gamma_j')$  and the matrix  $P$  is

$$P = \begin{bmatrix} \iota & I & 0 & \dots & 0 & 0 \\ \iota & 0 & I & \dots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ \iota & 0 & 0 & \dots & I & 0 \\ \iota & 0 & 0 & \dots & 0 & I \end{bmatrix}$$

where  $\iota$  is a column vector of ones of the same dimension as  $\beta$ ,  $I$  is an identity matrix of the dimension of  $\gamma_j$  and  $0$  is a matrix of dimensions  $dim(\beta) \times dim(\gamma_j)$ .

Finally, the structural parameters  $\theta$  can be consistently and efficiently estimated by minimizing the following quadratic form:<sup>8</sup>

$$\widehat{\theta}^{MD} = \arg \min_{\theta} Q(\theta) = \left[ \widehat{\delta} - h(\theta) \right]' V^{-1} \left[ \widehat{\delta} - h(\theta) \right]. \quad (13)$$

It is easy to see that the solution to the minimization of this quadratic form is

$$\widehat{\theta}^{MD} = [P'P]^{-1} P' \widehat{\delta}, \quad (14)$$

---

<sup>8</sup>By choosing any other positive definite weighting matrix in this quadratic form, one still obtains consistent estimates.

where  $V$  is replaced by the consistent estimators obtained in the first step. The expression of the Variance-Covariance matrix of  $\widehat{\theta}^{MD}$  is

$$Var\left(\widehat{\theta}^{MD}\right) = \left(P'^{-1}P\right)^{-1}. \quad (15)$$

This procedure is known to be asymptotically equivalent to obtain estimates by maximizing the log likelihood  $\mathcal{L}$  on the entire set of parameters  $\theta$  (see Chamberlain, 1982 and 1984 and references in there). If  $N \rightarrow \infty$  but  $T$  and  $J$  are fixed, then the asymptotic properties derived in those references are applicable to our case. These are the relevant conditions for us since we are interested in situations in which  $N$  is large relative to  $T$  and  $J$ . Then  $\widehat{\theta}^{MD}$  is asymptotically equivalent to  $\widehat{\theta}^{MLE}$ .

Appendix A shows that the solution for each component of  $\theta$  can be expressed as

$$\widehat{\beta}^{MD} = \left[ \sum_{j=1}^J \left( Var(\widehat{\delta}_j^{[c]}) \right)^{-1} \right]^{-1} \sum_{j=1}^J \left[ \left( Var(\widehat{\delta}_j^{[c]}) \right)^{-1} \widehat{\delta}_j^{[c]} \right] \quad (16)$$

$$\widehat{\gamma}_j^{MD} = \widehat{\delta}_j^{[nc]} - Cov\left(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]}\right)' \left( Var(\widehat{\delta}_j^{[c]}) \right)^{-1} \left( \widehat{\delta}_j^{[c]} - \widehat{\beta}^{MD} \right). \quad (17)$$

Note that the estimate of the common parameters,  $\widehat{\beta}^{MD}$ , it is a weighted average only of the estimates of those parameters by sub-panels,  $\widehat{\delta}_j^{[c]}$ , whereas the MD estimate of the other parameters,  $\widehat{\gamma}_j^{MD}$ , does not only depends on  $\widehat{\delta}_j^{[nc]}$  but they are also “adjusted” by the distance between  $\widehat{\delta}_j^{[c]}$  and  $\widehat{\beta}^{MD}$ .

## 5 Implementation and practical issues

In this Section we show how to implement the estimators presented in previous sections with specific assumptions about parametric distributions. We have chosen one of the most common distributions and models used in empirical work, but the general framework and estimation procedures proposed in the previous sections are applicable to other non-linear models and parametric distributions.

### 5.1 Unbalancedness correlated with the individual effect

Let us consider the following dynamic binary choice model:

$$y_{it} = 1 \{ \alpha y_{it-1} + X_{it}'\beta + \eta_i + \varepsilon_{it} \geq 0 \} \quad (18)$$

$$-\varepsilon_{it} | y_i^{t-1}, X_i, \eta_i, S_i \underset{iid}{\sim} N(0, 1). \quad (19)$$



The probability of a given random sample of  $N$  unit observations is

$$\Pr(S'_1 Y_1, \dots, S'_N Y_N | X_1, \dots, X_N, S_1, \dots, S_N) = \prod_{i=1}^N \Pr(S'_i Y_i | M_i X_i, S_i).$$

For each  $i = 1, \dots, N$ ,

$$\begin{aligned} \Pr(s_{i1} y_{i1}, \dots, s_{iT} y_{iT} | M_i X_i, S_i, \eta_i) &= \\ &= \prod_{t=1}^T \Pr(y_{it} | s_{it-1} y_{it-1}, M_i X_i, S_i, \eta_i)^{s_{it} s_{it-1}} \Pr(y_{it} | M_i X_i, S_i, \eta_i)^{s_{it} (1-s_{it-1})} \\ &= \prod_{t=t_i+1}^{t_i+T_i-1} \Pr(y_{it} | y_{it-1}, M_i X_i, S_i, \eta_i) \Pr(y_{it_i} | M_i X_i, S_i, \eta_i), \end{aligned} \quad (20)$$

If one decides to make a distributional assumption about the conditional density of the first observation  $\Pr(y_{it_i} | M_i X_i, S_i, \eta_i)$  and about  $\eta_i$ , we can write the probability in (20) as

$$\int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i-1} \Pr(y_{it} | y_{it-1}, M_i X_i, S_i, \eta_i) \Pr(y_{it_i} | M_i X_i, S_i, \eta_i) h(\eta_i | M_i X_i, S_i) d\eta_i. \quad (21)$$

where, from the model in equations (18) and (19),  $\Pr(y_{it} | y_{it-1}, M_i X_i, S_i, \eta_i)$  in this equation is

$$\begin{aligned} \Pr(y_{it} = 1 | y_{it-1}, M_i X_i, S_i, \eta_i) &= \Pr(-\varepsilon_{it} \leq \alpha y_{it-1} + \beta_0 + X'_{it} \beta + \eta_i | y_{it-1}, M_i X_i, S_i, \eta_i) \\ &= \Pr(-\varepsilon_{it} \leq \alpha y_{it-1} + \beta_0 + X'_{it} \beta + \eta_i | y_{it-1}, M_i X_i, \eta_i) \\ &= \Phi(\alpha y_{it-1} + \beta_0 + X'_{it} \beta + \eta_i). \end{aligned} \quad (22)$$

To solve the initial conditions problem, we follow Heckman's (1981) and use as an approximation to the true model for the first observation, the same parametric form as the conditional density for the rest of the observations. Then, using normal distributions,

$$\begin{aligned} \Pr(y_{it_i} = 1 | M_i X_i, S_i, \eta_i) &= \Pr(y_{it} = 1 | X_{it}, S_i, \eta_i, s_{it-1} = 0, s_{it} = 1) \\ &= \Phi(\delta_{0S_i} + X'_{it_i} \lambda_{S_i} + \mu_{S_i} \eta_i), \end{aligned} \quad (23)$$

where we have different parameters for each value of  $S_i$ . If, instead, we allow only for correlation between  $t_i$  and  $\eta_i$ , equation (23) will be different for each  $t_i$ .

For  $h(\eta_i | M_i X_i, S_i)$  we can follow Chamberlain (1980) to allow for correlation between the individual effect and the explanatory variables:

$$\eta_i | M_i X_i, S_i \sim N\left(\beta_{0\eta_{S_i}} + \overline{M_i X_i}' \beta_{\eta_{S_i}}, \sigma_{\eta_{S_i}}^2\right), \quad (24)$$

where  $\overline{M_i X_i}$  contains the within-means of the time-varying explanatory variables for the periods that they are observed, i.e.  $\overline{M_i X_i}' = \frac{1}{T_i} \sum_{t=t_i}^{t_i+T_i-1} x_{it}$ . Notice that (24) allows for correlation between the sample selection process,  $S_i$ , and the permanent unobserved heterogeneity  $\eta_i$  because the parameters are specific to each  $S_i$ .

If one decides to consider the distribution conditional on the initial period observation, we can write the probability in (20) as

$$\left[ \int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i-1} \Pr(y_{it}|y_{it-1}, M_i X_i, S_i, \eta_i) h(\eta_i|y_{it_i}, M_i X_i, S_i) d\eta_i \right] \Pr(y_{it_i}|M_i X_i, S_i). \quad (25)$$

To solve the initial conditions problem in this case we can follow Wooldridge (2005) and specify an approximation for the density of  $\eta_i|y_{it_i}, M_i X_i, S_i$  in (25). Continuing with the Normal case, we have<sup>9</sup>

$$\eta_i|y_{it_i}, M_i X_i, S_i \sim N \left( \pi_{0S_i} + \pi_{1S_i} y_{it_i} + \overline{M_i X_i}' \pi_{2S_i}, \sigma_{\eta S_i}^2 \right). \quad (26)$$

Note that here  $\overline{M_i X_i}' = \frac{1}{T_i-1} \sum_{t=t_i+1}^{t_i+T_i} x_{it}$  for reasons given in Rabe-Hesketh and Skrondal (2013).

## Estimation

Previous models can be estimated by Maximum Likelihood (ML) and by Minimum Distance (MD).

The contribution to the likelihood function for individual  $i$  in model (21) is given by

$$L_i = \int_{\eta_i} \Phi \left( \delta_{0S_i} + X_{it_i}' \lambda_{S_i} + \mu_{S_i} \eta_i \right) (2y_{it_i} - 1) \left\{ \prod_{t=t_i+1}^{t_i+T_i-1} \Phi \left[ (\alpha y_{it-1} + \beta_0 + X_{it}' \beta + \eta_i) (2y_{it} - 1) \right] \right\} h(\eta_i|M_i X_i, S_i) d\eta_i, \quad (27)$$

where  $h(\eta_i|M_i X_i, S_i)$  is the distribution in (24) or any other distribution of  $\eta_i|M_i X_i, S_i$  like a discrete finite distribution.

In model (25) the contribution to the likelihood function for individual  $i$  is given by

$$L_i = \int \prod_{t=t_i+1}^{t_i+T_i-1} \Phi \left[ \left( \alpha y_{it-1} + X_{it}' \beta + \pi_{0S_i} + \pi_{1S_i} y_{it_i} + \overline{M_i X_i}' \pi_{2S_i} + a \right) (2y_{it} - 1) \right] \frac{1}{\sigma_{\eta S_i}} \phi \left( \frac{a}{\sigma_{\eta S_i}} \right) da. \quad (28)$$

---

<sup>9</sup>Notice that if in (26) we impose that the variance of the distribution of  $\eta_i|y_{it_i}, M_i X_i, S_i$  is constant across sub-panels, the estimation by ML becomes easier since it can be obtained using standard software for the simple random-effects probit model. See Albarran et. al. (2017) for details.

The MLE maximizes  $\mathcal{L} = \sum_{i=1}^N \log L_i$  with respect to

$$\theta \equiv \left( \alpha, \beta', \{\pi_{0j}\}_{j=1}^J, \{\pi_{1j}\}_{j=1}^J, \{\pi_{2j}\}_{j=1}^J, \{\sigma_{\eta j}\}_{j=1}^J \right).$$

For balanced panels, it is well known since Wooldridge (2005) that modeling conditional on the first observation of the dependent variable plus the normality assumption for  $\eta_i|y_{i1}, M_i X_i$  produces a simple specification that can be estimated with standard random-effects probit software. Also, for the model that follows Heckman’s approach to the initial conditions problem, Arulampalam and Stewart (2009) propose and implement a procedure using built-in commands in econometric software. However, in the unbalanced case maximizing the likelihood in (27) or (28) is cumbersome and cannot be done using such standard built-in commands.<sup>10</sup>

Regarding the MD Estimator, we obtain in a first stage  $\hat{\delta} = \left( \hat{\delta}'_1, \hat{\delta}'_2, \dots, \hat{\delta}'_J \right)'$  by maximizing  $\mathcal{L}_j = \sum_{i \in \{i: S_i = S(j)\}} \log L_i$  for each  $j = 1, \dots, J$ . Then, in the second step, we obtain  $\hat{\theta}^{MD}$  using formulas (16) and (17). The MD Estimator is simpler than MLE because we can use standard random-effects probit software and because it is much faster. Further details on computation time are given in Section 6.2. On the other hand, in specific finite samples, the MD estimator may suffer from lack of variability in individual subpanels. In such a case, the MLE could be obtained (since there could exist enough variability in the whole sample), whereas the MD estimator could fail because in the first step  $\hat{\delta}_j$  could not be computed for all the subpanels. Based on the results reported in our simulation exercises and in our application, this seems to be an infrequent problem.

### Selecting a balanced sub-sample

Making the panel balanced using the sub-set of periods at which all individuals are observed implies assuming a common normal distribution. From a practical point of view this produces a potential problem.

If there is correlation between  $\eta_i$  and  $S_i$  and the distribution of  $\eta_i|y_{i \max_{j \in [1, N]} t_j}, M_i X_i, S_i$  is Normal for each sub-panel, making the panel balanced and assuming that  $\eta_i|y_{i \max_{j \in [1, N]} t_j}, X_{it_m}^{T_m}$  follows a normal distribution –which would allow to use the simple practical solution explained in Section 5.1 of Wooldridge (2005)– is incorrect because  $\eta_i|y_{i \max_{j \in [1, N]} t_j}, X_{it_m}^{T_m}$

---

<sup>10</sup>Although in theory it is possible to obtain these ML estimates by using the ‘gllamm’ and/or ‘gsem’ commands in Stata 13 (or higher), in practice this is not computationally feasible in many cases. See the Albarraan et. al. (2017) for details.

would not follow a normal distribution.<sup>11</sup>

This also poses a problem for using the comparison between the estimates obtained taking this balanced sub-sample with the estimates obtained ignoring the unbalancedness to decide whether or not the unbalancedness is ignorable. If normality about the distribution of  $\eta_i$  is incorrectly assumed in both cases, these two estimators will tend to produce similar biased estimates. Therefore, the comparison between them may lead to the incorrect conclusion that the unbalancedness can be ignored.

## 5.2 Unbalancedness independent of the individual effect

If we assume that  $S_i$  is independent of  $\eta_i$  (condition 1), then  $h(\eta_i|M_iX_i, S_i) = h(\eta_i|M_iX_i)$ . If in addition we assume that the distribution is the same regardless of the periods for which we observe  $X_i$  (such that condition 2 in section 3.1 is satisfied) then we have:

$$\eta_i|M_iX_i, S_i \sim N\left(\beta_{0\eta} + \overline{M_iX_i}'\beta_\eta, \sigma_\eta^2\right). \quad (29)$$

If we specify the likelihood based on expression (21), under conditions 1 and 2, even though the unbalancedness is not ignorable, there is a simplification in terms of computation because the distribution of  $\eta_i$  has the same parameters for all the sub-panels.<sup>12</sup>

In contrast with that, if we use the likelihood based on (25), conditions 1 and 2 do not lead to a conditional distribution of  $\eta_i$  that is common to all sub-panels. As previously noticed, even if we assume that the sample selection process  $S_i$  is independent of  $\eta_i$ ,  $h(\eta_i|y_{it_i}, M_iX_i, S_i)$  will be different for each  $t_i$ , i.e. it will be:

$$\eta_i|y_{it_i}, M_iX_i, S_i \sim N\left(\pi_{0t_i} + \pi_{1t_i}y_{it_i} + \overline{M_iX_i}'\pi_{2t_i}, \sigma_{\eta t_i}^2\right), \quad (30)$$

unless the process is not dynamic or it is in its steady state since  $t = 0$ , or  $y_{t_i}$  comes from the same exogenous distribution for all units and  $t_i$  (conditions 3 and 4). As can be seen in (30),  $\eta_i|y_{it_i}, M_iX_i, S_i$  still has different parameters depending on when each sub-panel starts even under independence of the unbalancedness from  $\eta_i$ . This implies a

---

<sup>11</sup>Of course, balancing the panel will work if the assumption about the distribution of  $y_{i \max_{j \in [1, N]} t_j}, X_{i, mt}^{mT}$  was the correct one:  $h(\eta_i|y_{i \max_{j \in [1, N]} t_j}, X_{i, mt}^{mT}) = \sum_{j=1}^J h(\eta_i|y_{i \max_{j \in [1, N]} t_j}, X_{i, mt}^{mT}, S_i = S^{(j)}) \Pr(S_i = S^{(j)}|y_{i \max_{j \in [1, N]} t_j}, X_{i, mt}^{mT})$ . When  $h(\eta_i|y_{i \max_{j \in [1, N]} t_j}, X_{i, mt}^{mT}, S_i)$  is the normal density, is a mixture of normals with as many components as sub-panels. This would be a difficult, though not infeasible, model to estimate in practice. Certainly it is more difficult to implement than the case that assumes normality.

<sup>12</sup>For example, this makes feasible obtaining the MLE from (27) using the ‘gllamm’ and ‘gsem’ commands in Stata, as explained in Albarran et. al (2017).

more complicated structure of the Likelihood and, therefore, computation of the MLE in this case is not simpler than in the general situation without independence.

### 5.3 Average Marginal effects

The Average Marginal Effects (AMEs), which are ultimately the parameters of interest, are based on

$$E[\Phi(\alpha y_{it-1} + X'_{it}\beta + \eta_i)], \quad (31)$$

where the expectation is taken with respect to the joint distribution of  $\eta_i$  and all covariates other than the one with respect to which the marginal effect is computed.

Consider the model with a specification of the distribution of the individual heterogeneity as in (26). Using that  $\eta_i = \pi_{0S_i} + \pi_{1S_i}y_{it_i} + \overline{M_i X_i}' \pi_{2S_i} + \xi_i$  and following Wooldridge (2005), the expression (31) becomes

$$E \left[ \Phi \left( \frac{\alpha y_{it-1} + X'_{it}\beta + \pi_{0S_i} + \pi_{1S_i}y_{it_i} + \overline{M_i X_i}' \pi_{2S_i}}{\sqrt{1 + \sigma_{\eta S_i}^2}} \right) \right], \quad (32)$$

where this expectation is taken with respect to the distribution of the covariates conditional on the unbalancedness structure,  $\{S^{(1)}, \dots, S^{(J)}\}$ . It is worth noting that this expression and, therefore, the AMEs depend on the correlation between the unbalancedness and this individual effect. Therefore, when this correlation is neglected, biased estimates of the AMEs will be obtained.

The AME for a continuous regressor is the derivative of (32) with respect to that regressor, and the AME for a discrete regressor is the difference in expression (32) for a unitary change in that regressor.

The estimated AME,  $\widehat{AME}$ , can be simply obtained by replacing the population expectation in (32) with the sample mean. For instance, since the AME for the lagged dependent variable is (32) evaluated at  $y_{it-1} = 1$  minus that expression evaluated at  $y_{it-1} = 0$ , the corresponding  $\widehat{AME}$  of  $y_{it-1}$  is

$$\begin{aligned} \widehat{AME}_{y_{it-1}} &= \frac{1}{N} \sum_{i=1}^N \Phi \left( \frac{\alpha + X'_{it}\beta + \pi_{0S_i} + \pi_{1S_i}y_{it_i} + \overline{M_i X_i}' \pi_{2S_i}}{\sqrt{1 + \sigma_{\eta S_i}^2}} \right) \\ &\quad - \frac{1}{N} \sum_{i=1}^N \Phi \left( \frac{X'_{it}\beta + \pi_{0S_i} + \pi_{1S_i}y_{it_i} + \overline{M_i X_i}' \pi_{2S_i}}{\sqrt{1 + \sigma_{\eta S_i}^2}} \right). \end{aligned} \quad (33)$$

Finally, the standard errors of  $\widehat{AME}$  can be computed using the Delta method.

## 5.4 Summary of estimators and notation

We summarize the different estimators that could be used in practice when having an unbalanced panel and present the notation that will be used in next sections.

**Ignore Unbal.:** Standard ML estimators ignoring the unbalancedness and using all observations as if the panel were balanced.

**Bal. Periods:** Standard ML estimators making the panel balanced using the sub-set of periods at which all individuals are observed.

**Bal. Units:** Standard ML estimators making the panel balanced using the sub-set of individuals that are observed the same number of periods.

**Unbal. ML:** ML estimators that account for the unbalancedness ( $S$ ) and allow for correlation between  $S$  and  $\eta$ .

**Unbal. MD:** MD estimators that account for the unbalancedness ( $S$ ) and allow for correlation between  $S$  and  $\eta$ .

**Unbal.  $t_i$ \_ML:** ML estimators that account for the unbalancedness ( $S$ ) and allow for correlation between the  $S$  and  $\eta_i$  but only through the moment at which we first observe each individual,  $t_i$ . The number of periods each individual is observed is assumed to be independent of  $\eta_i$ .<sup>13</sup>

**Unbal.  $t_i$ \_MD:** The same as  $t_i$ \_ML, but estimating by Minimum Distance.

**Unbal. Indep\_ML:** ML estimator when assuming independence between  $S$  and  $\eta$ .

## 6 Simulations: Finite sample performance

In this section we use Monte Carlo techniques to illustrate the behavior of the estimators. We are particularly interested in the finite sample performance of the estimators under different degrees of unbalancedness. In the simulations we consider the model without other covariates because this model already contains all the problems we want to address

---

<sup>13</sup>For the model that specifies the distribution of  $\eta_i$  conditional on the initial observation this also corresponds with the case in which we assume that the unbalancedness is independent of  $\eta_i$ . See equation (30) and the comments that follow that equation, and comments in subsection 5.2.

and it reduces computational time. Nonetheless, we also present simulation results with exogenous covariates based on the data used in the empirical application.

## 6.1 Data Generating Process

The sub-panels may vary in both the period individuals enter and when they leave the sample. The degree of unbalancedness in the sample is governed by  $J$ , which, as defined in Section 2, indicates the number of sub-panels. The set of individuals that are observed the same periods form a sub-panel. We set  $J = 0$  when the panel is balanced. In our baseline Data Generating Process (DGP), if  $J = 2$ , i.e. if there are two sub-panels: the first half of units ( $\frac{N}{2}$ ) are observed from 1 to  $T - 1$  and the second half of units are observed from 2 to  $T$ . If  $J = 4$ , the first quarter of units are observed from 1 to  $T - 1$ , the second quarter of units are observed from 1 to  $T - 2$ , the third are observed from 2 to  $T$ , and the last quarter of units is observed from 3 to  $T$ . And the same for higher values of  $J$ . Table B.1 in Appendix B shows this structure of unbalancedness up to  $J = 6$  for a case with  $T = 6$ . Given this way of generating the unbalancedness,  $J$  can only take even values. In our simulation, we impose the following restrictions on the values of  $J$ : (i) the maximum value is  $J_{\max} = \min\{2 * T - 3, \frac{N}{30}\}$ , where  $2 * T - 3$  guarantees that all sub-panels have at least 3 periods and  $\frac{N}{30}$  guarantees that there is at least 30 units in all sub-panels, and (ii) the minimum value is  $J_{\min} = \max\{2 * T - 15, 0\}$ , where the restriction  $2 * T - 15$  is to have sub-panels with less than 8 periods.<sup>14</sup>

Given that unbalancedness structure, we generate observations for our baseline specification as follows. There are  $T$  periods in the sample and  $N$  individuals. If the panel is balanced, then we observe all individuals from 1 to  $T$ . If the panel is unbalanced, each individual  $i$  belongs to one of the  $J$  subpanels and it is observed from  $t_i$  until  $t_i + T_i - 1$ . That is,  $S_i \in j$  where  $j$  can go from 1 to  $J$ . The first  $\frac{N}{J}$  individuals that we generate belongs to sub-panel  $j = 1$ , the second  $\frac{N}{J}$  individuals (i.e.  $i$  from  $\frac{N}{J} + 1$  to  $2 * \frac{N}{J}$ ) belongs

---

<sup>14</sup>When the time length is long, fixed effects approaches may be preferable. For example, simulations in Carro (2007) show cases where a modified MLE fixed effects estimator performs well with 8 periods.

to sub-panel  $j = 2$ , and so on. Our baseline DGP is:

$$y_{it} = 1\{\alpha y_{it-1} + \eta_i + \varepsilon_{it} \geq 0\} \quad (34)$$

$$\varepsilon_{it} \underset{iid}{\sim} N(0, 1) \quad (35)$$

$$\eta_i | S_i \in j \underset{iid}{\sim} N(\mu_{\eta_j}, \sigma_{\eta_j}^2) \quad (36)$$

$$\mu_{\eta_j} = \mu_\eta + (1.3 * J / (J - 1)) * ((j/J) - (J + 1) / (2 * J)) \quad (37)$$

$$\sigma_{\eta_j} = 0.25 + (j - 1) * ((\sigma_\eta - 0.2) / (J - 1)) \quad (38)$$

$$y_{i0} = 1\{\pi_0 + ((j/J) - (J + 1) / (2 * J)) + \pi_1 \eta_i + v_{i0} \geq 0\}, \quad v_{i0} \underset{iid}{\sim} N(0, 1), \quad (39)$$

where  $\alpha = 0.75$ ,  $N = 1,000$ ,  $\mu_\eta = 0$ ,  $\sigma_\eta^2 = 1$ ,  $\pi_0 = -1.25$ , and  $\pi_1 = 0.5$ , so the initial condition and  $\eta_i$  are both correlated with the unbalancedness process. We have run 1,000 replications for each DGP considered in this section.

$\mu_{\eta_j}$  and  $\sigma_{\eta_j}^2$  are generated in a way such that in all the unbalancedness structures ( $J$ ) that we simulate:  $E_j(\mu_{\eta_j}) = \mu_\eta = 0$ ,  $E_j(\sigma_{\eta_j}) = 0.6$ ,  $\mu_{\eta_j} \in [-1, 1]$ ,  $\sigma_{\eta_j} \in [0.2, 1]$ , and  $\mu_{\eta_j}$  and  $\sigma_{\eta_j}$  are increasing in  $S$ . Thus, the value of  $\eta_i$  is more likely to be larger the larger the value of  $j$ , i.e. for the last sub-panels.

After the baseline DGP is simulated, it is changed to evaluate the finite sample performance along the following dimensions:

1. Unbalancedness only from the left, i.e., sub-panels differ only on the period they start but all are observed until  $T$ . In this case  $J$  can take both even and odd values. Table B.1 contains examples of this unbalanced structure. Apart from the balanced case ( $J = 0$ ),  $J$  goes from  $J_{\min} = \max\{T - 6, 4\}$  to  $J_{\max} = \min\{T - 2, \frac{N}{30}\}$ .  $J_{\min}$  cannot be smaller than 4 because since the unbalancedness is only from the left, a smaller  $J$  would be a case too close to a balanced situation and we have not considered it. The way in which we generate  $\eta_i$  here implies that individuals with higher  $\eta_i$  tend not to be observed the first periods.
2. Unbalancedness process and initial condition both generated exogenously:  $\eta_i | S_i \in j \underset{iid}{\sim} N(\mu_\eta, \sigma_\eta^2)$ ,  $\mu_\eta = 0$ ,  $\sigma_\eta^2 = 1$ , so that  $\eta_i$  does not depend on  $j$ , and  $y_{i0} = 1\{\pi_0 + v_{i0} \geq 0\}$ . We also consider exogenous unbalancedness process but maintaining the endogeneity of the initial condition.
3. We have also considered  $N = 500$ .



4. We have also considered  $\alpha = 0.5$  and  $\alpha = 1$  to evaluate the sensitivity to different degrees of persistence.

## 6.2 Monte Carlo results

For the sake of brevity not all estimators are used in all the simulation exercises. Our general criteria has been to study in each simulated DGP the performance of estimators whose assumptions correspond with those made in the DGP. For instance, even though the “Unbal. ML” estimator will give consistent estimates in all the cases considered in this paper, when the unbalancedness is generated at random, only the estimators based this assumption (or a weaker version of it) are used. Nonetheless, for completeness, there will be a few simulations shown in Appendix B in which other estimators, including those that are known to be incorrect, are used too.

Table 1 and 2 shows the results for our baseline specification, in which both the initial condition and the unbalancedness are correlated with  $\eta_i$ , when estimating the  $\alpha$  parameter and the AME of the lagged dependent variable, respectively. Since the true AME (slightly) varies with the sample drawn in each Monte Carlo simulation, Tables 2 reports the true expected AME along with the estimated AME and the Root Mean Square Error (RMSE) of the estimators. In all cases we deal with the initial conditions problem by specifying the density of the unobserved effect conditional on the first observation.<sup>15</sup>

In Table 1 we observe that all the four approaches considered provide estimated values of the parameter  $\alpha$  very close to its true value. However, there exists some relevant points that are worth noting. The solution that employs standard methods after balancing the sample using the subset of periods at which all individuals are observed (“Bal. Periods”), has two important drawbacks compared to the approaches that account for the unbalancedness. First, this solution cannot be employed in many cases, including some where the unbalancedness is moderate: for  $J = 4$  with  $T = 6$  or  $J = 6$  with  $T = 8$ . Second, this solution implies an important loss of efficiency in terms of RMSE of the parameter and of the marginal effect when it can be employed, even for moderate unbalancedness. For instance, Table 1 shows that for  $T = 8$  and  $J = 4$  the RMSE of the estimated parameter

---

<sup>15</sup>Table B.6 in Appendix B presents simulations dealing with the initial conditions problem by specifying the density of the first observation conditional on  $\eta_i$  and the density of  $\eta_i$ , and by specifying the density on  $\eta_i$  conditional on the first observation. We find that the results are very similar, so we focus on the latter in all the other simulations.

Table 1: Simulation results on the estimation of  $\alpha$ . Baseline case. Double Unbalancedness.

	Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML	Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML
	$\hat{\alpha}$				RMSE			
T=4								
J=0	0.7502	0.7491	0.7502	0.7491	0.0860	0.0872	0.0860	0.0872
J=2		0.7568	0.7361	0.7185		0.1855	0.1241	0.1283
T=6								
J=0	0.7508	0.7476	0.7508	0.7508	0.0579	0.0589	0.0579	0.0579
J=2	0.7211	0.7544	0.7492	0.7468	0.0925	0.1016	0.0685	0.0675
J=4		0.7448	0.7485	0.7432		0.1504	0.0714	0.0707
J=6		0.7500	0.7505	0.7356		0.1803	0.0811	0.0810
T=8								
J=2	0.7282	0.7519	0.7510	0.7504	0.0610	0.0830	0.0514	0.0511
J=4	0.7457	0.7525	0.7512	0.7499	0.0925	0.1181	0.0506	0.0504
J=6		0.7476	0.7468	0.7440		0.1387	0.0557	0.0555
J=8		0.7415	0.7490	0.7432		0.1615	0.0598	0.0590
J=10		0.7500	0.7512	0.7407		0.1887	0.0638	0.0638
T=10								
J=6	0.7502	0.7551	0.7515	0.7507	0.1004	0.1186	0.0430	0.0433
J=8		0.7500	0.7485	0.7472		0.1441	0.0449	0.0448
J=10		0.7413	0.7533	0.7505		0.1580	0.0494	0.0489
J=12		0.7536	0.7482	0.7429		0.1661	0.0502	0.0501
J=14		0.7464	0.7516	0.7430		0.1784	0.0549	0.0544
T=15								
J=16		0.7517	0.7492	0.7484		0.1532	0.0353	0.0352

Note: In the baseline case, the initial condition and the unbalancedness are both correlated with  $\eta$ ,  $\alpha = 0.75$  and  $N = 1,000$ .

of the “Bal. Periods” estimator is around 0.09 compared with around 0.05 for the “Unbal. MD” and “Unbal. MLE”. For the AME, see Table 2, these figures are around 0.04 and 0.02, respectively. Regarding the “Bal. Units” estimator, we find that the RMSE when estimating  $\alpha$  is much higher than in the case of any other estimator due to the loss of observations when using this estimator. Furthermore in the estimation of the marginal effect the “Bal. Unit” has not only an efficiency loss but also a bias problem. As explained in Section 3.3, the estimated marginal effect is inconsistent, and this is precisely what we can see in Table 2. As a result the RMSE is twice to five times larger than the “Unbal. MD” and the “Unbal. MLE”.

With respect to the comparison between “Unbal. MD” and “Unbal. MLE”, as expected, the behavior of these estimators is very similar, both in terms of the estimated

Table 2: Simulation results on the estimation of the AMEs. Baseline case. Double Unbalancedness.

		Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML	Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML
	AME	$\widehat{AME}$				RMSE			
T=4									
J=0	0.2020	0.2096	0.2088	0.2096	0.2015	0.0298	0.0298	0.0298	0.0286
J=2	0.2194		0.1813	0.2201	0.2072		0.0621	0.0460	0.0441
T=6									
J=0	0.2021	0.2096	0.2089	0.2096	0.2024	0.0206	0.0206	0.0206	0.0187
J=2	0.2194	0.2152	0.1813	0.2259	0.2183	0.0315	0.0479	0.0291	0.0228
J=4	0.2278		0.1825	0.2320	0.2249		0.0624	0.0281	0.0252
J=6	0.2292		0.1847	0.2320	0.2230		0.0677	0.0312	0.0289
T=8									
J=2	0.2195	0.2183	0.1811	0.2279	0.2192	0.0197	0.0451	0.0206	0.0169
J=4	0.2280	0.2414	0.1836	0.2349	0.2279	0.0381	0.0559	0.0205	0.0176
J=6	0.2297		0.1833	0.2336	0.2272		0.0612	0.0215	0.0195
J=8	0.2300		0.1820	0.2333	0.2269		0.0673	0.0228	0.0205
J=10	0.2300		0.1848	0.2329	0.2253		0.0707	0.0240	0.0224
T=10									
J=6	0.2300	0.2484	0.1857	0.2368	0.2300		0.0558	0.0174	0.0149
J=8	0.2305		0.1845	0.2354	0.2290		0.0625	0.0171	0.0155
J=10	0.2304		0.1823	0.2362	0.2303		0.0671	0.0190	0.0171
J=12	0.2306		0.1858	0.2339	0.2276		0.0664	0.0185	0.0172
J=14	0.2303		0.1852	0.2342	0.2270		0.0694	0.0209	0.0190
T=15									
J=16	0.2312		0.1849	0.2372	0.2304		0.0651	0.0137	0.0118

Note: See note in Table 1.

parameters and the marginal effects. Although when estimating the model by ML we make use an efficient use of all the observations in the sample, estimating this model is computationally cumbersome and takes a lot of time because all parameters are jointly estimated. The MLE can take between 150 and 1,600 times more computing time than the MD, depending on the number of periods and subpanels. Note that the computation time will further increase when adding covariates. In contrast to that, the MD estimation is much faster. On the other hand, we face a potential problem of lack of variability in certain sub-panels, although the percentage of simulations that achieved convergence for the MD estimator is very high. In Table B.11, we see that the percentage of failures is below 1%. Higher failure rates only appear in a few cases when considering very high

degree of state dependence.

We have also simulated the baseline DGP with left-side unbalancedness. Results are reported in Table B.2 and B.3 in Appendix B. The comparison of the different estimators leads to the same conclusions as in the double unbalancedness case, both for the parameter  $\alpha$  and for the AMEs.

Finally, Table 3 presents the simulation results for a situation in which the initial condition and the unbalancedness are uncorrelated with  $\eta_i$ . As in the baseline case we present the results from the two solutions that balance the sample. The results show a similar pattern than in the previous DGPs, but the performance of the “Bal. Periods” estimator is slightly better. This is because in this case it is easier to approximate the relation between  $\eta$  and  $S$  with a common distribution of  $\eta$  that does not change across sub-panels. Third column in Table 3 presents MD estimates allowing for correlation between the unbalancedness and  $\eta_i$  only through the moment at which we first observe each individual,  $t_i$ . Last column presents the ML estimates using the assumption of independence between the unbalancedness and  $\eta$ , which corresponds with the assumption in this DGP, but in this case using the distribution of the first observation conditional on the unobserved effect and the distribution of the unobserved effect to deal with the initial conditions problem. As in the baseline case, we can see that both MD and ML estimates behave very similarly, being the former computationally much faster. Given this, the remaining simulation results that can be found in Appendix B do not include ML estimates.

From Tables B.4 to B.10 in Appendix B, we present a number of simulation results where we have sequentially changed different parameters of the baseline specification as explained in Section 6.1. Although the RMSE of all the solutions is increased when the sample size decreases, the results remain basically unchanged.

## 7 An application to export market participation

We illustrate previous methods by estimating a model for firms’ export market participation decision. We use data for Spanish manufacturing firms, the Business Strategies Survey (*Encuesta sobre Estrategias Empresariales*, ESEE). The survey, sponsored by the Spanish Ministry of Industry and published by the Fundación Empresa Pública, is under-

Table 3: Simulation results on the estimation of  $\alpha$ . Baseline case but the initial condition and the unbalancedness are uncorrelated with  $\eta$ . Double Unbalancedness.

	Bal. Periods	Bal. Units	Unbal. $t_i\_MD$	Unbal. Indep\_ML	Bal. Periods	Bal. Units	Unbal. $t_i\_MD$	Unbal. Indep\_ML
$\alpha = 0.75$	$\hat{\alpha}$				RMSE			
T=4								
J=0	0.7498	0.7498	0.7498	0.7498	0.0844	0.0844	0.0844	0.0844
J=2		0.7564	0.7528	0.7416		0.2592	0.1433	0.1261
T=6								
J=0	0.7501	0.7501	0.7501	0.7501	0.0566	0.0566	0.0566	0.0566
J=2	0.7549	0.7531	0.7502	0.7488	0.1085	0.1152	0.0728	0.0709
J=4		0.7540	0.7505	0.7488		0.1685	0.0825	0.0783
J=6		0.7556	0.7585	0.7484		0.2054	0.0939	0.0861
T=8								
J=2	0.7517	0.7488	0.7501	0.7495	0.0673	0.0827	0.0548	0.0545
J=4	0.7583	0.7465	0.7508	0.7498	0.1166	0.1173	0.0584	0.0578
J=6		0.7450	0.7512	0.7502		0.1407	0.0628	0.0620
J=8		0.7454	0.7513	0.7473		0.1588	0.0699	0.0679
J=10		0.7448	0.7542	0.7457		0.1781	0.0755	0.0729
T=10								
J=6	0.7565	0.7526	0.7520	0.7513	0.1155	0.1199	0.0514	0.0509
J=8		0.7537	0.7511	0.7501		0.1365	0.0541	0.0536
J=10		0.7533	0.7519	0.7508		0.1505	0.0573	0.0572
J=12		0.7558	0.7535	0.7496		0.1620	0.0607	0.0594
J=14		0.7534	0.7540	0.7525		0.1751	0.0653	0.0645
T=15								
J=16		0.7541	0.7509	0.7499		0.1416	0.0404	0.0406

taken annually since 1990 and constitutes an unbalanced panel.<sup>16</sup>

We use data for the period 1990 to 1999, because after 1999 there was a change in the sample design.<sup>17</sup> We have dropped those observations for which relevant information is missing and those firms affected, in the corresponding year, by some process of absorption,

<sup>16</sup>There are different reasons why firms disappear from the sample: liquidations, merges, acquisitions, or activity changes to non-manufacturing sectors. There are also firms that stop collaborating for unknown reasons. Refreshment samples try to compensate the drop outs with new entrants, so as to keep approximately the initial sample size. However, the entrance of new firms is not implemented every year, nor in a systematic way. For more details on the survey, see, for example, Fariñas and Jaumandreu (1999).

<sup>17</sup>Data before and after the change are hardly comparable, because big firms are overrepresented for several waves after 1999. Including also this period after 1999 would imply a more severe unbalanced structure, which would go in favor of our proposed estimators. However, we preferred not to use this data because we did not want the unbalancedness structure to be artificially driven by a change in the sample design.

merger or split. We also eliminate the firms that are not observed in consecutive years and those with less than 3 observations in the sample period. Our final sample consists of an unbalanced panel of 1,920 firms and 13,203 observations. There are between 1,200 and 1,450 firms per year approximately and 36% of the firms are in the panel during the whole sample period. Table 4 shows the unbalancedness structure of our sample. We have 34 different subpanels, some of them with missing observations to the right and others with missing observations to the left. We have dropped out those subpanels with less than 100 observations (less than 30 firms). This only represents a loss of less than 5% of the sample, while making it feasible to obtain the MD estimator. We end up with 14 different subpanels and with 12,683 observations and 1,807 firms.

Table 4: Unbalancedness structure of the total sample

Subpanel	Number of firms	Pattern by year									
		1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
$S = 1$	143	x	x	x	.	.	.	.	.	.	.
$S = 2$	100	x	x	x	x	.	.	.	.	.	.
$S = 3$	102	x	x	x	x	x	.	.	.	.	.
$S = 4$	66	x	x	x	x	x	x	.	.	.	.
$S = 5$	63	x	x	x	x	x	x	x	.	.	.
$S = 6$	48	x	x	x	x	x	x	x	x	.	.
$S = 7$	79	x	x	x	x	x	x	x	x	x	.
$S = 8$	699	x	x	x	x	x	x	x	x	x	x
$S = 9$	65	.	x	x	x	x	x	x	x	x	x
$S = 10$	34	.	.	x	x	x	x	x	x	x	x
$S = 11$	37	.	.	.	x	x	x	x	x	x	x
$S = 12$	34	.	.	.	.	x	x	x	x	x	x
$S = 13$	91	.	.	.	.	.	.	x	x	x	x
$S = 14$	246	.	.	.	.	.	.	.	x	x	x
$S = 1$ to 14	1,807										
$S = 15$	16	.	.	.	.	.	.	x	x	x	.
$S = 16$	12	.	x	x	x	x	.	.	.	.	.
...		other patterns									
$S = 15$ to 34	113										
All subpanels	1,920										

The starting point for estimation is an equation of the form

$$y_{it} = \mathbf{1}(\alpha y_{it-1} + X'_{it}\beta + \eta_i + v_{it} \geq 0), \quad (40)$$

where  $y_{it} = 1$  if the  $i - th$  firm exported in year  $t$ . Our empirical model is based on a

simple model of optimization for a firm facing the export decision. A profit-maximizing firm makes its decision based on the expected profits from exporting, now and in the future, taking into account the sunk costs of entering the new market, which depends on previous export behavior, and other variable costs (see Roberts and Tybout, 1997). On the other hand, export decision is also affected by a number of time-invariant firm-specific characteristics (for instance, product quality or managerial ability) that cannot be directly observed. The choice of variables included in the vector  $X$  largely follows the previous literature on the determinants of firm’s export decisions. See Table C.1 in Appendix C for the definition and descriptive statistics of these variables.<sup>18</sup> Exporters in our total sample are on average larger and older companies, spend more on R&D and have a higher proportion of skilled workers. These characteristics are in line with the previous literature. With respect to the persistence, 46% of firms in the sample exported each year, while approximately 36% of firms never exported and about 18% showed variation in their exporting behavior.

With this data set we cannot perform the estimates using the balanced sample formed by the periods at which all the firms are observed (“Bal. Units”), because we do not have any observation in that set (see Table 4). We have performed estimates using the balanced sample with the firms that are observed all the periods (those in subpanel  $S = 8$ ). Tables 5, 6 and C.2 in Appendix B presents the estimation results. Columns labeled “Bal. Units” shows estimates using the previously commented balanced sample. Columns labeled “Ignore Unbal.” present the results from a model that ignores the unbalancedness. And, finally, columns labeled “Unbal. MD” show the results from the model that accounts for the unbalancedness and allows for its correlation with the unobserved effect. We consider only the MD estimates because the estimation by ML is computationally cumbersome and it presents similar performance than the MD ones, as shown in the simulations. In all cases we model the unobserved heterogeneity conditional on the initial condition and the time average of the exogenous variables as in equation (26).

Table C.2 presents three estimates of the common parameters of the correlated random effects probit model. Given that the comparison between estimated parameters is not obvious because of the differences in scale, we focus on Average Marginal Effects (AME).

---

<sup>18</sup>Typically the literature includes a measure of firm’s productivity. Our data set does not offer a good measure for it, but it has been accounted for by including firms’ specific effects.

Moreover, marginal effects are usually the parameters of interest in nonlinear models. These are presented in Table 5 for the lagged export status variable. First row presents the AME for the entire sample. As explained in the previous sections, the estimator that ignores the unbalancedness and the estimator that takes the balanced sample are incorrect, since they do not allow the unobservables to have different distributions across subpanels. Coincidentally both estimators provide similar results, probably because the observations in the balanced sample are 56% of the sample used to estimate the model ignoring unbalancedness. This is an example where comparing these two estimators leads to the incorrect conclusion about the possibility of ignoring the unbalancedness.

Regarding the “Unbal. MD” estimator, we find that the estimated AME for the entire sample is around 4 percentage points greater than the one from the estimator that ignores the unbalancedness. This difference is statistically significant even though the AME for the entire sample tends to mask biases in opposite directions in different subsamples.<sup>19</sup> This can be seen in more detail if we analyze the average marginal effects by subgroups (see Table 5) and by subpanels (see Table 6). In particular, we find statistically different results for younger firms and also for firms that do not export in the first period. The last row in Table 5 presents the estimates excluding the largest subpanel ( $S = 8$ ). This is to show that if we had a dataset without a panel that dominates so much, the differences between the “Ignore Unbal.” and “Unbal. MD.” are more significant, as it can be seen from the Test of Difference in the last column.

If we look at the AMEs by subpanel, there are five subpanels in which the MD gives statistically significant AMEs and the differences are even larger than for the total sample. Furthermore, while the MD estimates range between 0.1095 and 0.4399, the corresponding estimates for the model ignoring the unbalancedness only range from 0.2108 to 0.2689. There is a great deal of variation on the marginal effect of lagged export across subpanels that is not captured by the “Ignore Unbal.” estimator. These results indicate that the model that ignores the unbalancedness incorrectly imposes, among other restrictions, independence between the distribution of the unobserved heterogeneity and the unbalancedness.

---

<sup>19</sup> In the last column of Tables 5 and 6 we have performed a Hausman-type test of the difference between the “Unbal. MD” estimates and the estimates that ignore the unbalancedness, using the variance-covariance matrix of the MD estimates only instead of subtracting from it the variance of the “Ignore Unbal.” estimator. 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 variance-covariance matrix of the difference.



Table 5: Estimated Average marginal effects of Lagged Export.

	Bal. Units (1)	Ignore Unbal. (2)	Unbal. MD (3)	Test of Diff. (2) vs (3)
Total sample	0.2423 (0.0290)	0.2351 (0.0234)	0.2776 (0.0254)	*
Subsample, by age <sup>††</sup>				
Age < 12	0.2590 (0.0313)	0.2528 (0.0251)	0.3181 (0.0290)	**
Age 12-24	0.2735 (0.0314)	0.2573 (0.0250)	0.2994 (0.0266)	
Age > 24	0.2121 (0.0268)	0.2032 (0.0212)	0.2307 (0.0234)	
Subsample, by I.C.				
Export <sub>t<sub>i</sub></sub> = 1	0.1640 (0.0257)	0.1808 (0.0209)	0.2064 (0.0234)	
Export <sub>t<sub>i</sub></sub> = 0	0.2811 (0.0269)	0.2811 (0.0269)	0.3391 (0.0287)	**
Subpanels $S \neq 8$		0.2358 (0.0236)	0.3267 (0.0328)	***

Note: Standard errors are reported in parentheses. The implementation of the test of difference is discussed in footnote 19. Asterisks indicate the difference is significantly different from zero at \*10%; \*\*5%; \*\*\*1%.

†† The age of approximately one third of the sample is lower than 12 and around 40% of the firms are 24 or older.

**Simulation evidence on the properties of the estimators** We simulated data calibrated to the ESEE sample to study the properties of the estimators in the empirical application. This also has the additional interest of exhibiting some Monte Carlo results with covariates in the dynamic model, which complement those reported in Section 6.

The data generating process in sub-section 6.1 is extended here to incorporate exogenous covariates. Thus, the main equation becomes

$$y_{it} = 1\{\alpha y_{it-1} + X'_{it}\beta + \eta_i + \varepsilon_{it} \geq 0\}, t = t_i + 1, \dots, t_i + T_i \quad (41)$$

$$\varepsilon_{it}|y_{it_i}, X_i, S_i \in j \underset{iid}{\sim} N(0, 1) \quad (42)$$

$$\eta_i|y_{it_i}, X_i, S_i \in j \sim N\left(\pi_{0j} + \pi_{1j}y_{it_i} + \overline{\overline{X}}_i \pi_{2j}, \sigma_{\eta_j}^2\right), \quad (43)$$

where  $X_{it}$  denotes the vector of exogenous regressors,  $\overline{\overline{X}}_i$  contains the within-means (from period  $t_i + 1$  to  $t_i + T_i$ ) of the time-varying explanatory regressors and  $y_{it_i}$  is the first observed value of the endogenous variable for the individual  $i$  (i.e., her initial condition).

Table 6: Estimated Average marginal effects of Lagged Export. By Subpanels

	Bal. Units (1)	Ignore Unbal. (2)	Unbal. MD (3)	Test of Diff. (2) vs (3)
Subpanels				
$S = 1$		0.2414 (0.0245)	0.2903 (0.0904)	
$S = 2$		0.2338 (0.0239)	0.4380 (0.0442)	***
$S = 3$		0.2470 (0.0247)	0.4144 (0.0776)	**
$S = 4$		0.2108 (0.0218)	0.2539 (0.1033)	
$S = 5$		0.2340 (0.0239)	0.3477 (0.0732)	
$S = 6$		0.2230 (0.0222)	0.1095 (0.0209)	***
$S = 7$		0.2182 (0.0223)	0.3441 (0.0477)	***
$S = 8$	0.2423 (0.0290)	0.2336 (0.0233)	0.2413 (0.0245)	
$S = 9$		0.2195 (0.0221)	0.2758 (0.0793)	
$S = 10$		0.2612 (0.0257)	0.2634 (0.1403)	
$S = 11$		0.2689 (0.0260)	0.3256 (0.0830)	
$S = 12$		0.2674 (0.0251)	0.3144 (0.1175)	
$S = 13$		0.2563 (0.0250)	0.4399 (0.0393)	***
$S = 14$		0.2374 (0.0239)	0.3765 (0.0877)	

Note: See note in Table 5.

In this Monte Carlo exercise, the initial conditions, the exogenous regressors  $X_{it}$  and their means  $\overline{X}_i$  are kept fixed at their observed sample values across simulations. In order to make the simulated model as parsimonious as possible, we have excluded those variables that were insignificant in the main estimated equation (41) and in the auxiliary regressions (43) for all the subpanels  $j = 1, \dots, 14$ . Therefore, we consider a model only with “Size”, the share of “Medium skill” workers, the “Age” of the firm and a time trend as covariates  $X_{it}$  (and  $\overline{X}_i$ , accordingly), and it is estimated using the “Unbal. MD” in the entire sample. The parameters required for simulations,  $\alpha$ ,  $\beta$ ,  $\pi_{0j}$ ,  $\pi_{1j}$ ,  $\pi_{2j}$  and  $\sigma_{\eta j}^2$  are

taken from these estimates. In each of our 1,000 simulations, we obtained random draws of two mutually independent standard normal variables,  $\varepsilon_{it}$  and  $\xi_i$ . Thus the simulated values of  $y_{it}$  are computed by plugging  $\varepsilon_{it}$  and  $\eta_i = \pi_{0j} + \pi_{1j}y_{it_i} + \overline{X}_i \pi_{2j} + \sigma_{\eta j} * \xi_i$  into (41).

Table 7 contains the simulation results for the lagged dependent variable coefficients and average marginal effects. The results obtained in this experiment confirms that the “Unbal. MD” outperforms both “Bal. Units” and “Ignore Unbal.” estimator. First, in part A of Table 7, we can see that our proposed “Unbal. MD” estimator performs clearly better in terms of both bias and RMSE than the two alternatives to estimate in the total sample the state-dependence parameter  $\alpha$  and, specially, the AME. Then, in part B of Table 7 we check the extent to which each estimator is able to capture heterogeneity in the AME across subgroups. Of course, the “Bal. Units” estimator does a nice work in the only subpanel that is using, but neglects the other ones. On the other hand the estimator that ignores the unbalancedness can provide different AMEs across subgroups, but the estimated AMEs are substantially biased in some of them. By contrast, the “Unbal. MD” estimator performs reasonably well overall.

## 8 Conclusions

In this paper we consider the estimation of dynamic non-linear correlated random effects models when using unbalanced panel data. We identify two types of problems: (i) an inconsistency in the estimates of the coefficients when the unbalancedness is ignored; and (ii) an efficiency loss and/or an inconsistency in the estimates when using different balanced versions of the unbalanced original data. These problems are specially severe when the unbalanced process is correlated with the individual effect.

We propose a general model that accounts for the unbalancedness that can be arbitrarily correlated with the permanent unobserved heterogeneity. We show that this model can be estimated by Maximum Likelihood (ML) and also by Minimum Distance (MD). Monte Carlo experiments and an empirical illustration show that our proposed estimation approaches perform better both in terms of bias and RMSE than the approaches that ignore the unbalancedness or that balance the sample. Both the ML and the MD estimators have comparative advantages and disadvantages. Its computational simplicity leads us to favor the MD approach.

Table 7: Simulation results based on the MD results obtained in the empirical application

A. Total Sample							
True parameter		Bal. Units		Ignore Unbal.		Unbal. MD	
		Estimated	RMSE	Estimated	RMSE	Estimated	RMSE
$\alpha = 1.5153$		1.5145	0.0962	1.4651	0.0900	1.5413	0.0796
$AME = 0.2721$		0.2332	0.0483	0.2449	0.0358	0.2539	0.0305

  

B. By Subgroups							
True AME		Bal. Units		Ignore Unbal.		Unbal. MD	
		$\widehat{AME}$	RMSE	$\widehat{AME}$	RMSE	$\widehat{AME}$	RMSE
By Age							
< 12	0.3055	0.2495	0.0639	0.2606	0.0513	0.2698	0.0463
12 – 24	0.2895	0.2606	0.0425	0.2647	0.0352	0.2761	0.0292
> 24	0.2310	0.2065	0.0363	0.2172	0.0258	0.2249	0.0245
By Initial Conditions							
$y_{t_i} = 1$	0.2167	0.1644	0.0580	0.1974	0.0287	0.2041	0.0281
$y_{t_i} = 0$	0.3200	0.2962	0.0412	0.2857	0.0430	0.2970	0.0364
By Subpanels							
$S = 8$	0.2334	0.2332	0.0280	0.2442	0.0255	0.2387	0.0266
$S \neq 8$	0.3196			0.2456	0.0778	0.2726	0.0576

The comparison between the sets of estimates presented in the empirical application emphasizes the point that different individuals behave differently due to the heterogeneity in the distribution of the unobservables across subpanels. It also reveals the importance of accounting for it to give a proper estimate of the marginal effect of the explanatory variables in a dynamic non-linear model.

The model and estimation solutions proposed in this paper include different possible random effects specifications: from the simple model in which the random heterogeneous intercepts are independent to the most general correlated random coefficients model. Moreover, it can easily accommodate models for which the state dependence follows higher order Markov chains and models with autocorrelation in the error term.

## References

- [1] Akee, R. K, W. E. Copeland, G. Keeler, A. Angold, and E.J. Costello (2010): “Parents’ Incomes and Children’s Outcomes: A Quasi-Experiment Using Transfer Pay-

- ments from Casino Profits”, *American Economic Journal: Applied Economics* 2:1, 86–115.
- [2] Albarran, P., R. Carrasco and J. Carro (2017): “Using Stata to Estimate Dynamic Nonlinear Random Effects Models with Unbalanced Panels”, *mimeo*.
- [3] Arellano, M. (2003): “Discrete Choice with Panel Data”, *Investigaciones Económicas*, vol. XXVII (3), 423-458.
- [4] Arellano, M. and J. Hahn (2007): “Understanding Bias in Nonlinear Panel Models: Some Recent Developments”, in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Volume 3*, edited by Richard Blundell, Whitney Newey, and Torsten Persson. Cambridge University Press.
- [5] Arellano, M. and B. Honoré (2001): “Panel Data Models: Some Recent Developments”, in J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, Vol. 5. North-Holland, 3229-3296.
- [6] Arulampalam, W. and M. B. Stewart (2009): “Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators”, *Oxford Bulletin of Economics and Statistics*, 71, 659-681.
- [7] Bhattacharya, D. (2008): “Inference in panel data models under attrition caused by unobservables”, *Journal of Econometrics*, 144, 430-446.
- [8] Bernard, A. B. and J. B. Jensen (2004): “Why Some Firms Export”, *The Review of Economics and Statistics*, 86(2), 561-569.
- [9] Campa, J. M. (2004): “Exchange rates and trade: How important is hysteresis in trade?”, *European Economic Review*, 48, 527-548.
- [10] Carro, J. M. (2007): “Estimating dynamic panel data discrete choice models with fixed effects”, *Journal of Econometrics*, 140, 503-528.
- [11] Chamberlain, G., (1980): “Analysis of Covariance with Qualitative Data”, *The Review of Economic Studies*, 47 (1), 225-238.
- [12] Chamberlain, G., (1982): “Multivariate Regression Models for Panel Data”, *Journal of Econometrics*, 18, 5-46.

- [13] Chamberlain, G., (1984): “Panel data”, in Griliches, Z., Intriligator, M.D., (Eds.), *Handbook of Econometrics*, Vol. 2. North-Holland, Amsterdam.
- [14] Contoyannis, P., A. M. Jones and N. Rice (2004): “The Dynamics of Health in the British Household Panel Survey”, *Journal of Applied Econometrics*, 19, 473-503.
- [15] Fariñas, J. C. and J. Jaumandreu (1999): “Diez años de Encuesta sobre Estrategias Empresariales”, *Economía Industrial*, 329, 29-42.
- [16] 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*, Manski, C., McFadden, D. (eds). MIT Press: Cambridge, MA, 114–178.
- [17] Hyslop, D. R. (1999): “State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women”, *Econometrica*, 67, 1255-1294.
- [18] Magnus, J. R. and H. Neudecker (1999): “Matrix Differential Calculus with Applications in Statistics and Econometrics”, Wiley, England.
- [19] Pacini, D. and F. Windmeijer (2015): “Moment conditions for AR(1) panel data models with missing outcomes”, Discussion Paper 15/660, Department of Economics, University of Bristol.
- [20] Rabe-Hesketh, S., and A. Skrondal (2013): “Avoiding biased versions of Wooldridge’s simple solution to the initial conditions problem”, *Economics Letters*, 120, 346-349.
- [21] Roberts, M., and J. Tybout, J. (1997): “The decision to export in Colombia: An empirical model of entry with Sunk Costs”, *American Economic Review*, 87(4), 545-564.
- [22] Stewart, M. B. (2007): “The interrelated dynamics of unemployment and low-wage employment”, *Journal of Applied Econometrics*, 22, 511- 531.
- [23] Wooldridge, J. M. (2005): “Simple Solutions to the Initial Conditions Problem for Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity”, *Journal of Applied Econometrics*, 20, 39-54.

- [24] Wooldridge, J. M. (2007): “Inverse probability weighted estimation for general missing data problems”, *Journal of Econometrics* 141, pp. 1281-1301
- [25] Wooldridge, J. M. (2010): “Correlated Random Effects Models with Unbalanced Panels”, mimeo.

## A Appendix: MD Estimator

We show that (16) and (17) are the result to the minimization problem in (13). Given that the matrix  $V$  is block diagonal, that problem becomes

$$\min_{\theta} Q(\theta) = \sum_{j=1}^J \left[ \begin{pmatrix} \widehat{\delta}_j^{[c]} - \boldsymbol{\beta} \\ \widehat{\delta}_j^{[nc]} - \gamma_j \end{pmatrix}' \left( \text{Var}(\widehat{\delta}_j) \right)^{-1} \begin{pmatrix} \widehat{\delta}_j^{[c]} - \boldsymbol{\beta} \\ \widehat{\delta}_j^{[nc]} - \gamma_j \end{pmatrix} \right], \quad (44)$$

where  $\text{Var}(\widehat{\delta}_j)$ , that is the  $j$ th-block in the diagonal of  $\text{Var}(\widehat{\delta})$ , is:

$$\text{Var}(\widehat{\delta}_j) = \begin{pmatrix} \text{Var}(\widehat{\delta}_j^{[c]}) & \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]}) \\ \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]})' & \text{Var}(\widehat{\delta}_j^{[nc]}) \end{pmatrix}. \quad (45)$$

Based on the result on the inverse of block matrices (cf. p. 11 in Magnus and Neudecker, 1999):

$$\left( \text{Var}(\widehat{\delta}_j) \right)^{-1} \equiv \begin{pmatrix} V_{11j} & V_{12j} \\ V_{12j}' & D_j^{-1} \end{pmatrix}, \quad (46)$$

where

$$V_{11j} = \text{Var}(\widehat{\delta}_j^{[c]})^{-1} + \text{Var}(\widehat{\delta}_j^{[c]})^{-1} \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]}) D_j^{-1} \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]})' \text{Var}(\widehat{\delta}_j^{[c]})^{-1} \quad (47)$$

$$V_{12j} = -\text{Var}(\widehat{\delta}_j^{[c]})^{-1} \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]}) D_j^{-1} \quad (48)$$

$$D_j = \text{Var}(\widehat{\delta}_j^{[nc]}) - \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]})' \text{Var}(\widehat{\delta}_j^{[c]})^{-1} \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]}). \quad (49)$$

We can derive the first order conditions of the minimization problem in (44) as:

$$\frac{\partial Q(\theta)}{\partial \gamma_j} = -2V_{12j}' \left( \widehat{\delta}_j^{[c]} - \boldsymbol{\beta} \right) - 2D_j^{-1} \left( \widehat{\delta}_j^{[nc]} - \gamma_j \right) = 0; j = 1, \dots, J \quad (50)$$

$$\frac{\partial Q(\theta)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^J \left[ -2V_{11j} \left( \widehat{\delta}_j^{[c]} - \boldsymbol{\beta} \right) - 2V_{12j} \left( \widehat{\delta}_j^{[nc]} - \gamma_j \right) \right] = 0 \quad (51)$$

Substituting the expression of  $V_{12j}'$  in (50), we obtain

$$\left( \widehat{\delta}_j^{[nc]} - \widehat{\gamma}_j^{MD} \right) = D_j D_j^{-1} \text{Cov}(\widehat{\delta}_j^{[c]}, \widehat{\delta}_j^{[nc]})' \text{Var}(\widehat{\delta}_j^{[c]})^{-1} \left( \widehat{\delta}_j^{[c]} - \widehat{\boldsymbol{\beta}}^{MD} \right). \quad (52)$$

from which the expression for the MD estimator of each  $\widehat{\gamma}_j^{MD}$  in (17) follows.

On the other hand, we can insert the expression of  $\left( \widehat{\delta}_j^{[nc]} - \widehat{\gamma}_j^{MD} \right)$  from (52) into (51) to get

$$\sum_{j=1}^J \left[ V_{11j} \left( \widehat{\delta}_j^{[c]} - \widehat{\boldsymbol{\beta}}^{MD} \right) - V_{12j} D_j V_{12j}' \left( \widehat{\delta}_j^{[c]} - \widehat{\boldsymbol{\beta}}^{MD} \right) \right] = 0. \quad (53)$$



Then using the expressions for  $V_{11j}$  and  $V_{12j}$  it simplifies to

$$\left( \sum_{j=1}^J Var \left( \widehat{\delta}_j^{[c]} \right)^{-1} \right) \widehat{\beta}^{MD} = \sum_{j=1}^J Var \left( \widehat{\delta}_j^{[c]} \right)^{-1} \widehat{\delta}_j^{[c]} \quad (54)$$

and we obtain the expression for the MD estimator of the common parameters in (16).

Notice that for estimation the elements of  $Var \left( \widehat{\delta}_j \right)$  in (45) are replaced by the consistent estimators of them that we obtain in the first step.

## B Appendix: Other simulation tables

Table B.1: Examples of (un)balanced structures

$t =$		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>Balanced case</b>							
<b>J = 0,</b>	For $N$ units	x	x	x	x	x	x
<b>Double unbalancedness</b>							
<b>J = 2,</b>	For $N/2$ units	x	x	x	x	x	.
	For $N/2$ units	.	x	x	x	x	x
<b>J = 4,</b>	For $N/4$ units	x	x	x	x	x	.
	For $N/4$ units	x	x	x	x	.	.
	For $N/4$ units	.	x	x	x	x	x
	For $N/4$ units	.	.	x	x	x	x
<b>J = 6,</b>	For $N/6$ units	x	x	x	x	x	.
	For $N/6$ units	x	x	x	x	.	.
	For $N/6$ units	x	x	x	.	.	.
	For $N/6$ units	.	x	x	x	x	x
	For $N/6$ units	.	.	x	x	x	x
	For $N/6$ units	.	.	.	x	x	x
<b>Left-side Unbalancedness</b>							
<b>J = 2,</b>	For $N/2$ units	x	x	x	x	x	x
	For $N/2$ units	.	x	x	x	x	x
<b>J = 3,</b>	For $N/3$ units	x	x	x	x	x	x
	For $N/3$ units	.	x	x	x	x	x
	For $N/3$ units	.	.	x	x	x	x
<b>J = 4,</b>	For $N/4$ units	x	x	x	x	x	x
	For $N/4$ units	.	x	x	x	x	x
	For $N/4$ units	.	.	x	x	x	x
	For $N/4$ units	.	.	.	x	x	x

Note: “x” denotes that individuals are observed in that period and “.” that they are not.

Table B.2: Simulation results on the estimation of  $\alpha$ . Baseline case. Left-side Unbalancedness

	Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML	Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML
	$\hat{\alpha}$				RMSE			
T=4								
J=0	0.7502	0.7502	0.7502	0.7502	0.0860	0.0860	0.0860	0.0860
J=2		0.7325	0.7430	0.7376		0.1103	0.0947	0.0942
T=6								
J=0	0.7508	0.7508	0.7508	0.7508	0.0579	0.0580	0.0580	0.0580
J=4		0.7416	0.7490	0.7449		0.1014	0.0673	0.0668
T=8								
J=4	0.7388	0.7494	0.7516	0.7505	0.0723	0.0778	0.0509	0.0507
J=5	0.7386	0.7440	0.7480	0.7462	0.0931	0.0902	0.0509	0.0506
J=6		0.7427	0.7510	0.7480		0.0983	0.0564	0.0560
T=10								
J=4	0.7386	0.7489	0.7479	0.7476	0.0504	0.0667	0.0399	0.0398
J=5	0.7415	0.7439	0.7488	0.7482	0.0581	0.0768	0.0419	0.0419
J=6	0.7438	0.7472	0.7493	0.7484	0.0746	0.0829	0.0438	0.0437
J=7	0.7427	0.7500	0.7501	0.7488	0.0956	0.0914	0.0456	0.0456
J=8		0.7456	0.7498	0.7476		0.0960	0.0467	0.0466
T=15								
J=9	0.7438	0.7460	0.7486	0.7483	0.0518	0.0760	0.0336	0.0337
J=10	0.7495	0.7458	0.7496	0.7491	0.0608	0.0835	0.0349	0.0350
J=11	0.7487	0.7538	0.7521	0.7515	0.0765	0.0860	0.0344	0.0344
J=12	0.7404	0.7472	0.7483	0.7473	0.0997	0.0892	0.0362	0.0361
J=13		0.7493	0.7491	0.7478		0.0940	0.0380	0.0380

Table B.3: Simulation results on the estimation of the AMEs. Baseline case. Left-side Unbalancedness.

		Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML	Bal. Periods	Bal. Units	Unbal. MD	Unbal. ML
	AME	$\widehat{AME}$				RMSE			
T=4									
J=0	0.2023	0.2096	0.2096	0.2096	0.2023	0.0298	0.0298	0.0298	0.0284
J=2	0.2278		0.2684	0.2306	0.2227		0.0613	0.0362	0.0334
T=6									
J=0	0.2020	0.2097	0.2096	0.2096	0.2024	0.0206	0.0206	0.0206	0.0187
J=4	0.2398		0.2732	0.2424	0.2372		0.0528	0.0260	0.0244
T=8									
J=4	0.2364	0.2361	0.2764	0.2419	0.2364	0.0265	0.0507	0.0260	0.0178
J=5	0.2401	0.2388	0.2745	0.2433	0.2385	0.0356	0.0494	0.0203	0.0183
J=6	0.2438		0.2738	0.2472	0.2426		0.0495	0.0215	0.0206
T=10									
J=4	0.2347	0.2364	0.2766	0.2395	0.2338	0.0181	0.0495	0.0157	0.0139
J=5	0.2376	0.2404	0.2745	0.2422	0.2367	0.0216	0.0478	0.0161	0.0147
J=6	0.2404	0.2432	0.2758	0.2444	0.2396	0.0286	0.0480	0.0172	0.0155
J=7	0.2431	0.2441	0.2773	0.2470	0.2423	0.0378	0.0494	0.0174	0.0164
J=8	0.2461		0.2750	0.2492	0.2449		0.0472	0.0178	0.0171
T=15									
J=9	0.2412	0.2469	0.2757	0.2459	0.2406	0.0203	0.0452	0.0128	0.0118
J=10	0.2429	0.2496	0.2754	0.2474	0.2424	0.0242	0.0456	0.0133	0.0123
J=11	0.2449	0.2499	0.2784	0.2498	0.2453	0.0300	0.0469	0.0136	0.0122
J=12	0.2469	0.2470	0.2757	0.2500	0.2456	0.0397	0.0446	0.0135	0.0129
J=13	0.2489		0.2768	0.2516	0.2477		0.0456	0.0143	0.0138

Table B.4: Simulation results on the estimation of  $\alpha$ . Baseline case but with  $N = 500$ .

Panel A: Double Unbalancedness							
		Bal.	Bal.	Unbal.	Bal.	Bal.	Unbal.
		Periods	Units	MD	Periods	Units	MD
$\alpha = 0.75$		$\hat{\alpha}$			RMSE		
T=4	J=0	0.7472	0.7472	0.7472	0.1280	0.1280	0.1280
	J=2		0.7479	0.7192		0.2602	0.1718
T=6	J=0	0.7518	0.7518	0.7518	0.0796	0.0796	0.0796
	J=2	0.7250	0.7523	0.7502		0.1225	0.0923
	J=4		0.7443	0.7468		0.2171	0.1008
	J=6		0.7489	0.7479		0.2604	0.1050
T=8	J=2	0.7277	0.7497	0.7465	0.0815	0.1085	0.0675
	J=4	0.7460	0.7503	0.7506		0.1325	0.0728
	J=6		0.7455	0.7477		0.1999	0.0769
	J=8		0.7429	0.7525		0.2372	0.0844
	J=10		0.7387	0.7485		0.2530	0.0842
T=10	J=6	0.7521	0.7516	0.7533	0.1363	0.1682	0.0599
	J=8		0.7543	0.7508		0.1875	0.0627
	J=10		0.7416	0.7472		0.2261	0.0664
	J=12		0.7346	0.7480		0.2474	0.0717
	J=14		0.7434	0.7490		0.2627	0.0735
T=15	J=16		0.7475	0.7484		0.2162	0.0500

  

Panel B: Left-side Unbalancedness							
		Bal.	Bal.	Unbal.	Bal.	Bal.	Unbal.
		Periods	Units	MD	Periods	Units	MD
$\alpha = 0.75$		$\hat{\alpha}$			RMSE		
T=4	J=0	0.7472	0.7472	0.7472	0.1280	0.1280	0.1280
	J=2		0.7178	0.7327		0.1582	0.1318
T=6	J=0	0.7518	0.7518	0.7518	0.0796	0.0796	0.0796
	J=4		0.7300	0.7487		0.1396	0.0922
T=8	J=4	0.7393	0.7419	0.7485	0.0995	0.1085	0.0699
	J=5	0.7355	0.7420	0.7498		0.1319	0.0719
	J=6		0.7340	0.7519		0.1332	0.0774
T=10	J=4	0.7421	0.7491	0.7498	0.0745	0.0927	0.0580
	J=5	0.7425	0.7442	0.7491		0.0817	0.0584
	J=6	0.7454	0.7455	0.7504		0.1020	0.0622
	J=7	0.7421	0.7422	0.7497		0.1293	0.0640
	J=8		0.7442	0.7514		0.1374	0.0650
T=15	J=9	0.7437	0.7466	0.7472	0.0718	0.1059	0.0473
	J=10	0.7449	0.7476	0.7484		0.0860	0.0474
	J=11	0.7543	0.7556	0.7550		0.1044	0.0505
	J=12	0.7483	0.7478	0.7502		0.1381	0.0508
	J=13		0.7526	0.7493		0.1248	0.0518

Table B.5: Simulation results on the estimation of the AMEs. Baseline case but with  $N = 500$ .

Panel A: Double Unbalancedness								
			Bal. Periods	Bal. Units	Unbal. MD	Bal. Periods	Bal. Units	Unbal. MD
		AME	$\widehat{AME}$			RMSE		
T=4	J=0	0.2020	0.2094	0.2094	0.2094	0.0431	0.0431	0.0431
	J=2	0.2195		0.1801	0.2148		0.0778	0.0581
T=6	J=0	0.2022	0.2106	0.2106	0.2106	0.0276	0.0276	0.0276
	J=2	0.2194	0.2165	0.1811	0.2280	0.0428	0.0559	0.0331
	J=4	0.2277		0.1819	0.2332		0.0761	0.0358
	J=6	0.2291		0.1838	0.2324		0.0851	0.0370
T=8	J=2	0.2194	0.2182	0.1800	0.2271	0.0275	0.0498	0.0243
	J=4	0.2281	0.2420	0.1827	0.2352	0.0529	0.0644	0.0267
	J=6	0.2298		0.1828	0.2352		0.0737	0.0275
	J=8	0.2302		0.1843	0.2372		0.0811	0.0307
	J=10	0.2301		0.1820	0.2335		0.0868	0.0296
T=10	J=6	0.2299	0.2491	0.1842	0.2381	0.0572	0.0669	0.0226
	J=8	0.2306		0.1872	0.2378		0.0706	0.0233
	J=10	0.2306		0.1837	0.2361		0.0809	0.0240
	J=12	0.2308		0.1814	0.2356		0.0856	0.0255
	J=14	0.2304		0.1842	0.2346		0.0879	0.0257
T=15	J=16	0.2313		0.1855	0.2374		0.0791	0.0187

  

Panel B: Left-side Unbalancedness								
			Bal. Periods	Bal. Units	Unbal. MD	Bal. Periods	Bal. Units	Unbal. MD
		AME	$\widehat{AME}$			RMSE		
T=4	J=0	0.2020	0.2094	0.2094	0.2094	0.0431	0.0431	0.0431
	J=2	0.2276		0.2625	0.2285		0.0723	0.0466
T=6	J=0	0.2022	0.2106	0.2106	0.2106	0.0276	0.0276	0.0276
	J=4	0.2397		0.2679	0.2440		0.0613	0.0340
T=8	J=4	0.2364	0.2367	0.2731	0.2417	0.0369	0.0561	0.0255
	J=5	0.2402	0.2377	0.2735	0.2449	0.0501	0.0589	0.0265
	J=6	0.2439		0.2702	0.2483		0.0577	0.0288
T=10	J=4	0.2345	0.2378	0.2766	0.2408	0.0275	0.0554	0.0217
	J=5	0.2376	0.2406	0.2746	0.2428	0.0304	0.0544	0.0215
	J=6	0.2404	0.2443	0.2751	0.2457	0.0393	0.0567	0.0230
	J=7	0.2431	0.2442	0.2732	0.2472	0.0507	0.0583	0.0235
	J=8	0.2461		0.2740	0.2501		0.0597	0.0241
T=15	J=9	0.2416	0.2467	0.2756	0.2454	0.0274	0.0527	0.0171
	J=10	0.2430	0.2479	0.2758	0.2470	0.0332	0.0530	0.0175
	J=11	0.2449	0.2552	0.2792	0.2506	0.0413	0.0585	0.0188
	J=12	0.2469	0.2501	0.2759	0.2507	0.0544	0.0556	0.0187
	J=13	0.2491		0.2773	0.2518		0.0550	0.0189

Table B.6: Simulation results. Baseline case but with the Unbalancedness and the initial condition uncorrelated with  $\eta$  and  $N = 500$ .

Panel A: Double Unbalancedness

		Bal. Units	Bal. Units*	MD	Indep_ML*	Bal. Units	Bal. Units*	MD	Indep_ML*
$\alpha = 0.75$		$\hat{\alpha}$				RMSE			
T=4	J=0	0.7532	0.7532	0.7532	0.7532	0.1212	0.1213	0.1212	0.1213
	J=2			0.7633	0.7409			0.2115	0.1848
T=6	J=0	0.7502	0.7502	0.7502	0.7502	0.0833	0.0833	0.0833	0.0833
	J=2	0.7567	0.7538	0.7557	0.7530	0.1620	0.1617	0.1046	0.1025
	J=4			0.7588	0.7513			0.1178	0.1100
	J=6			0.7715	0.7543			0.1366	0.1253
T=8	J=2	0.7518	0.7500	0.7513	0.7505	0.1002	0.0999	0.0801	0.0796
	J=4	0.7616	0.7561	0.7513	0.7504	0.1673	0.1665	0.0856	0.0850
	J=6			0.7518	0.7493			0.0912	0.0892
	J=8			0.7558	0.7460			0.0984	0.0956
	J=10			0.7594	0.7474			0.1072	0.1057
T=10	J=6	0.7672	0.7608	0.7505	0.7500	0.1652	0.1638	0.0699	0.0697
	J=8			0.7513	0.7494			0.0739	0.0732
	J=10			0.7526	0.7471			0.0791	0.0783
	J=12			0.7572	0.7459			0.0843	0.0853
	J=14			0.7590	0.7530			0.0887	0.0906
T=15	J=16			0.7475	0.7438			0.0559	0.0564

\* indicates that this estimator has been computed using Heckman's approach.

Table B.6: Simulation results. Baseline case but with the Unbalancedness and the initial condition uncorrelated with  $\eta$  and  $N = 500$ . (cont'd)

Panel B: Left-side Unbalancedness

		Bal. Units	Bal. Units*	MD	Indep_ML	Bal. Units	Bal. Units*	MD	Indep_ML		
$\alpha = 0.75$		$\hat{\alpha}$				RMSE					
T=4	J=0	0.7532	0.7532	0.7532	0.7532	0.1212	0.1213	0.1212	0.1213		
	J=2			0.7590	0.7482			0.1595	0.1482		
T=6	J=0	0.7502	0.7502	0.7502	0.7502	0.0833	0.0833	0.0833	0.0833		
	J=4			0.7667	0.7514			0.1304	0.1230		
T=8	J=4	0.7537	0.7501	0.7520	0.7493	0.1229	0.1224	0.0895	0.0876		
	J=5	0.7553	0.7502	0.7570	0.7492			0.1667	0.1667	0.0979	0.0944
	J=6			0.7590	0.7474					0.1068	0.1035
T=10	J=4	0.7506	0.7488	0.7502	0.7494	0.0836	0.0836	0.0681	0.0679		
	J=5	0.7473	0.7447	0.7491	0.7480			0.0986	0.0981	0.0719	0.0717
	J=6	0.7468	0.7435	0.7494	0.7458			0.1199	0.1194	0.0766	0.0754
	J=7	0.7467	0.7414	0.7527	0.7441	0.1648	0.1653	0.0827	0.0814		
	J=8			0.7553	0.7480			0.0875	0.0829		
T=15	J=9	0.7540	0.7520	0.7510	0.7497	0.0842	0.0838	0.0548	0.0547		
	J=10	0.7533	0.7507	0.7506	0.7484			0.0961	0.0955	0.0575	0.0570
	J=11	0.7563	0.7523	0.7519	0.7452			0.1219	0.1210	0.0606	0.0606
	J=12	0.7568	0.7527	0.7540	0.7406			0.1655	0.1665	0.0629	0.0643
	J=13			0.7557	0.7399					0.0663	0.0648

\* indicates that this estimator has been computed using Heckman's approach.



Table B.7: Simulation results. Baseline case but with Smaller state dependence ( $\alpha = 0.50$ ) and  $N = 500$ .

Panel A: Double Unbalancedness					
		Bal. Units	Unbal. MD	Bal. Units	Unbal. MD
$\alpha = 0.50$		$\hat{\alpha}$		RMSE	
T=4	J=0	0.5014	0.5014	0.1199	0.1199
	J=2		0.5107		0.2168
T=6	J=0	0.4964	0.4964	0.0801	0.0801
	J=2	0.4991	0.4999	0.1516	0.1007
	J=4		0.5020		0.1150
	J=6		0.5168		0.1330
T=8	J=2	0.5024	0.5017	0.0942	0.0771
	J=4	0.5109	0.5018	0.1550	0.0821
	J=6		0.5020		0.0871
	J=8		0.5060		0.0962
	J=10		0.5108		0.1059
T=10	J=6	0.5128	0.4998	0.1488	0.0660
	J=8		0.4996		0.0701
	J=10		0.5001		0.0743
	J=12		0.5037		0.0789
	J=14		0.5061		0.0845
T=15	J=16		0.4964		0.0543

  

Panel B: Left-side Unbalancedness					
		Bal. Units	Unbal. MD	Bal. Units	Unbal. MD
$\alpha = 0.50$		$\hat{\alpha}$		RMSE	
T=4	J=0	0.5014	0.5014	0.1199	0.1199
	J=2		0.5063		0.1597
T=6	J=0	0.4964	0.4964	0.0801	0.0801
	J=4		0.5079		0.1199
T=8	J=4	0.5041	0.5020	0.1141	0.0874
	J=5	0.5074	0.5048	0.1542	0.0933
	J=6		0.5097		0.1018
T=10	J=4	0.5003	0.5000	0.0775	0.0626
	J=5	0.4968	0.4992	0.0920	0.0665
	J=6	0.4956	0.4998	0.1093	0.0870
	J=7	0.4795	0.5032	0.1517	0.0751
	J=8		0.5060		0.0805
T=15	J=9	0.5006	0.4995	0.0798	0.0534
	J=10	0.5016	0.4987	0.0919	0.0559
	J=11	0.5030	0.4998	0.1136	0.0590
	J=12	0.5065	0.5013	0.1539	0.0604
	J=13		0.5035		0.0627

Table B.8: Simulation results. Baseline case but with Higher state dependence ( $\alpha = 1$ ) and  $N = 500$ .

Panel A: Double Unbalancedness

		Bal. Units	Unbal. MD	Bal. Units	Unbal. MD
$\alpha = 1$		$\hat{\alpha}$		RMSE	
T=4	J=0	1.0029	1.0029	0.1206	0.1206
	J=2		1.0167		0.2136
T=6	J=0	1.0016	1.0016	0.0856	0.0856
	J=2	1.0112	1.0072	0.1691	0.1096
	J=4		1.0161		0.1249
	J=6		1.0220		0.1381
T=8	J=2	1.0030	1.0017	0.1042	0.0829
	J=4	1.0140	1.0010	0.1761	0.0889
	J=6		1.0037		0.0981
	J=8		1.0078		0.1051
	J=10		1.0106		0.1152
T=10	J=6	1.0184	1.0014	0.1854	0.0730
	J=8		1.0030		0.0772
	J=10		1.0060		0.0834
	J=12		1.0104		0.0879
	J=14		1.0100		0.0929
T=15	J=16		1.0001		0.0592

Panel B: Left-side Unbalancedness

		Bal. Units	Unbal. MD	Bal. Units	Unbal. MD
$\alpha = 1$		$\hat{\alpha}$		RMSE	
T=4	J=0	1.0029	1.0029	0.1206	0.1206
	J=2		1.0089		0.1608
T=6	J=0	1.0016	1.0016	0.0856	0.0856
	J=4		1.0232		0.1344
T=8	J=4	1.0050	1.0033	0.1367	0.0923
	J=5	1.0044	1.0095	0.1865	0.1012
	J=6		1.0113		0.1135
T=10	J=4	1.0031	1.0015	0.0923	0.0718
	J=5	1.0012	1.0010	0.1086	0.0765
	J=6	1.0005	1.0040	0.1352	0.0823
	J=7	1.0034	1.0076	0.1878	0.0890
	J=8		1.0081		0.0940
T=15	J=9	1.0055	1.0033	0.0934	0.0604
	J=10	1.0051	1.0027	0.1061	0.0626
	J=11	1.0080	1.0048	0.1350	0.0661
	J=12	1.0085	1.0067	0.1873	0.0691
	J=13		1.0065		0.0718

Table B.9: Simulation results. Baseline case but with the Unbalancedness uncorrelated with  $\eta$  and  $N = 500$ .

Panel A: Double Unbalancedness

		Bal. Units	Unbal. MD	Bal. Units	Unbal. MD
$\alpha = 0.75$		$\hat{\alpha}$		RMSE	
T=4	J=0	0.7535	0.7535	0.1258	0.1258
	J=2		0.7658		0.2207
T=6	J=0	0.7506	0.7506	0.0852	0.0852
	J=2	0.7580	0.7565	0.1642	0.1067
	J=4		0.7596		0.1203
	J=6		0.7727		0.1403
T=8	J=2	0.7518	0.7519	0.1007	0.0815
	J=4	0.7615	0.7521	0.1676	0.0873
	J=6		0.7528		0.0928
	J=8		0.7569		0.1006
	J=10		0.7608		0.1098
T=10	J=6	0.7672	0.7505	0.1657	0.0709
	J=8		0.7514		0.0749
	J=10		0.7526		0.0803
	J=12		0.7574		0.0855
	J=14		0.7594		0.0905
T=15	J=16		0.7482		0.0562

Panel B: Left-side Unbalancedness

		Bal. Units	Unbal. MD	Bal. Units	Unbal. MD
$\alpha = 0.75$		$\hat{\alpha}$		RMSE	
T=4	J=0	0.7535	0.7535	0.1258	0.1259
	J=2		0.7596		0.1659
T=6	J=0	0.7506	0.7506	0.0852	0.0852
	J=4		0.7684		0.1329
T=8	J=4	0.7537	0.7525	0.1230	0.0903
	J=5	0.7554	0.7577	0.1664	0.0989
	J=6		0.7594		0.1075
T=10	J=4	0.7506	0.7499	0.0837	0.0690
	J=5	0.7473	0.7489	0.0987	0.0725
	J=6	0.7468	0.7492	0.1199	0.0770
	J=7	0.7466	0.7525	0.1648	0.0831
	J=8		0.7552		0.0879
T=15	J=9	0.7540	0.7511	0.0842	0.0547
	J=10	0.7533	0.7508	0.0961	0.0574
	J=11	0.7563	0.7521	0.1219	0.0605
	J=12	0.7568	0.7541	0.1655	0.0629
	J=13		0.7589		0.0664

Table B.10: Simulation results on the estimation of the AMEs. Baseline case but with the Unbalancedness uncorrelated with  $\eta$  and  $N = 500$ .

Panel A: Double Unbalancedness							
		Bal. Units		Unbal. MD	Bal. Units		Unbal. MD
		AME	$\widehat{AME}$		RMSE		
T=4	J=0	0.2019	0.2034	0.2034	0.0416	0.0416	0.0764
	J=2	0.2019		0.2092			
T=6	J=0	0.2021	0.2024	0.2024	0.0273	0.0273	0.0362
	J=2	0.2021	0.2088	0.2052			
	J=4	0.2021		0.2068			
	J=6	0.2021		0.2093			
T=8	J=2	0.2021	0.2039	0.2031	0.0339	0.0265	0.0286
	J=4	0.2021	0.2109	0.2034			
	J=6	0.2021		0.2038			
	J=8	0.2021		0.2048			
	J=10	0.2021		0.2045			
T=10	J=6	0.2020	0.2133	0.2025	0.0625	0.0229	0.0244
	J=8	0.2020		0.2030			
	J=10	0.2020		0.2032			
	J=12	0.2020		0.2043			
	J=14	0.2020		0.2040			
T=15	J=16	0.2021		0.2018		0.0180	

  

Panel B: Left-side Unbalancedness							
		Bal. Units		Unbal. MD	Bal. Units		Unbal. MD
		AME	$\widehat{AME}$		RMSE		
T=4	J=0	0.2019	0.2034	0.2034	0.0416	0.0416	0.0589
	J=2	0.2019		0.2065			
T=6	J=0	0.2021	0.2024	0.2024	0.0273	0.0273	0.0462
	J=4	0.2021		0.2086			
T=8	J=4	0.2021	0.2058	0.2042	0.0436	0.0304	0.0335
	J=5	0.2021	0.2081	0.2060			
	J=6	0.2021		0.2055			
T=10	J=4	0.2020	0.2036	0.2026	0.0285	0.0226	0.0239
	J=5	0.2020	0.2028	0.2023			
	J=6	0.2020	0.2032	0.2026			
	J=7	0.2020	0.2048	0.2037			
	J=8	0.2020		0.2037			
T=15	J=9	0.2021	0.2051	0.2031	0.0286	0.0178	0.0188
	J=10	0.2021	0.2051	0.2030			
	J=11	0.2021	0.2071	0.2035			
	J=12	0.2021	0.2093	0.2040			
	J=13	0.2021		0.2039			

Table B.11: Percentage of Monte Carlo Simulations that achieved convergence for the Minimum Distance estimation

		Baseline	Uncorrelated IC and Unbalancedness	Higher State Dependence	Uncorrelated Unbalancedness
		Unbal. MD in Tables 1 & B.2	Unbal. MD in Table B.6	Unbal. MD in Table B.8	Unbal. MD in Table B.9
Panel A: Double Unbalancedness					
T=4	J=0	100.0	100.0	100.0	100.0
	J=2	99.0	100.0	100.0	100.0
T=6	J=0	100.0	100.0	100.0	100.0
	J=2	100.0	100.0	100.0	100.0
	J=4	100.0	100.0	99.9	100.0
	J=6	99.9	99.8	99.5	99.9
T=8	J=2	100.0	100.0	100.0	100.0
	J=4	100.0	100.0	100.0	100.0
	J=6	100.0	100.0	100.0	100.0
	J=8	100.0	100.0	100.0	100.0
	J=10	99.8	99.1	95.5	99.1
T=10	J=6	100.0	100.0	100.0	100.0
	J=8	100.0	100.0	100.0	100.0
	J=10	100.0	100.0	100.0	100.0
	J=12	100.0	100.0	99.7	100.0
	J=14	100.0	95.0	89.2	95.1
T=15	J=16	100.0	100.0	99.9	100.0
Panel B: Left-side Unbalancedness					
T=4	J=0	100.0	100.0	100.0	100.0
	J=2	99.9	100.0	100.0	100.0
T=6	J=0	100.0	100.0	100.0	100.0
	J=4	100.0	100.0	99.7	100.0
T=8	J=4	100.0	100.0	100.0	100.0
	J=5	100.0	100.0	100.0	100.0
	J=6	100.0	99.9	99.3	99.9
T=10	J=4	100.0	100.0	100.0	100.0
	J=5	100.0	100.0	100.0	100.0
	J=6	100.0	100.0	100.0	100.0
	J=7	100.0	100.0	100.0	100.0
	J=8	99.9	99.5	98.3	99.5
T=15	J=9	100.0	100.0	100.0	100.0
	J=10	100.0	100.0	100.0	100.0
	J=11	100.0	100.0	99.9	100.0
	J=12	100.0	100.0	99.7	100.0
	J=13	99.8	96.8	91.3	96.8

Note: In other specifications all simulations converged or the percentage of convergence very was close to 100%.

## C Appendix: Other application tables

Table C.1: Definition of variables and Descriptive statistics

Variable	Definition	Total	Exporters	Non-exporters
Export	Export activity (binary indicator)	0.52 (0.50)		
Size	(Number of employees)/100	1.77 (5.7)	2.95 (7.6)	0.48 (1.1)
R&D intensity	(R&D expenditure)/Sales	0.006 (0.02)	0.009 (0.02)	0.002 (0.01)
High skill	Share of workers with a university degree	2.93 (5.6)	3.7 (5.7)	2.0 (5.4)
Med. skill	Share of workers with a high-school degree	4.23 (6.9)	5.2 (7.3)	3.1 (6.2)
Age	Years since firm's creation /10	2.40 (2.3)	2.98 (2.5)	1.78 (1.9)
Number of Observations		13,203	6,887	6,316

Note: For each variable, Mean and Standard Deviation (in parentheses) are reported

Table C.2: Export decision estimates

Variables	Bal. Units	Ignore Unbal.	Unbal. MD
Export <sub>t-1</sub>	1.544 (0.094)	1.459 (0.075)	1.534 (0.074)
Size <sub>t</sub>	0.107 (0.119)	0.035 (0.085)	0.029 (0.086)
Trend	0.100 (0.015)	0.097 (0.011)	0.095 (0.013)
R&D_int <sub>t</sub>	-2.535 (1.994)	-1.412 (1.562)	-2.263 (1.752)
High_skill <sub>t</sub>	-0.010 (0.011)	-0.009 (0.008)	-0.011 (0.008)
Med_skill <sub>t</sub>	0.002 (0.007)	0.002 (0.005)	0.006 (0.005)
Age <sub>t</sub>	0.039 (0.030)	0.045 (0.021)	0.025 (0.021)
Export <sub>t<sub>i</sub></sub>	2.618 (0.239)	2.475 (0.177)	
$\overline{\text{Size}}$	0.106 (0.128)	0.211 (0.093)	
$\overline{\text{R\&D\_int}}$	6.695 (4.871)	7.123 (3.215)	
$\overline{\text{High\_skill}}$	0.000 (0.019)	0.003 (0.012)	
$\overline{\text{Med\_skill}}$	0.038 (0.015)	0.028 (0.009)	
Const.	-2.646 (0.165)	-2.693 (0.126)	
Num. obs.	6,291	10,876	10,876

Note: Standard errors are reported in parentheses.