UNIVERSIDAD CARLOS III DE MADRID working papers

UC3M Working Papers
Economics
15-03
ISSN 2340-5031
February 2015

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Estimation of Dynamic Nonlinear Random Effects Models with Unbalanced Panels[1]

Pedro Albarran[2], Raquel Carrasco[3] and Jesus M. Carro[4]

Abstract

This paper presents and evaluates estimation methods for dynamic nonlinear correlated random effects (CRE) models with unbalanced panels. Accounting for the unbalancedness is crucial in dynamic non-linear models and it cannot be ignored even if the process that produces it is completely at random. Available approaches to estimate dynamic CRE models accounting for the initial conditions problem were developed for balanced panels and they do not work with unbalanced panels. In this type of dynamic models, just ignoring the unbalancedness produces inconsistent estimates of the parameters. Another potential "solution", used by some practitioners, is to take the sub-sample that constitutes a balanced panel and then to estimate the model using the available methods. Nonetheless, this approach is not feasible in some cases because the constructed balanced panel might not contain enough number of common periods across individuals. Moreover, when feasible, it discards useful information, which, as we show, leads to important efficiency losses. In this paper we consider several scenarios in which the sample selection process can be arbitrarily correlated with the permanent unobserved heterogeneity. The approaches we propose exploit all the observations available, can be implemented using standard solutions to the initial conditions problem, and can be easily applied in the context of commonly used models, such as dynamic binary choice models.

JEL classification: C23, C25
Keywords: Unbalanced Panels, correlated random effects, dynamic non-linear models.

[1] We are grateful to Wiji Arulampalam, Ricardo Mora, and participants at the 2013 Econometric Society Asian Meeting and the XXXIX SAEe 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 ECO2009-11165 and ECO2012-31358.
[2] Universidad de Alicante. E-mail: albarran@ua.es
[3] Department of Economics, Universidad Carlos III de Madrid. E-mail: rcarras@eco.uc3m.es
[4] 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 correlated random effects (CRE) models with unbalanced panels. The CRE approach represents a simple method to estimate this type of models. Examples of applications using it are Hyslop (1999), Contoyannis et. al.(2004), Stewart (2007), and Akee et. al.(2010). Although at a cost of imposing restrictive parametric assumptions on the conditional distribution of the heterogeneity parameters, it is not subject to the incidental parameters problem that the fixed-effects (FE) approach suffers in nonlinear models.[1] Bias-corrected versions of FE estimators address the incidental parameters problem, but they usually require a greater number of periods for the bias adjustments to work than the available in many data sets.[2] Under these circumstances, correlated random effects methods can be regarded as an useful alternative.

In a dynamic setting the main drawback of the CRE approach is that it gives rise to the well-known initial conditions problem. Heckman (1981) and Wooldridge (2005) propose solutions to deal with this problem, but these solutions are developed for balanced panels and, in general, they cannot be directly implemented with unbalanced panel data which, in practice, are the norm. For example, in large panel data sets like the PSID for the U.S or the GESOEP for Germany, there are individuals who drop out (potentially non-randomly) of the sample. In other cases, like in the so called "rotating panels", the unbalancedness is generated by the sample design and, therefore, the missingness is completely at random (for instance, in the Monthly Retail Trade Survey for the U.S, or in the Household Budget Continuous Survey for Spain).

The initial conditions problem present in dynamic models with balanced panel data is augmented 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, the unbalancedness

[1] For the purpose of this paper, FE methods are those that leave the distribution of the individual specific parameter and its relation with the explanatory variables unrestricted, while CRE methods are those that impose a certain amount of structure in the conditional distribution of the individual effects. For a review of the literature on non-linear FE models see Arellano and Honoré (2001) and Arellano (2003).

[2] Some examples of bias-correcting the fixed effects estimators in dynamic models are Carro (2007), Fernandez-Val (2009), Bester and Hansen (2009), and Carro and Traferri (2014). Arellano and Hahn (2007) offer a review of this literature, further references and a general framework in which the various approaches can be included.

cannot be ignored for consistent estimation of dynamic models even if it is completely at random (i.e. independent of the process of the observables and the unobservables), whereas unbalancedness completely at random allows us to ignore the unbalancedness for consistent estimation of a static model. One solution typically applied by practitioners is to take the subset of observations constituting a balanced panel, and then to use the available methods (see Wooldridge, 2005, pp. 44). Nonetheless, this approach is in some cases unfeasible because the constructed balanced panel might not contain enough number of common periods across individuals. Moreover, when feasible, reducing the data set to make the panel balanced will discard useful information, which may imply important efficiency losses.

It is important to note that previous problems are still present in the traditional RE model that assumes that the heterogeneity is independent of the time-varying covariates. The CRE case that we cover in detail contains the RE as a particular case. Even if we extend the independency assumption of the RE to the unbalanced process, the dynamic nature of the model would still give the inconsistency problems previously pointed out when the panel is unbalanced.[3]

To the best of our knowledge only Wooldridge (2010) addresses the issue of estimating CRE models with unbalanced panels, but considering only static models. Specifically, he proposes several strategies for allowing the unobserved heterogeneity to be correlated with the observed covariates and the selection mechanism for unbalanced panels. However, the assumption of strict exogeneity of the covariates is very restrictive, and the solutions in Wooldridge (2010) cannot be directly extended to dynamic models because, as said, the unbalancedness also affects how we have to deal with the initial conditions.

In this paper we present several scenarios in which the sample selection process can be arbitrarily correlated with the time invariant unobserved heterogeneity. Our study goes beyond a theoretical discussion on how to address the estimation of dynamic CRE models with unbalanced panels and proposes practical solutions, describing how they can be implemented using standard software typically used by practitioners. Although our implementation section focuses on the dynamic binary probit model, our

[3]Examples of papers using the RE approach are Arulampalam and Stewart (2009), Campa (2004), or Bernard and Jensen (2004).

theoretical discussion covers other commonly used models, such as the ordered probit or the Tobit model, and our proposals for implementing the estimation in practice can be directly extended to these other nonlinear models.

The paper is organized as follows. Section 2 presents the model, the likelihood that accounts for the unbalancedness, and the consequences of that for existing estimators. Section 3 describes how the estimation of dynamic models with unbalanced panels can be implemented and other practical issues. In Section 4 we study the finite sample properties of the proposed estimators in a binary probit model with a single lagged dependent variable by means of Monte Carlo simulations. Finally, Section 5 concludes.

2 Model framework

2.1 The general case

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

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

where $i = 1, ..., N$ represents cross-sectional units, y_{it} is the potentially observed outcome, and X_{it} are potentially observed 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 the unbalanced one, 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.

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

and

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

is the number of periods we observe for unit i. Another characteristic of the panels considered is that all the observations for unit i are consecutive. This means that

$$s_{it} = 1 \; \forall \; t \in [t_i, t_i + T_i]$$
$$s_{it} = 0 \; \forall \; t < t_i \text{ or } t > t_i + T_i$$

We denote by J de number of different S_i sequences that we have in the panel. Finally, we consider panels where N is large and T and J are small relative to N.

We are interested in the conditional distribution $F(y_{it} \mid y_{it-1}, X_i, \eta_i)$ where η_i denotes the vector of permanent unobserved heterogeneity. Through this paper we make the following assumption:

Assumption 1:

$$F(y_{it} \mid y_{it-1}, X_i, \eta_i, S_i) = F(y_{it} \mid y_{it-1}, X_i, \eta_i)$$

It means that the sample selection process s_{it} is strictly exogenous with respect to the idiosyncratic shocks to y_{it}, although it is allowed to be correlated with η_i and the observed covariates. This assumption is present also in most analysis with unbalanced panels taking the fixed effects approach.

Let $f(y_{it} \mid y_{it-1}, X_i, \eta_i, S_i; \beta)$ be the correctly specified density for the conditional distribution on previous equation and $h(\eta_i|X_i, S_i; \beta_\eta)$ a correctly specified density of $F(\eta_i|X_i, S_i)$.

The density of $(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT})$ for a given individual is

$$\begin{aligned} f\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i\right) &= \prod_{t=1}^{T} f\left(y_{it}|y_{it-1}, X_i, S_i\right)^{s_{it}s_{it-1}} f\left(y_{it}|X_i, S_i\right)^{s_{it}(1-s_{it-1})} \qquad (1) \\ &= \prod_{t=t_i+1}^{t_i+T_i} f\left(y_{it}|y_{it-1}, X_i, S_i\right) f\left(y_{it_i}|X_i, S_i\right) \end{aligned}$$

Previous equation can be written as

$$f\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i\right) = \int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i} f\left(y_{it}|y_{it-1}, X_i, S_i, \eta_i\right) f\left(y_{it_i}|X_i, S_i, \eta_i\right) h(\eta_i|X_i, S_i) d\eta_i, \qquad (2)$$

or as

$$f(s_{i1}y_{i1},\ldots,s_{iT}y_{iT}|X_i,S_i)=\left[\int_{\eta_i}\prod_{t=t_i+1}^{t_i+T_i}f(y_{it}|y_{it-1},X_i,S_i,\eta_i)\,h(\eta_i|y_{it_i},X_i,S_i)d\eta_i\right]f(y_{it_i}|X_i,S_i) \tag{3}$$

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.

Given that previous equations depend on unobservables, η_i, if the start of the sample does not coincide with the start of the stochastic process, the first observation will not be independent of η_i. Moreover, $f(y_{it_i}|X_i,S_i,\eta_i)$ in equation (2) or $h(\eta_i|y_{it_i},X_i,S_i)$ in equation (3) are different for each sub-panel with different observed periods S_i. Writing an equation for $f(y_{i1}|X_i,\eta_i)$ and $h(\eta_i|X_i,S_i)$, or for $h(\eta_i|y_{i1},X_i)$, as Heckman (1981) and Wooldridge (2005) do respectively for the balanced case, is not enough to solve the initial conditions problem. In general, unless $t_i=1$, $f(y_{it_i}|X_i,S_i,\eta_i)\neq f(y_{i1}|X_i,\eta_i)$ and $h(\eta_i|y_{it_i},X_i,S_i)\neq h(\eta_i|y_{i1},X_i)$.

The framework considered so far includes situations in which the selection mechanism S_i is correlated with the individual effect, η_i. If we write the likelihood of the data using expression (2), a different distribution of the initial conditions and of the unobserved effects for each sub-panel are required. That is, the densities $f(y_{it_i}|X_i,S_i,\eta_i;\delta_{S_i})$ and $h(\eta_i|X_i,S_i;\beta_{\eta S_i})$ in (2) depend on a vector of parameters different for each sub-panel. This implies that $f(y_{it_i}\mid X_i=x,\eta_i=\eta,S_i)\neq f(y_{jt_j}\mid X_j=x,\eta_j=\eta,S_j)$ for $S_i\neq S_j$, and they will be the same whenever $S_i=S_j=s$.

In equation (3), we need to specify the density of η_i conditional on the initial observation, $h(\eta_i|y_{it_i},X_i,S_i;\pi_{S_i})$. This will depend on different parameters for each sub-panel, and we can discard $f(y_{it_i}|X_i,S_i)$ because that term is outside the integral.

Notice that the unbalancedness S_i is completely defined by two elements: the period each sub-panel starts, t_i, and the number of periods of each sub-panel, T_i. Depending on the unbalancedness structure, we could have that the correlation between S_i and η_i is only through t_i.[4] In that case, previous densities will depend on different vectors

[4] Notice that, unless determined by sample design, whether η_i is correlated with S_i or only with t_i should be part of the assumptions made about the process.

of parameters for each starting period (i.e. different t_i instead of different S_i).

2.2 Unbalancedness independent of the individual effect

In this case, in addition to **Assumption 1**, we assume,

Assumption 2: Unbalancedness is independent of η_i.

This assumption is relevant, for instance, when having rotating panels. Assumption 2 means that $h(\eta_i|X_i, S_i) = h(\eta_i|X_i)$, that is, the conditional distribution $\eta_i|X_i, S_i$ does not depend on S_i. However, even under Assumption 2 $f(y_{it_i} \mid X_i, \eta_i, S_i)$ 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, even though we have assumed that the sample selection process S_i is independent of η_i, the density of the unobserved effects conditional on the initial conditions, $h(\eta_i|y_{it_i}, X_i, S_i)$ will be different for each t_i, unless the process is not dynamic or it is in its steady state since period $t = 0$.

2.3 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. Suppose that the correct joint density of $(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i)$ is given by equation (2). Instead, the density used to write the likelihood when ignoring the unbalancedness is

$$f\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i\right) = \int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i} f\left(y_{it}|y_{it-1}, X_i, \eta_i\right) f\left(y_{it_i}|X_i, \eta_i\right) h(\eta_i|X_i) d\eta_i. \quad (4)$$

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

$$f\left(y_{it}|y_{it-1}, X_i, S_i, \eta_i\right) = f\left(y_{it}|y_{it-1}, X_i, \eta_i\right)$$

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

(i) $h(\eta_i|X_i)$ in (4) to satisfy

$$h(\eta_i|X_i) = \int_{S_i} h(\eta_i|X_i, S_i) dG_{S_i},$$

(ii) Regarding $f\left(y_{it_i}|X_i, S_i, \eta_i\right)$, model (4) imposes that all units have the same distribution for the initial condition regardless the period t_i at which they enter the panel. That is, (4) is imposing that

$$f\left(y_{i1}|X_i, \eta_i\right) = f\left(y_{i2}|X_i, \eta_i\right) = ...,$$

and these densities are different, except if

1. 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)

   and

2. S_i is independent from the shocks to the initial conditions.

Unless these two conditions are both satisfied, the estimates obtained by ignoring the unbalancedness are inconsistent.

Notice that the assumption that S_i is independent from the shocks to the initial conditions is not enough to ensure that the conditional densities for each initial observational period coincide. For example, suppose that we have two individuals, 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. Then,

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

$$\Pr(y_{i2} \mid \eta_i, S_i) = \sum_{y1} \Pr(y_{i2} \mid y_{i1}, \eta_i) \cdot \Pr(y_{i1} \mid \eta_i)$$

The two probabilities are different unless y_{i1} is at the steady state.

Also, notice that S_i being independent from η_i (Assumption 2), is not enough to allow us to ignore the unbalancedness. Under Assumption 2, condition (i) above is satisfied, but, again, that does not imply that condition (ii) is satisfied.

2.4 Taking the balanced sub-sample

Wooldridge (2005) points out that a potential solution to the unbalancedness under independence between S_i and the idiosyncratic shocks to y_{it} is to use the subset of observations constituting a balanced panel. Then, one could apply to that balanced sample the standard Heckman's or Wooldridge's methods to solve 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 unfeasible.[5] Let us look at this approach in more detail.

Suppose that the correct conditional density of $s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i$ is given by (3), excluding the term for the initial observations $f\left(y_{it_i}|X_i, S_i\right)$. Instead of that, the likelihood function that is maximized when making the panel balanced is

$$f\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i\right) = \int_{\eta_i} \prod_{t=\max t_i+1}^{\min(t_i+T_i)} f\left(y_{it}|y_{it-1}, X_i, \eta_i\right) h(\eta_i|y_{i\max t_i}, X_i)d\eta_i \qquad (5)$$

Under Assumption 1 $f\left(y_{it}|y_{it-1}, X_i, S_i, \eta_i\right) = f\left(y_{it}|y_{it-1}, X_i, \eta_i\right)$. Thus, in order to have a consistent ML Estimator of the parameters of the conditional distribution of $y_{it}|y_{it-1}, X_i, \eta_i$ based on (5) we need

$$h(\eta_i|y_{i\max t_i}, X_i) = \int_{S_i} h(\eta_i|y_{i\max t_i}, X_i, S_i)dG_{S_i}.$$

So, as long as the $h(\eta_i|y_{i\max t_i}, X_i)$ we specify satisfies this condition and we have enough periods in the balanced sample, the MLE based on (5) will be consistent, though less efficient. However, depending on the nature of $h(\eta_i|y_{i\max t_i}, X_i, S_i)$ (i.e. depending on the nature of the relations between η_i and S_i and the evolution of the distribution of y_{it} across periods and sub-panels) approximating $h(\eta_i|y_{i\max t_i}, X_i)$ may require a complex distribution even if $h(\eta_i|y_{i\max t_i}, X_i, S_i)$ is the standard normal distribution.[6]

For completeness we should mention another way of obtaining a balanced sub-panel from the original sample. Since the entire unbalanced panel contains several

[5] For example, in a rotating panel with $T = 5$ with three subpanels where each subpanel lasts for three periods (i.e. $T_i = 3$), the first subpanel starts at $t_i = 1$, the second at $t_i = 2$, and the third at $t_i = 3$, the subpanels only have one period in common, less than the 3 periods needed for estimation.

[6] See next Section for a discussion of the problems with the practical implementation of this approach.

balanced sub-panels, one can just take one of these (of course, one as long as possible in the time dimension). 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. For example, it may be the case that nobody is observed in all the periods of the panel, but a group of individuals is observed from the second to the last wave of the panel; we can take this group as our balanced sub-sample.

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 potential infeasibility problem when there are no enough periods in which all individuals are observed. However, this does not allow to identify the average marginal effect of covariates because, although the common parameters of the conditional model may be correctly estimated using only this sub-sample, the conditional distribution of the heterogenous individual effects will only be valid for this particular sub-group of individuals. Unless Assumption 2 (Unbalancedness independent of η_i) is imposed, the distribution of η_i for this balanced sub-sample is different from the distribution of η_i for the entire sample. Given that we focus on methods that, if feasible, can be valid both under Assumption 2 and under the more general case that allows for correlation between η_i and the unbalancedness, in the rest of the paper we will only consider the way of obtaining a balanced sub-sample described in the previous paragraphs of this section.

3 Implementation and other practical issues

In this Section we show how the results from previous Section can be applied with specific assumptions about parametric distributions. We have chosen one of the most common distribution assumed in empirical works, but it can be generalized to other non-linear models and parametric distributions.

3.1 A general case that allows for correlation between η_i and the unbalancedness

Let us consider the following dynamic discrete choice model:

$$y_{it} = 1\left\{\alpha y_{it-1} + \beta_0 + X'_{it}\beta + \eta_i + \varepsilon_{it} \geq 0\right\} \tag{6}$$

$$-\varepsilon_{it}|\, y_{it-1}, X_i, \eta_i, S_i \underset{iid}{\sim} N(0,1) \tag{7}$$

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

$$\Pr\left(S'_1Y_1, \ldots, S'_NY_N|\, X_1, \ldots, X_N, S_1, \ldots, S_N\right) = \prod_{i=1}^{N} \Pr\left(S'_iY_i|X_i, S_i\right) = \prod_{i=1}^{N} \Pr\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i\right)$$

Thus, for each $i = 1, ..., N$,

$$\begin{aligned}\Pr\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i\right) &= \prod_{t=1}^{T} \Pr\left(y_{it}|y_{it-1}, X_i, S_i\right)^{s_{it}s_{it-1}} \Pr\left(y_{it}|X_i, S_i\right)^{s_{it}(1-s_{it-1})} \\ &= \prod_{t=t_i+1}^{t_i+T_i} \Pr\left(y_{it}|y_{it-1}, X_i, S_i\right) \Pr\left(y_{it_i}|X_i, S_i\right),\end{aligned} \tag{8}$$

where $\Pr\left(y_{it}|y_{it-1}, X_i, S_i, \eta_i\right)$ is given by the model in equations (6) and (7). Therefore, we have

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

If we decide to make a distributional assumption about the conditional density of the first observation $\Pr\left(y_{it_i}|X_i, S_i, \eta_i\right)$ we can write the probability in (8) as

$$\Pr\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i\right) = \int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i} \Pr\left(y_{it}|y_{it-1}, X_i, S_i, \eta_i\right) \Pr\left(y_{it_i}|X_i, S_i, \eta_i\right) h(\eta_i|X_i, S_i) d\eta_i \tag{9}$$

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

$$\begin{aligned}\Pr\left(y_{it_i} = 1|X_i, S_i, \eta_i\right) &= \Pr\left(y_{it} = 1|X_{it}, S_i, \eta_i, s_{it-1} = 0, s_{it} = 1\right) \\ &= \Phi\left(\delta_{0S_i} + X'_{it_i}\delta_{S_i} + \mu_{S_i}\eta_i\right),\end{aligned} \tag{10}$$

where we have different distributions for each value of S_i. If, instead, we allow only for correlation between t_i and η_i, equation (10) will be different for each t_i.

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

$$\eta_i|X_i, S_i \sim N\left(\overline{X}_i'\beta_{\eta S_i}, \sigma^2_{\eta S_i}\right), \tag{11}$$

where $\overline{X}_i$ contains the within-means of the time-varying explanatory variables. Notice that (11) allows for correlation between the sample selection process, S_i, and the permanent unobserved heterogeneity η_i. If we assume that S_i is independent of η_i, then, $h(\eta_i|X_i, S_i) = h(\eta_i|X_i)$ and

$$\eta_i|X_i, S_i \sim N\left(\overline{X}_i'\beta_\eta, \sigma^2_\eta\right) \tag{12}$$

Alternatively, we could assume that $\eta_i|X_i, S_i$ depends only on t_i but not on the rest of S_i, that is, but not on the duration of the sub-panel.

If we decide to consider the distribution conditional on the initial period observation, we can write the probability in (8) as

$$\Pr\left(s_{i1}y_{i1}, \ldots, s_{iT}y_{iT}|X_i, S_i\right) = \left[\int_{\eta_i} \prod_{t=t_i+1}^{t_i+T_i} \Pr\left(y_{it}|y_{it-1}, X_i, S_i, \eta_i\right) h(\eta_i|y_{it_i}, X_i, S_i) d\eta_i\right] \Pr\left(y_{it_i}|X_i, S_i\right) \tag{13}$$

To solve the initial conditions problem in this case we can follow Wooldridge (2005) and specify an approximation for the density of $\eta_i|y_{it_i}, X_i, S_i$ in (13), discarding $\Pr\left(y_{it_i}|X_i, S_i\right)$ since that term is outside the integral. Continuing with the Normal case, we have

$$\eta_i|y_{it_i}, X_i, S_i \sim N\left(\pi_{0S_i} + \pi_{1S_i}y_{it_i} + \overline{X}_i'\pi_{2S_i}, \sigma^2_{\eta S_i}\right) \tag{14}$$

or

$$\eta_i|y_{it_i}, X_i, S_i \sim N\left(\pi_{0t_i} + \pi_{1t_i}y_{it_i} + \overline{X}_i'\pi_{2t_i}, \sigma^2_{\eta t_i}\right) \tag{15}$$

depending on whether we allow for correlation between S_i and η_i, or only between the moment at which we first observe each individual, t_i, and η_i.[7] As previously noticed, even if we assume that the sample selection process S_i is independent of η_i,

[7] Note that here $\overline{X}_i = \frac{1}{T_i - 1}\sum_{t=t_i+1}^{t_i+T_i} x_{it}$ for the reasons given in Rabe-Hesketh and Skrondal (2013).

$h(\eta_i|y_{it_i}, X_i, S_i)$ will be different for each t_i, i.e. it will be as in (15), 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 or rule for all units and t_i).

Previous models can be estimated by Maximum Likelihood (ML). The contribution to the likelihood function for individual i in model (9) is given by

$$L_i = \int_{\eta_i} \Phi\left(\delta_{0S_i} + X'_{it_i}\delta_{S_i} + \mu_{S_i}\eta_i\right)(2y_{it_i} - 1) \tag{16}$$
$$\left\{\prod_{t=t_i+1}^{t_i+T_i} \Phi\left[(\alpha y_{it-1} + \beta_0 + X'_{it}\beta + \eta_i)(2y_{it} - 1)\right]\right\} h(\eta_i|X_i, S_i)d\eta_i,$$

where $h(\eta_i|X_i, S_i)$ is the distribution in (11) or in (12) or any other distribution of $\eta_i|X_i, S_i$ like a discrete finite distribution. In model (13) the contribution to the likelihood function for individual i is given by

$$L_i = \int \prod_{t=t_i+1}^{t_i+T_i} \Phi\left[\left(\alpha y_{it-1} + X'_{it}\beta + \pi_{0S_i} + \pi_{1S_i}y_{it_i} + \overline{X}'_i\pi_{2S_i} + a\right)(2y_{it} - 1)\right] \frac{1}{\sigma_{\eta S_i}}\phi\left(\frac{a}{\sigma_{\eta S_i}}\right) da \tag{17}$$

The log-likelihood function is $\mathcal{L} = \sum_{i=1}^{N} \log L_i$ and it will be maximized with respect to $\left(\alpha, \beta', \{\pi_{0j}\}_{j=1}^{J}, \{\pi_{1j}\}_{j=1}^{J}, \{\pi_{2j}\}_{j=1}^{J}, \{\sigma_{\eta j}\}_{j=1}^{J}\right)$.

For balanced panels, it is well known since Wooldridge (2005) that modelling conditional on the first observation of the dependent variable plus the normality assumption for $\eta_i|y_{i1}, X_i$ produces a very simple estimation method that can be implemented with standard random-effects probit software.[8] Also, for the model that follows the Heckman's solution to the initial conditions problem, Arulampalam and Stewart (2009) propose an implementation procedure using the 'gllamm' command in Stata. However, in the unbalanced case maximizing the likelihood in (16) or (17) is cumbersome and cannot be done using the standard built-in commands in econometric software.[9]

Simpler and direct implementation: MD estimation The computational problems with the maximization of the log likelihood $\mathcal{L} = \sum_{i=1}^{N} \log L_i$ come from having

[8] See Stewart (2007) for an application where this is estimated using the Stata command for standard random-effects probit models, 'xtprobit'.

[9] Altough 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 Appendix for details.

parameters that are specific to each sub-panel.[10] The estimation of the model for each sub-panel separately takes us back to the same situation we face when having a balanced panel. To estimate the model in each sub-panel we can use standard software like the following commands in Stata: the 'xtprobit' command when estimating (13), and the 'gllamm' and 'gsem' commands when estimating (9). Once we have estimated the model for each sub-panel, we compute the weighted average of the estimates of the parameters that are common: α and β'. See the Appendix for details on how all this can be done.

A simpler distributional assumption Notice that if in (13), instead of making the assumption (14) or (15), we impose that the variance of the distribution of $\eta_i|y_{it_i}, X_i, S_i$ is constant across different sub-panels, that is

$$\eta_i|y_{it_i}, X_i, S_i \sim N\left(\pi_{0S_i} + \pi_{1S_i}y_{it_i} + \overline{X}_i'\pi_{2S_i}, \sigma_\eta^2\right), \tag{18}$$

or

$$\eta_i|y_{it_i}, X_i, S_i \sim N\left(\pi_{0t_i} + \pi_{1t_i}y_{it_i} + \overline{X}_i'\pi_{2t_i}, \sigma_\eta^2\right), \tag{19}$$

the estimation by ML becomes much easier since it can be done by using the simple and fast "xtprobit" command in Stata as we show in the Appendix.

Practical problems when taking the balanced sub-sample: At this point it is worth mentioning the potential problems of assuming Normal distributions when making the panel balanced to deal with the initial conditions problem. We have seen in the previous section that the only theoretical problem of this approach is that it disregards information (sometimes too much information so it is not possible to implement it). But from a practical point of view, if there is correlation between η_i and S_i and the distribution of $\eta_i|y_{i\max t_i}, X_i, S_i$ is Normal for each sub-panel, making the panel balanced and assuming that $\eta_i|y_{i\max t_i}, X_i$ follows a normal distribution –which would allow to use the simple practical solution by Wooldridge (2005) explained in Section 5.1– is incorrect: $\eta_i|y_{i\max t_i}, X_i$ does not follow a normal distribution in that

[10] By sub-panel we mean the part of the panel formed by units that have the same S_i. This means that in each unbalanced panel we have J subpanels.

case.[11] This also poses a problem for using the comparison between the estimates taking the balanced sub-sample with the estimates ignoring the unbalancedness to decide whether or not the unbalancedness is ignorable, as done in some applied papers. If normality about the distribution of η_i is incorrectly assumed in both cases, these two estimators will tend to produce similarly biased estimates. Therefore the comparison between them can lead to incorrectly conclude that the unbalancedness is ignorable.

3.2 Implementation when the unbalancedness is independent of the individual effect

If we specify the likelihood based on expression (9), under Assumption 2 there is a simplification in terms of computation because there is only one common distribution of η_i for all the sub-panels. It is the distribution in (12). This makes feasible obtaining the MLE from (16) using the 'gllamm' and 'gsem' commands in Stata. See the Appendix for details.

In contrast with that, if we use the likelihood based on (13), Assumption 2 does not lead to a conditional distribution of η_i that is common to all sub-panels. As can be seen in (15), $\eta_i|y_{it_i}, X_i, S_i$ still depends on when each sub-panel starts even under independence of the unbalancedness from η_i. This, as said, cannot be done using the standard built-in commands in econometric software.

3.3 Summary of estimators and notation

In this section we list all the different estimators that could be used in practice when having an unbalanced panel and present the notation that will be used in next section. The different estimators arise mostly from the assumptions we are willing to make about the relation between the unbalancedness and the permanent unobserved heterogeneity." H" and "W" denote that we use the Heckman's or the Wooldridge's approach to address the initial conditions problem, respectively.

[11] Of course, balancing the panel will work if the distribution of $\eta_i|y_{i\max t_i}, X_i$ assumed were the correct one: $h(\eta_i|y_{i\max t_i}, X_i) = \int_{S_i} h(\eta_i|y_{i\max t_i}, X_i, S_i) dG_{S_i}$, which, when $h(\eta_i|y_{i\max t_i}, X_i, S_i)$ is the normal density, is a mixture of normals with as many components as subpanels. This would be a difficult, though not unfeasible, model to estimate in practice. Certantly it is much more difficult to implement than the case that assumes normality.

A1H and A1W: Standard ML estimators for balanced panels using the subset of observations constituting a balanced panel.

A2H and A2W: ML estimators with unbalanced panels that allow for correlation between the unbalancedness (S) and η. They come from the likelihoods in (16) and (17).

A2H_MD and A2W_MD: The same as A2H and A2W, but estimating by Minimum Distance. See subsection 3.1.

A2bW: The same as A2W, but with the simpler assumption on the conditional distribution of η indicated in (18). This makes estimation by ML much simpler in practice than in A2W.

A3H and A3W: ML estimators with unbalanced panels that allow for correlation between the unbalancedness and η_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 η_i. For A3W this also corresponds with the case in which we assume that the unbalancedness is independent of η_i. See equation (15) and the comments that follow that equation, and comments in subsection 3.2.

A3H_MD and A3W_MD: The same as A3H and A3W, but estimating by Minimum Distance.

A3bW: It is like A3W, but with the simpler assumption on the conditional distribution of η indicated in (19). This makes estimation by ML much simpler in practice than in A3W.

A4H: ML estimator when assuming independence between the unbalancedness and η, that is when assuming that $\eta_i|X_i, S_i$ follows the distribution in (12). As commented in subsection 3.2, this simplifies the implementation of the ML estimator.

4 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.

4.1 Data Generating Process and unbalancedness

The baseline specification is:

$$y_{it} = 1\{\alpha y_{it-1} + \eta_i + \varepsilon_{it} \geq 0\} \quad t = 1, ..., T; i = 1, ..., N \tag{20}$$

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

$$\eta_i \underset{iid}{\sim} N(\mu_\eta, \sigma_\eta^2) \tag{21}$$

$$y_{i0} = 1\{\pi_0 + \pi_1 \eta_i + v_{i0} \geq 0\}, \quad v_{i0} \underset{iid}{\sim} N(0, 1), \tag{22}$$

where $\alpha = 0.75$, $N = 500$, $\mu_\eta = 0$, $\sigma_\eta^2 = 1$, $\pi_0 = -1.25$, and $\pi_1 = 0$, so the initial condition of the process is exogenous and it is not drawn from the steady state.[12]

The unbalancedness is randomly generated, independently of everything else, and the sub-panels vary in both when individuals enter and when they leave the sample. The degree of unbalancedness in the sample is governed by a parameter J, which indicates the number of sub-panels. The set of individuals that are observed the same periods form a sub-panel. $J = 0$ indicates that the panel is balanced. If $J = 2$, 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 observer from 2 to T. If $J = 4$, a 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 1 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. We also 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

[12] In the simulations we consider the model without other covariates because this model already contains all the problems we want to address and it reduces computational time. Actually, a model with strictly exogenous covariates may have a better performance and then this would be like a worse case scenario. In any case, both our theoretical study and our discussions on how to implement the estimators we propose include other covariates.

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.[13]

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. Here J can take both even and odd values. Table 2 contains an example of the unbalanced structure in this case. 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.

2. Different values of N. In particular we have considered $N = 200$ and $N = 1000$, in addition to the baseline case $N = 500$.

3. Different values of α, to evaluate the sensitivity to different degrees of persistence. We have considered $\alpha = 0.5$ and $\alpha = 1$.

4. Initial condition correlated with η_i: $\pi_1 = 0.5$.

5. Both initial condition and unbalancedness correlated with η_i: $\pi_1 = 0.5$ and η_i is generated as follows:
$$\eta_i | S_i \underset{iid}{\sim} N(\mu_{\eta S}, \sigma^2_{\eta S}),$$
where $\mu_{\eta S}$ and $\sigma^2_{\eta S}$ are different for each sub-panel, so that there is correlation between the values of η_i and being observed (S_i). $\mu_{\eta S}$ and $\sigma^2_{\eta S}$ are generated randomly but in a way such that $E_S(\sigma^2_{\eta S}) = \sigma^2_\eta = 1$, $E_S(\mu_{\eta S}) = \mu_\eta = 0$, and $\mu_{\eta S}$ is increasing in S. Thus, the value of η_i is more likely to be larger the larger the value of S, i.e. for the last sub-panels. In the left-side unbalancedness it means that individuals with higher η_i tend not to be observed the first periods. Notice that since η_i follows a normal distribution for each sub-panel, its aggregate

[13] When the time lenght is long, fixed effects approaches may be prefereable. For example, simulations in Carro (2007) show cases where a modified MLE fixed effects estimator performs well with 8 periods.

distribution over the entire set of individuals is not normal, but a mixture of normals.

4.2 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 A2H and A2W 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 in which other estimators, including those that are known to be incorrect, are used too.

Table 3 and Table 4 show the results on the finite-sample performance of several of the estimators discussed in this paper in our baseline specification. Under this setting, and irrespective of the unbalancedness, it is known that all the proposed approaches that do not ignore the unbalancedness give consistent estimates. We observe that all the five approaches considered here provide estimated values of the parameter α very close to its true value. However, there exists some other relevant points that are worth noting. Solution approaches that use Wooldridge's proposal to address the initial condition problem and those that use Heckman's proposal have similar performance in terms of Root Mean Square Error (RMSE), independently of T, J, and the type of unbalancedness. This similar performance is maintained even when Heckman's proposal allow to have an estimator that specifically uses the independence of the unbalancedness and η_i, A4H, whereas in the Wooldridge's proposal the estimator is the same as in the case of correlation between t_i and η_i. As a consequence of that, and given that this paper is not about comparing this two proposals to deal with the initial conditions problem, we will only use Wooldridge's proposal in the rest of simulations because it tends to be faster to compute.

More relevant to our aim, the solutions that employ standard methods after balancing the sample, namely A1H and A1W, have two important drawbacks compared to any of the other approaches. First, those solutions 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, those solutions imply an important loss of efficiency in terms of RMSE when they can be employed compared to the approaches proposed in this paper. Our proposals always dominates the usual solutions in terms of RMSE and they can have as less as one half of its RMSE. This is true both if we consider double unbalancedness (Table 3) or only left-side unbalancedness (Table 4) and, again, losses are remarkable even for moderate unbalancedness. For instance, Table 3 shows that for $T = 8$ and $J = 4$ the RMSE of A1H and A1W is around 0.17 compared with around 0.09 for A3W, A3Wb and A4H. Table 4 displays a similar picture: for $T = 8$ and $J = 5$, for instance, the RMSE of A1H and A1W is around 0.16, compared with around 0.09 for A3W, A3Wb and A4H.

In Table 5 we have the same baseline specification as before but with a smaller ($N = 200$) and a larger ($N = 1000$) sample size in Panels A and B, respectively. Although the RMSE of all the solutions is reduced (increased) when the sample size increases (decreases), the relative loss of efficiency of the approach that takes only the observations that constitute a balanced sample, A1W, remains as with $N = 500$. Also, the performance of A1W quickly deteriorates even with moderate unbalancedness. For instance, Table 5 Panel B shows small differences for $T = 8$ and $J = 2$ in RMSE (0.06 of A1W compared with 0.05 of A3W and A3Wb), but if unbalancedness is just a bit more intense, $J = 4$, the RMSE of A1W almost doubles to 0.11, whereas the RMSE of A3W and A3Wb barely changes. All this is in addition to the fact that the approach cannot be employed for many of the unbalanced structures. These results remain unchanged when in Table 6 and Table 7 we consider lower and higher state dependence, $\alpha = 0.50$ and $\alpha = 1$, respectively. Moreover, the RMSE does not seem to diminish with smaller state dependence.

In Table 8 the initial condition is correlated with η. As can be seen, this does not change the finite sample performance of the estimators. The endogeneity of the initial condition does not play a role here because the problem with unbalanced panels, when the unbalancedness is exogenous, comes from the dynamics of the model and not from the initial condition.

Finally, Table 9 presents a situation in which not only the initial condition but also the unbalancedness is correlated with η_i, as explained in point 5 in section 4.1. The results for this simulation show a similar pattern than in previous Tables regarding the comparison between the estimator that balance the sample (A1W) and our preferred solution (A2W_MD): A1W cannot be employed in many cases, and it implies an important loss of efficiency in terms of RMSE. However, the performance of A1W is now worse than in previous DGPs, reflecting the extra difficulty of having to approximate a potentially complicated relation between η and S assuming that there is a common distribution of η that does not change across sub-panels in a given period.

The estimator that accounts for the unbalancedness but imposing a common variance on the distribution of $\eta_i|S_i$ (A2Wb) performs worse than A2W_MD both in terms of bias and RMSE. This is not surprising, since the assumption of common variance across sub-panels is much less reasonable when there is correlation between η_i and the unbalancedness. As opposed to that, in previous DGPs the estimator that imposes common variance slightly outperforms the MD estimator because it makes an efficient use of all the information and, when η and S are independent, it seems to approximate reasonably well the true distribution of η.

Last column in Table 9 presents the ML estimates of the most general model that does not impose a common variance of the distribution of $\eta_i|S_i$ (A2W). Estimating this model is computationally cumbersome so we just report simulations up to $T = 8$ and $J = 8$. As opposed to the MD estimator, in this case there is no a potential problem of lack of variability in certain sub-panels. Table 10 reports the percentage of simulations that achieved convergence for the MD estimator. We see that the percentage of failures is below 10% up to simulations with a very high degree of unbalancedness, and even in this cases it does not seem to perform worse than the ML estimator, although this result could be due to the fact that a different maximization routine is used in both cases.

Average Marginal Effects So far we have discussed only how well our proposed approaches perform to estimate the parameter α. However, practitioners estimating non-linear models are ultimately interested in marginal effects. Therefore, we consider the finite-sample performance of the estimated Average Marginal Effect (AME) in the model specification of Table 8 to see if the conclusions reached for the estimation of the parameters are valid for the marginal effects. Since the true AME (slightly) varies with the sample drawn in each Monte Carlo simulation, Table 11 reports the true expected AME along with the estimated AME and the RMSE of the estimator. As can be seen in Table 11, the same conclusions we have reached with respect to α apply to the estimation of the marginal effects too.

5 Conclusions

The main results that emerged from our analysis are the following:

- First, we show that the approach that disregards information by balancing the sample presents important efficiency losses in comparison with the different versions of the approaches proposed in this paper that exploit the unbalancedness structure.

- Second, the problem is specially severe when the unbalanced process is correlated with the individual effect. Approximating the distribution of the individual effect conditional on the covariates and the unbalancedness can be very difficult if doing it for the entire sample, which could even affect the consistency if the approximation is poor. In contrast with that the methods we propose accommodate very easily the fact that the distributions of the individual effect may be totally different for each subpanel.

- Third, the unbalancedness and the dynamics of the model can produce an initial condition problem even if the initial condition of the process is exogenous.

- Finally, the approaches proposed in this paper can be implemented relatively easily using standard software and perform well, including the simple Minimum Distance estimation.

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

Payments from Casino Profits", *American Economic Journal: Applied Economics* 2:1, 86–115.

[2] Arellano, M.(2003): "Discrete Choice with Panel Data" *Investigaciones Económicas,* vol. XXVII (3), 423-458.

[3] 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.

[4] 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.

[5] 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.

[6] Bernard, A. B. and J. B. Jensen (2004): "Why Some Firms Export", *The Review of Economics and Statistic*, 86(2), 561-569.

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

[8] Campa, J. M. (2004): "Exchange rates and trade: How important is hysteresis in trade?", *European Economic Review*, 48, 527-548.

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

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

[11] Chamberlain, G., (1980): "Analysis of Covariance with Qualitative Data", *The Review of Economic Studies*, 47 (1), 225-238.

[12] 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

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

[14] 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.

[15] Hyslop, D. R. (1999): "State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women", *Econometrica*, 67, 1255-1294.

[16] 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.

[17] Stewart, M. B. (2007): "The interrelated dynamics of unemployment and low-wage employment", *Journal of Applied Econometrics,* 22, 511- 531.

[18] 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.

[19] Wooldridge, J. M. (2010): "Correlated Random Effects Models with Unbalanced Panels", mimeo.

6 Appendix

The different models can be estimated using standard software as, for instance, Stata. In what follows, we present the Stata codes used in each case, as well as the main problems we have found to implement them. Some models can be easily estimated using the command "xtprobit", while for others we have used the "gsem" and "gllamm" commands. Finally, in some cases it is necessary to write a specific likelihood maximizing program.

6.1 Using the balanced panel (A1H, A1W)

For the estimator A1W it is possible to obtain ML estimates using the "xtprobit" command. The likelihood function to be maximized for A1W is

$$L_i = \int \prod_{t=\max t_i+1}^{\min(t_i+T_i)} \Phi\left[\left(\alpha y_{it-1} + X_{it}'\beta + \pi_0 + \pi_1 y_{i\max t_i} + \overline{X}_i'\pi_2 + a\right)(2y_{it}-1)\right]\frac{1}{\sigma_\eta}\phi\left(\frac{a}{\sigma_\eta}\right)da,$$

Thus, if the variable *id* is the individual indicator, and $y0$ is the initial condition, this model can be estimated, after selecting the balanced sub-sample (A1W), using the following Stata code:

```
qbys id: gen y0 = y[1]
xtprobit y l.y y0 x m_x, re iter(#) intpoints(#)
```

where $l.y$ is the first lag of y, while x and m_x are vectors of the explanatory variables and their means, respectively.[14] The symbol "#" indicates the number of iterations (iter) and the number of quadrature points (intpoints). In our simulations these have been set to 50 and 24, respectively.

For the estimator A1H one can use the "gllamm" command, as in Arulampalam and Stewart (2009). Nonetheless, we have found that, in the absence of convergence problems, the "gsem" command reaches the optimum faster. Therefore, in our Monte Carlo study we have tried first the "gsem" command using Stata V13 and if convergence is not achieved after a certain number of iterations (10 iterations), we have switched to "gllamm". The likelihood function to be maximized is

$$L_i = \int_{\eta_i} \Phi\left(\delta_0 + X_{i\max t_i}'\delta + \mu\eta_i\right)(2y_{i\max t_i}-1)$$
$$\left\{\prod_{t=\max t_i+1}^{\min(t_i+T_i)} \Phi\left[\left(\alpha y_{it-1} + \beta_0 + X_{it}'\beta + \eta_i\right)(2y_{it}-1)\right]\right\} h(\eta_i|X_i)d\eta_i,$$

for A1H.

The syntax of the "gsem" command requires to specify two equations: one for the main dynamic equation and another for the initial condition. Moreover, we have to set a latent variable, the individual effect, common to both equations. Thus, if the variable *time* indicates the period number each observation corresponds to (time=1,2,...), we use the following Stata code:

```
gen yy1 = y if time>1
gen yy0 = y if time==1
gen ly = l.y
gen xx1 = x  if time>1
gen xx0 = x0 if time==1
```

[14] Note that here $\overline{X}_i = \frac{1}{T_i-1}\sum_{t=t_i+1}^{t_i+T_i} x_{it}$ for the reasons given in Rabe-Hesketh and Skrondal (2013).

```
gsem (yy1 <- ly xx1 I[id], probit)              ///
  (yy0 <- xx0 I[id], probit), intp(#) iter(10)
```

On the other hand, the "gllamm" command can be used to estimate this model, following the notation proposed by Arulampalam and Stewart (2009). Specifically, they suggest to combine the equations for the initial condition and for the rest of observations as follows. Thus, taking into account the assumption

$$\eta_i|X_i, S_i \sim N\left(\overline{X}'_i\beta_\eta, \sigma^2_\eta\right),$$

we can write

$$\Pr[y_{it} = 1|y_{it-1}, X_i, S_i, \eta_i] = \\ \Phi\left[(1-d_i^0)(\alpha y_{it-1} + X'_{it}\beta + \beta_0 + \overline{X}'_i\beta_\eta + b) + d_i^0(X'_{it_i}\delta + \delta_0 + \mu(\overline{X}'_i\beta_\eta + b))\right]$$

$$= \Phi\left[(1-d_i^0)\alpha y_{it-1} + (1-d_i^0)X'_{it}\beta + (1-d_i^0)\overline{X}'_i\beta_\eta + \beta_0 + d_i^0 X'_{it_i}\delta + d_i^0(\delta_0 - \beta_0) + \overline{X}'_i\beta_\eta + b + d_i^0(\overline{X}'_i\beta_\eta + b)(\mu - 1)\right],$$

where d_i^0 is a dummy variable equal to 1 for the first observation, and 0 otherwise, and $b \sim N\left(0, \sigma^2_\eta\right)$.

Before running gllamm we have to define one equation to specify the variables that multiply the random effect. The syntax is as follows:

```
gen d0=(time==1)
gen const = 1
eq etai: const d0
qbys id: gen Ly = y[_n-1]
replace Ly=0 if time==1
gllamm y Ly x1 d0 m_x x0,i(id) nrf(1) eqs(etai) nip(#) fam(binom) link(probit)   ///
  adapt trace iterate(#)
```

where the $nrf(1)$ option indicates that there is one random effect, and the equation "etai" specifies the variables associated to it: a constant variable, *const*, and the variable $d0$. The variable Ly equals to the first lag of the dependent variable y but taking the value 0 for the first observation, since the regressor is the interaction between y_{it-1} and $(1 - d_i^0)$.

6.2 Allowing for correlation between the unbalancedness and the individual effect (A2H, A2W, and A3H, A3W)

In these cases performing the ML estimates is computationally cumbersome. Although in principle these could be obtained using the "gsem" and the "gllamm" commands, this is so time consuming that makes infeasible in practice to perform a Monte Carlo study. Therefore, to obtain the ML estimates we have written the expressions for the likelihood functions in a specific likelihood maximizing program available upon request. Nonetheless, we provide an explanation on how to estimate these models with "gsem" and "gllamm" because for one estimate they could be feasible to implement.

For the model A2H, the likelihood function to be maximized is

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

The generalization of the "gsem" command to the unbalanced case with correlation basically consists on specifying one initial condition equation different for each sub-panel, while the dynamic equation for the rest of observations is common to all the individuals. For instance, suppose that we have two sub-panels and that JJ is a variable that indicates the sub-panel to which the individual belongs to. In our example, JJ can take the values 1 or 2. Before calling the gsem command we have to generate the initial conditions for each sub-panel, $y0\_1$ and $y0\_2$:

```
gen y0_1=y if time==1 & JJ==1
gen y0_2=y if time==1 & JJ==2
```

Equally, we have to generate $x0\_1$ and $x0\_2$:

```
gen x0_1=x if time==1 & JJ==1
gen x0_2=x if time==1 & JJ==2
```

Then, the gsem command is specified as follows:

```
xi:gsem(y1<-l.y x1 I[id], probit)    ///
  (y0_1<-J[id] x0_1, probit)         ///
  (y0_2<-K[id] x0_2, probit)
```

Notice that a different latent variable should be included in each equation to ensure that the unobserved effect follows a different distribution in each sub-panel.

Unfortunately, the gsem command has an important drawback for our purposes, because as the number of sub-panels increases the number of equations to include in the command also increases. Therefore, the estimation procedure followed by this command becomes increasingly complex and it often fails to achieve convergence.

For similar reasons, the implementation of the "gllamm" command for this model is also difficult. Following with the previous example of two sub-panels, using the gllamm command requires to state that there are two random effects, one for each sub-panel, each of them with a different constant, $const\_1$ and $const\_2$, and different dummy variables for each initial condition, $d0\_1$ and $d0\_2$. Therefore, we need to generate:

```
gen const_1=(JJ==1)
gen const_2=(JJ==2)
gen d0_1=(time==1 & JJ==1)
gen d0_2=(time==1 & JJ==2)
```

The Stata code is as follows:

```
eq etai_1:const_1 d0_1
eq etai_2:const_2 d0_2
gllamm y Ly x1 d0_1 d0_2 x0_1 x0_2 mx_1 mx_2, i(id) nrf(2) eqs(etai_1 etai_2)  ///
  nip(#) fam(binom) link(probit)adapt trace iterate(#)nocorrel,
```

where the *nocorrel* option specifies zero correlation between the two random effects.

For the model A2W, the likelihood function is the following:

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

Notice that this model can not be estimated with the xtprobit command. We have to use the gllamm command. It requires to specify that there are two conditional distributions for the random effects, one for each sub-panel and to include two different constants and initial conditions in the main equation, as follows:

```
eq etai_1:const_1
eq etai_2:const_2
gllamm y l.y x const_1 const_2 y0_1 y0_2 mx_1 mx_2, i(id) nrf(2) eqs(etai_1 etai_2)     ///
  nip(#) fam(binom) link(probit)adapt trace iterate(#) nocorrel noconst
```

where mx_1 and mx_2 are the vector of means of the explanatory variables interacted with $const\_1$ and $const\_2$, respectively.

As previously pointed out, as the number of sub-panels increases the implementation of these commands becomes infeasible.

Notice that if the unbalancedness is defined only in terms of different initial periods for each individual, and not also on the different duration of the sub-panels, the definition of the indicator JJ changes. That is, for estimating models A3H and A3W JJ takes different values depending only on the different initial periods available in the sample, while for estimating models A2H and A2W JJ takes different values depending on the combination of different initial and last periods.

Minimum Distance estimation An easy alternative estimation procedure is the Minimum Distance. MD estimation involves the estimation of the coefficients for each sub-panel in a first stage. In the second stage, the estimator is derived by minimizing the weighted difference between the coefficients obtained in the first stage. In Stata, the estimation of the model A2W_MD can be easily performed by using the "xtprobit" command for each sub-panel. The Stata code for the case of two sub-panels is the following:

```
forvalues Z=1/2 {
  xi:xtprobit y l.y y0 x m_x,re iter(#) intpoints(#), if JJ==‘Z’
  matB=e(b)
  matV=e(V)
  scalar accum1 = accum1+B[1,1]/V[1,1]
  scalar accum2 = accum2+1/V[1,1]
}
```

Notice that we compute the optimum MD estimator. Then, the MD estimates are obtained as

```
return scalar coef1 = accum1/accum2
return scalar SE1 = sqrt(1/accum2)
```

Equally, for the MD estimation of model A2H we can apply the "gsem" or the "gllamm" commands inside the loop for each sub-panel, as previously explained.

It is important to note that, although computationally feasible, the practical problem with the MD estimator is the potential lack of variability in a specific sub-panel. This problem is less likely to appear when using the whole sample as the ML does.

Constant variance of η_i Finally, the simplifying assumption that the variance of the conditional distribution of η_i is constant across sub-panels, makes the implementation of the ML of previous model (A2bW) feasible. That is, if we assume that

$$\eta_i | y_{it_i}, X_i, S_i \sim N\left(\pi_{0S_i} + \pi_{1S_i} y_{it_i} + \overline{X}'_i \pi_{2S_i}, \sigma^2_\eta\right),$$

ML estimates can be easily obtained by using the "xtprobit" command. For the two sub-panels case, we have to generate two different constants, $const\_1$ and $const\_2$, and two different initial conditions for each sub-panel, $y0\_1$ and $y0\_2$. The Stata code used is the following:

```
xtprobit y l.y x const_1 const_2 y0_1 y0_2 mx_1 mx_2, re iter(#) intpoints(#) noconst
```

6.3 Using unbalanced panels and assuming independence between the unbalancedness and the individual effect

In this case the estimator of the model that specifies the density of the unobserved effect conditional on the first observation to deal with the initial conditions problem when we assume independence is the same as the one that allows for correlation between t_i and η (A3W). Therefore, in this subsection we just focus on the model that specifies the density of the unobserved effect conditional on the first observation to deal with the initial conditions problem (**A4H**), for which ML estimates can be obtained by using the "gsem" or the "gllamm" commands. Notice that the difference with respect to the correlated case is that there is only one common distribution for the unobserved effects in all sub-panels. The likelihood function to be maximized is

$$L_i = \int_{\eta_i} \Phi\left(\delta_{0S_i} + X'_{it_i}\delta_{S_i} + \mu_{S_i}\eta_i\right)(2y_{it_i}-1)\left\{\prod_{t=t_i+1}^{t_i+T_i} \Phi\left[(\alpha y_{it-1} + \beta_0 + X'_{it}\beta + \eta_i)(2y_{it}-1)\right]\right\} h(\eta_i|X_i)d\eta_i \tag{23}$$

The Stata code to implement the "gsem" command for case in which we have two different sub-panels is

```
gsem(y1<-l.y x1 I[id],probit)   ///
  (y0_1<-I[id] x0_1, probit)    ///
  (y0_2<-I[id] x0_2, probit)
```

where the difference with respect to the correlated case is that the same latent variable, $I[id]$, is included in all the equations.

As in previous cases, the Stata code to implement the A4H estimator using the "gllamm" command can be easily understood using the Arulampalam and Stewart (2009) notation:

$$\Pr[y_{it}=1|y_{it-1},X_i,S_i,\eta_i] = \Phi\Big[(1-d^0_{S_i})\alpha y_{it-1} + (1-d^0_{S_i})X'_{it}\beta + (1-d^0_{S_i})\overline{X}'_i\beta_\eta + \beta_0 + d^0_{S_i}X'_{it_i}\delta$$
$$+ d^0_{S_i}(\delta_0-\beta_0) + \overline{X}'_i\beta_\eta + b + d^0_{S_i}(\overline{X}'_i\beta_\eta + b)(\mu-1)\Big]$$

Thus, it only requires to specify one equation for the random effect, with one constant, *const*, and two different dummy variables for the initial conditions, $d0\_1$ and $d0\_2$:

```
eq etai:const d0_1 d0_2
gllamm y Ly x1 d0_1 d0_2 x0_1 x0_2 mx_1 mx_2, i(id) nrf(1) eqs(etai)   ///
  nip(#) fam(binom) link(probit)adapt trace iterate(#)
```

7 Tables

Table 1: Example of double unbalancedness

	$t =$	**1**	**2**	**3**	**4**	**5**	**6**
$\mathbf{J = 0}$	For N units $s_i =$	1	1	1	1	1	1
$\mathbf{J = 2}$	For $N/2$ units $s_i =$	1	1	1	1	1	0
	For $N/2$ units $s_i =$	0	1	1	1	1	1
$\mathbf{J = 4}$	For $N/4$ units $s_i =$	1	1	1	1	1	0
	For $N/4$ units $s_i =$	1	1	1	1	0	0
	For $N/4$ units $s_i =$	0	1	1	1	1	1
	For $N/4$ units $s_i =$	0	0	1	1	1	1
$\mathbf{J = 6}$	For $N/6$ units $s_i =$	1	1	1	1	1	0
	For $N/6$ units $s_i =$	1	1	1	1	0	0
	For $N/6$ units $s_i =$	1	1	1	0	0	0
	For $N/6$ units $s_i =$	0	1	1	1	1	1
	For $N/6$ units $s_i =$	0	0	1	1	1	1
	For $N/6$ units $s_i =$	0	0	0	1	1	1

Table 2: Example of left-side unbalancedness

	$t =$	**1**	**2**	**3**	**4**	**5**	**6**	7
$\mathbf{J = 0}$	For N units $s_i =$	1	1	1	1	1	1	1
$\mathbf{J = 2}$	For $N/2$ units $s_i =$	0	1	1	1	1	1	1
	For $N/2$ units $s_i =$	0	0	1	1	1	1	1
$\mathbf{J = 3}$	For $N/3$ units $s_i =$	0	1	1	1	1	1	1
	For $N/3$ units $s_i =$	0	0	1	1	1	1	1
	For $N/3$ units $s_i =$	0	0	0	1	1	1	1
$\mathbf{J = 4}$	For $N/4$ units $s_i =$	0	1	1	1	1	1	1
	For $N/4$ units $s_i =$	0	0	1	1	1	1	1
	For $N/4$ units $s_i =$	0	0	0	1	1	1	1
	For $N/4$ units $s_i =$	0	0	0	0	1	1	1

Table 3: Monte Carlo Simulation results. Baseline Specification

$\alpha = 0.75$		A1H	A1W	A3W_MD	A3bW	A4H	A1H	A1W	A3W_MD	A3bW	A4H
				mean $\widehat{\alpha}$					RMSE		
T=4	J=0	0.7532	0.7532	0.7532	0.7532	0.7532	0.1213	0.1212	0.1212	0.1212	0.1213
	J=2			0.7633	0.7370	0.7409			0.2115	0.2042	0.1848
T=6	J=0	0.7502	0.7502	0.7502	0.7502	0.7502	0.0833	0.0833	0.0833	0.0833	0.0833
	J=2	0.7538	0.7567	0.7557	0.7498	0.7530	0.1617	0.1620	0.1046	0.1034	0.1025
	J=4			0.7588	0.7467	0.7513			0.1178	0.1138	0.1100
	J=6			0.7715	0.7474	0.7543			0.1366	0.1292	0.1253
T=8	J=2	0.7500	0.7518	0.7513	0.7483	0.7505	0.0999	0.1002	0.0801	0.0796	0.0796
	J=4	0.7561	0.7616	0.7513	0.7469	0.7504	0.1665	0.1673	0.0856	0.0850	0.0850
	J=6			0.7518	0.7451	0.7493			0.0912	0.0897	0.0892
	J=8			0.7558	0.7431	0.7460			0.0984	0.0961	0.0956
	J=10			0.7594	0.7420	0.7474			0.1072	0.1048	0.1057
T=10	J=6	0.7608	0.7672	0.7505	0.7472	0.7500	0.1638	0.1652	0.0699	0.0697	0.0697
	J=8			0.7513	0.7469	0.7494			0.0739	0.0731	0.0732
	J=10			0.7526	0.7452	0.7471			0.0791	0.0777	0.0783
	J=12			0.7572	0.7457	0.7459			0.0843	0.0820	0.0853
	J=14			0.7590	0.7458	0.7530			0.0887	0.0858	0.0906
T=15	J=16			0.7475	0.7442	0.7438			0.0559	0.0557	0.0564

Note: In the baseline specification $\alpha = 0.75$, $N = 500$, $\mu_\eta = 0$, $\sigma^2_\eta = 1$, $\pi_0 = -1.25$, $\pi_1 = 0$ so the initial condition of the process is exogenous and not drawn from the steady state, and there is Double Unbalancedness at random. See equations (20) - (22) and comments that follow them for more details.

Table 4: Monte Carlo Simulation results. Baseline Specification with Left-side Unbalancedness

$\alpha = 0.75$		A1H	A1W	A3W_MD	A3bW	A4H	A1H	A1W	A3W_MD	A3bW	A4H
		mean $\widehat{\alpha}$					RMSE				
T=4	J=0	0.7532	0.7532	0.7532	0.7532	0.7532	0.1213	0.1212	0.1212	0.1212	0.1213
	J=2			0.7590	0.7375	0.7482			0.1595	0.1548	0.1482
T=6	J=0	0.7502	0.7502	0.7502	0.7502	0.7502	0.0833	0.0833	0.0833	0.0833	0.0833
	J=4			0.7667	0.7373	0.7514			0.1304	0.1249	0.1230
T=8	J=4	0.7501	0.7537	0.7520	0.7414	0.7493	0.1224	0.1229	0.0895	0.0881	0.0876
	J=5	0.7502	0.7553	0.7570	0.7394	0.7492	0.1667	0.1662	0.0979	0.0938	0.0944
	J=6			0.7590	0.7364	0.7474			0.1068	0.1036	0.1035
T=10	J=4	0.7488	0.7506	0.7502	0.7445	0.7494	0.0836	0.0836	0.0681	0.0679	0.0679
	J=5	0.7447	0.7473	0.7491	0.7423	0.7480	0.0981	0.0986	0.0719	0.0717	0.0717
	J=6	0.7435	0.7468	0.7494	0.7401	0.7458	0.1194	0.1199	0.0766	0.0760	0.0754
	J=7	0.7414	0.7467	0.7527	0.7386	0.7441	0.1653	0.1648	0.0827	0.0813	0.0814
	J=8			0.7553	0.7374	0.7480			0.0875	0.0858	0.0829
T=15	J=9	0.7520	0.7540	0.7510	0.7469	0.7497	0.0838	0.0842	0.0548	0.0547	0.0547
	J=10	0.7507	0.7533	0.7506	0.7457	0.7484	0.0955	0.0961	0.0575	0.0574	0.0570
	J=11	0.7523	0.7563	0.7519	0.7451	0.7452	0.1210	0.1219	0.0606	0.0599	0.0606
	J=12	0.7527	0.7568	0.7540	0.7444	0.7406	0.1665	0.1655	0.0629	0.0622	0.0643
	J=13			0.7557	0.7443	0.7399			0.0663	0.0654	0.0648

Note: In the baseline specification $\alpha = 0.75$, $N = 500$, $\mu_\eta = 0$, $\sigma^2_\eta = 1$, $\pi_0 = -1.25$, $\pi_1 = 0$ so the initial condition of the process is exogenous and not drawn from the steady state. Here there is only Left-side Unbalancedness at random.

Table 5: Monte Carlo Simulation results. Baseline Specification with $N = 200$ and $N = 1000$

Panel A: N=200

$\alpha = 0.75$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			mean $\widehat{\alpha}$			RMSE	
T=4	J=0	0.7525	0.7525	0.7525	0.1866	0.1866	0.1866
	J=2		0.7798	0.7360		0.3327	0.3263
T=6	J=0	0.7496	0.7496	0.7496	0.1289	0.1289	0.1289
	J=2	0.7490	0.7530	0.7451	0.2420	0.1623	0.1599
	J=4		0.7667	0.7452		0.1884	0.1822
	J=6		0.7658	0.7387		0.2151	0.2108
T=8	J=2	0.7525	0.7531	0.7497	0.1491	0.1211	0.1205
	J=4	0.7513	0.7555	0.7477	0.2621	0.1337	0.1313
	J=6		0.7645	0.7494		0.1440	0.1398
T=10	J=6	0.7534	0.7509	0.7458	0.2653	0.1116	0.1098

Panel B: N=1000

$\alpha = 0.75$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			mean $\widehat{\alpha}$			RMSE	
T=4	J=0	0.7487	0.7487	0.7487	0.0839	0.0839	0.0839
	J=2		0.7549	0.7392		0.1497	0.1454
T=6	J=0	0.7477	0.7477	0.7477	0.0573	0.0573	0.0573
	J=2	0.7519	0.7483	0.7432	0.1076	0.0707	0.0702
	J=4		0.7494	0.7411		0.0782	0.0771
	J=6		0.7573	0.7400		0.0925	0.0881
T=8	J=2	0.7517	0.7518	0.7488	0.0659	0.0533	0.0530
	J=4	0.7590	0.7517	0.7472	0.1157	0.0563	0.0560
	J=6		0.7527	0.7468		0.0607	0.0602
	J=8		0.7560	0.7465		0.0690	0.0665
	J=10		0.7586	0.7455		0.0738	0.0719
T=10	J=6	0.7619	0.7516	0.7482	0.1158	0.0489	0.0486
	J=8		0.7512	0.7472		0.0515	0.0512
	J=10		0.7517	0.7466		0.0546	0.0530
	J=12		0.7529	0.7456		0.0586	0.0574
	J=14		0.7544	0.7449		0.0632	0.0615
T=15	J=16		0.7493	0.7468		0.0413	0.0412
	J=18		0.7496	0.7465		0.0427	0.0425
	J=20		0.7506	0.7467		0.0443	0.0440
	J=22		0.7511	0.7459		0.0471	0.0465
	J=24		0.7511	0.7452		0.0497	0.0488

Note: In the baseline specification $\alpha = 0.75$, $\mu_\eta = 0$, $\sigma_\eta^2 = 1$, $\pi_0 = -1.25$, $\pi_1 = 0$ so the initial condition of the process is exogenous and not drawn from the steady state, and there is Double Unbalancedness at random.

Table 6: Monte Carlo Simulation results. Smaller state dependence: $\alpha = 0.50$

Panel A: Double Unbalancedness

$\alpha = 0.50$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			mean $\widehat{\alpha}$			RMSE	
T=4	J=0	0.5014	0.5014	0.5014	0.1199	0.1199	0.1199
	J=2		0.5107	0.4887		0.2168	0.2094
T=6	J=0	0.4964	0.4964	0.4964	0.0801	0.0801	0.0801
	J=2	0.4991	0.4999	0.4966	0.1516	0.1007	0.1000
	J=4		0.5020	0.4946		0.1150	0.1119
	J=6		0.5168	0.4959		0.1330	0.1268
T=8	J=2	0.5024	0.5017	0.5001	0.0942	0.0771	0.0767
	J=4	0.5109	0.5018	0.4994	0.1550	0.0821	0.0817
	J=6		0.5020	0.4984		0.0871	0.0862
	J=8		0.5060	0.4980		0.0962	0.0938
	J=10		0.5108	0.4964		0.1059	0.1028
T=10	J=6	0.5128	0.4998	0.4981	0.1488	0.0660	0.0658
	J=8		0.4996	0.4974		0.0701	0.0698
	J=10		0.5001	0.4959		0.0743	0.0737
	J=12		0.5037	0.4959		0.0789	0.0780
	J=14		0.5061	0.4952		0.0845	0.0824
T=15	J=16		0.4964	0.4946		0.0543	0.0543

Panel B: Left-side Unbalancedness

$\alpha = 0.50$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			mean $\widehat{\alpha}$			RMSE	
T=4	J=0	0.5014	0.5014	0.5014	0.1199	0.1199	0.1199
	J=2		0.5063	0.4844		0.1597	0.1552
T=6	J=0	0.4964	0.4964	0.4964	0.0801	0.0801	0.0801
	J=4		0.5079	0.4812		0.1199	0.1164
T=8	J=4	0.5041	0.5020	0.4939	0.1141	0.0874	0.0858
	J=5	0.5074	0.5048	0.4917	0.1542	0.0933	0.0908
	J=6		0.5097	0.4895		0.1018	0.0992
T=10	J=4	0.5003	0.5000	0.4959	0.0775	0.0626	0.0625
	J=5	0.4968	0.4992	0.4941	0.0920	0.0665	0.0665
	J=6	0.4956	0.4998	0.4926	0.1093	0.0870	0.0697
	J=7	0.4795	0.5032	0.4918	0.1517	0.0751	0.0739
	J=8		0.5060	0.4907		0.0805	0.0792
T=15	J=9	0.5006	0.4995	0.4962	0.0798	0.0534	0.0534
	J=10	0.5016	0.4987	0.4948	0.0919	0.0559	0.0561
	J=11	0.5030	0.4998	0.4941	0.1136	0.0590	0.0588
	J=12	0.5065	0.5013	0.4933	0.1539	0.0604	0.0603
	J=13		0.5035	0.4937		0.0627	0.0625

Note: As in the baseline specification, $N = 500$, $\mu_\eta = 0$, $\sigma^2_\eta = 1$, $\pi_0 = -1.25$, $\pi_1 = 0$ so the initial condition of the process is exogenous and not drawn from the steady state, and the unbalancedness is at random. However, here $\alpha = 0.50$.

Table 7: Monte Carlo Simulation results. Higher state dependence: $\alpha = 1$

Panel A: Double Unbalancedness

$\alpha = 1$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			mean $\widehat{\alpha}$			RMSE	
T=4	J=0	1.0029	1.0029	1.0029	0.1206	0.1206	0.1206
	J=2		1.0167	0.9890		0.2136	0.2065
T=6	J=0	1.0016	1.0016	1.0016	0.0856	0.0856	0.0857
	J=2	1.0112	1.0072	0.9987	0.1691	0.1096	0.1080
	J=4		1.0161	0.9952		0.1249	0.1181
	J=6		1.0220	0.9944		0.1381	0.1331
T=8	J=2	1.0030	1.0017	0.9969	0.1042	0.0829	0.0822
	J=4	1.0140	1.0010	0.9937	0.1761	0.0889	0.0879
	J=6		1.0037	0.9917		0.0981	0.0953
	J=8		1.0078	0.9895		0.1051	0.1014
	J=10		1.0106	0.9884		0.1152	0.1111
T=10	J=6	1.0184	1.0014	0.9954	0.1854	0.0730	0.0724
	J=8		1.0030	0.9947		0.0772	0.0757
	J=10		1.0060	0.9932		0.0834	0.0807
	J=12		1.0104	0.9934		0.0879	0.0852
	J=14		1.0100	0.9931		0.0929	0.0892
T=15	J=16		1.0001	0.9939		0.0592	0.0584

Panel B: Left-side Unbalancedness

$\alpha = 1$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			mean $\widehat{\alpha}$			RMSE	
T=4	J=0	1.0029	1.0029	1.0029	0.1206	0.1206	0.1206
	J=2		1.0089	0.9885		0.1608	0.1575
T=6	J=0	1.0016	1.0016	1.0016	0.0856	0.0856	0.0857
	J=4		1.0232	0.9898		0.1344	0.1277
T=8	J=4	1.0050	1.0033	0.9891	0.1367	0.0923	0.0900
	J=5	1.0044	1.0095	0.9864	0.1865	0.1012	0.0976
	J=6		1.0113	0.9832		0.1135	0.1083
T=10	J=4	1.0031	1.0015	0.9936	0.0923	0.0718	0.0715
	J=5	1.0012	1.0010	0.9913	0.1086	0.0765	0.0759
	J=6	1.0005	1.0040	0.9902	0.1352	0.0823	0.0803
	J=7	1.0034	1.0076	0.9883	0.1878	0.0890	0.0857
	J=8		1.0081	0.9861		0.0940	0.0912
T=15	J=9	1.0055	1.0033	0.9976	0.0934	0.0604	0.0595
	J=10	1.0051	1.0027	0.9958	0.1061	0.0626	0.0620
	J=11	1.0080	1.0048	0.9951	0.1350	0.0661	0.0645
	J=12	1.0085	1.0067	0.9948	0.1873	0.0691	0.0674
	J=13		1.0065	0.9945		0.0718	0.0706

Note: As in the baseline specification, $N = 500$, $\mu_\eta = 0$, $\sigma^2_\eta = 1$, $\pi_0 = -1.25$, $\pi_1 = 0$ so the initial condition of the process is exogenous and not drawn from the steady state, and the unbalancedness is at random. However, here $\alpha = 1$.

Table 8: Monte Carlo Simulation results. Initial condition correlated with η

Panel A: Double Unbalancedness

$\alpha = 0.75$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			$\widehat{\alpha}$			RMSE	
T=4	J=0	0.7535	0.7535	0.7535	0.1258	0.1258	0.1258
	J=2		0.7658	0.7411		0.2207	0.2134
T=6	J=0	0.7506	0.7506	0.7506	0.0852	0.0852	0.0852
	J=2	0.7580	0.7565	0.7519	0.1642	0.1067	0.1057
	J=4		0.7596	0.7489		0.1203	0.1164
	J=6		0.7727	0.7499		0.1403	0.1323
T=8	J=2	0.7518	0.7519	0.7497	0.1007	0.0815	0.0812
	J=4	0.7615	0.7521	0.7488	0.1676	0.0873	0.0867
	J=6		0.7528	0.7475		0.0928	0.0915
	J=8		0.7569	0.7458		0.1006	0.0982
	J=10		0.7608	0.7447		0.1098	0.1074
T=10	J=6	0.7672	0.7505	0.7481	0.1657	0.0709	0.0707
	J=8		0.7514	0.7480		0.0749	0.0742
	J=10		0.7526	0.7463		0.0803	0.0790
	J=12		0.7574	0.7471		0.0855	0.0832
	J=14		0.7594	0.7474		0.0905	0.0874
T=15	J=16		0.7482	0.7455		0.0562	0.0560

Panel B: Left-side Unbalancedness

$\alpha = 0.75$		A1W	A3W_MD	A3Wb	A1W	A3W_MD	A3Wb
			$\widehat{\alpha}$			RMSE	
T=4	J=0	0.7535	0.7535	0.7535	0.1258	0.1259	0.1258
	J=2		0.7596	0.7387		0.1659	0.1611
T=6	J=0	0.7506	0.7506	0.7506	0.0852	0.0852	0.0852
	J=4		0.7684	0.7406		0.1329	0.1268
T=8	J=4	0.7537	0.7525	0.7435	0.1230	0.0903	0.0888
	J=5	0.7554	0.7577	0.7419	0.1664	0.0989	0.0943
	J=6		0.7594	0.7391		0.1075	0.1038
T=10	J=4	0.7506	0.7499	0.7454	0.0837	0.0690	0.0688
	J=5	0.7473	0.7489	0.7433	0.0987	0.0725	0.0724
	J=6	0.7468	0.7492	0.7413	0.1199	0.0770	0.0764
	J=7	0.7466	0.7525	0.7401	0.1648	0.0831	0.0816
	J=8		0.7552	0.7390		0.0879	0.0861
T=15	J=9	0.7540	0.7511	0.7477	0.0842	0.0547	0.0546
	J=10	0.7533	0.7508	0.7466	0.0961	0.0574	0.0573
	J=11	0.7563	0.7521	0.7460	0.1219	0.0605	0.0598
	J=12	0.7568	0.7541	0.7455	0.1655	0.0629	0.0620
	J=13		0.7589	0.7454		0.0664	0.0654

Note: As in the baseline specification, $\alpha = 0.75$, $N = 500$, $\mu_\eta = 0$, $\sigma^2_\eta = 1$, $\pi_0 = -1.25$, and the unbalancedness is at random. However, here $\pi_1 = 0.5$ so the initial condition of the process is correlated with η.

Table 9: Monte Carlo Simulation results. The initial condition and the unbalancedness are both correlated with η

Panel A: Double Unbalancedness

$\alpha = 0.75$		A1W	A2W_MD	A2Wb	A2W	A1W	A2W_MD	A2Wb	A2W
			$\widehat{\alpha}$				RMSE		
T=4	J=0	0.7535	0.7535	0.7534	0.7535	0.1258	0.1258	0.1258	0.1257
	J=2		0.7444	0.7555	0.7280		0.2254	0.2618	0.2284
T=6	J=0	0.7506	0.7506	0.7506	0.7506	0.0852	0.0852	0.0852	0.0853
	J=2	0.7398	0.7464	0.7520	0.7418	0.1709	0.1124	0.1264	0.1122
	J=4		0.7409	0.7565	0.7290		0.1178	0.1267	0.1199
	J=6		0.7494	0.7623	0.7250		0.1348	0.1449	0.1398
T=8	J=2	0.7465	0.7493	0.7570	0.7488	0.1099	0.0847	0.1105	0.0867
	J=4	0.7465	0.7477	0.7590	0.7434	0.1765	0.0889	0.0923	0.0885
	J=6		0.7517	0.7644	0.7418		0.0880	0.0967	0.0917
	J=8		0.7529	0.7656	0.7361		0.0964	0.1041	0.0978
	J=10		0.7380	0.7637			0.1025	0.1119	
T=10	J=6	0.7527	0.7520	0.7617		0.1751	0.0740	0.0779	
	J=8		0.7443	0.7571			0.0755	0.0811	
	J=10		0.7470	0.7612			0.0799	0.0869	
	J=12		0.7434	0.7591			0.0873	0.0912	
	J=14		0.7339	0.7581			0.0887	0.0948	
T=15	J=16		0.7437	0.7559			0.0610	0.0641	

Panel B: Left-side Unbalancedness

$\alpha = 0.75$		A1W	A2W_MD	A2Wb	A1W	A2W_MD	A2Wb
			$\widehat{\alpha}$			RMSE	
T=4	J=0	0.7535	0.7535	0.7534	0.1258	0.1258	0.1258
	J=2		0.7440	0.7553		0.1670	0.1939
T=6	J=0	0.7506	0.7506	0.7506	0.0852	0.0852	0.0852
	J=4		0.7490	0.7610		0.1215	0.1290
T=8	J=4	0.7446	0.7460	0.7574	0.1242	0.0833	0.0913
	J=5	0.7428	0.7464	0.7575	0.1866	0.0914	0.1005
	J=6		0.7469	0.7612		0.0956	0.1031
T=10	J=4	0.7499	0.7489	0.7570	0.0973	0.0734	0.0764
	J=5	0.7458	0.7490	0.7589	0.1097	0.0734	0.0763
	J=6	0.7465	0.7481	0.7581	0.1389	0.0757	0.0809
	J=7	0.7459	0.7457	0.7568	0.1787	0.0824	0.0865
	J=8		0.7479	0.7589		0.0842	0.0888
T=15	J=9	0.7514	0.7487	0.7564	0.0935	0.0607	0.0628
	J=10	0.7481	0.7445	0.7520	0.1115	0.0590	0.0622
	J=11	0.7551	0.7464	0.7557	0.1367	0.0618	0.0644
	J=12	0.7589	0.7472	0.7556	0.1857	0.0643	0.0666
	J=13		0.7477	0.7579		0.0677	0.0701

Note: As in the baseline specification, $\alpha = 0.75$, $N = 500$, $\mu_\eta = 0$, $\sigma_\eta^2 = 1$, $\pi_0 = -1.25$. However, here $\pi_1 = 0.5$ so the initial condition of the process is correlated with η, and the unbalancedness is also correlated with η the way it is explained in point 5 in section 4.1.

Table 10: Percentage of Monte Carlo Simulations that achieved convergence for the Minimum Distance estimation

		Baseline	Higher State Dependence	Correlated Ini. Condit.	Correlated IC and Unbalancedness
		A3W_MD in Tables 3 and 4	A3W_MD in Table 7	A3W_MD in Table 8	A2W_MD in Table 9
Panel A: Double Unbalancedness					
T=4	J=0	100.0	100.0	100.0	100.0
	J=2	100.0	100.0	100.0	99.0
T=6	J=0	100.0	100.0	100.0	100.0
	J=2	100.0	100.0	100.0	99.4
	J=4	100.0	99.9	100.0	97.5
	J=6	99.8	99.4	99.9	93.0
T=8	J=2	100.0	100.0	100.0	99.6
	J=4	100.0	100.0	100.0	98.3
	J=6	100.0	100.0	100.0	96.6
	J=8	100.0	100.0	100.0	92.5
	J=10	99.1	95.5	98.5	87.1
T=10	J=6	100.0	100.0	100.0	96.5
	J=8	100.0	100.0	100.0	95.0
	J=10	100.0	100.0	100.0	92.0
	J=12	100.0	99.7	100.0	89.6
	J=14	95.0	89.2	93.9	79.2
T=15	J=16	100.0	99.9	100.0	86.0
Panel B: Left-side Unbalancedness					
T=4	J=0	100.0	100.0	100.0	100.0
	J=2	100.0	100.0	100.0	98.6
T=6	J=0	100.0	100.0	100.0	100.0
	J=4	100.0	99.7	100.0	96.2
T=8	J=4	100.0	100.0	100.0	97.2
	J=5	100.0	100.0	100.0	95.3
	J=6	99.9	99.3	99.9	92.7
T=10	J=4	100.0	100.0	100.0	96.3
	J=5	100.0	100.0	100.0	96.0
	J=6	100.0	100.0	100.0	92.9
	J=7	100.0	100.0	100.0	90.5
	J=8	99.5	98.3	99.5	86.8
T=15	J=9	100.0	100.0	100.0	90.1
	J=10	100.0	100.0	100.0	85.5
	J=11	100.0	99.9	100.0	83.7
	J=12	100.0	99.7	100.0	77.5
	J=13	96.8	91.3	96.8	73.8

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

Table 11: Monte Carlo Simulation results on the estimation of the AMEs. The initial condition is correlated with η

Panel A: Double Unbalancedness

		A1W		A3W_MD		A3Wb		A1W	A3W_MD	A3Wb
		AME	$\widehat{AME}$	AME	$\widehat{AME}$	AME	$\widehat{AME}$		RMSE	
T=4	J=0	0.2019	0.2034	0.2019	0.2034	0.2019	0.2034	0.0416	0.0416	0.0416
	J=2			0.2019	0.2092	0.2019	0.2014		0.0764	0.0732
T=6	J=0	0.2021	0.2024	0.2021	0.2024	0.2021	0.2024	0.0273	0.0273	0.0273
	J=2	0.2021	0.2088	0.2021	0.2052	0.2021	0.2033	0.0602	0.0362	0.0355
	J=4			0.2021	0.2068	0.2021	0.2022		0.0414	0.0397
	J=6			0.2021	0.2093	0.2021	0.2023		0.0476	0.0449
T=8	J=2	0.2021	0.2039	0.2021	0.2031	0.2021	0.2021	0.0339	0.0265	0.0262
	J=4	0.2021	0.2109	0.2021	0.2034	0.2021	0.2019	0.0612	0.0286	0.0281
	J=6			0.2021	0.2038	0.2021	0.2015		0.0306	0.0299
	J=8			0.2021	0.2048	0.2021	0.2009		0.0333	0.0323
	J=10			0.2021	0.2045	0.2021	0.2004		0.0360	0.0354
T=10	J=6	0.2020	0.2133	0.2020	0.2025	0.2020	0.2016	0.0625	0.0229	0.0227
	J=8			0.2020	0.2030	0.2020	0.2016		0.0244	0.0240
	J=10			0.2020	0.2032	0.2020	0.2009		0.0264	0.0256
	J=12			0.2020	0.2043	0.2020	0.2011		0.0279	0.0270
	J=14			0.2020	0.2040	0.2020	0.2011		0.0296	0.0285
T=15	J=16			0.2021	0.2018	0.2021	0.2011		0.0180	0.0178

Panel B: Left-side Unbalancedness

		A1W		A3W_MD		A3Wb		A1W	A3W_MD	A3Wb
		AME	$\widehat{AME}$	AME	$\widehat{AME}$	AME	$\widehat{AME}$		RMSE	
T=4	J=0	0.2019	0.2034	0.2019	0.2034	0.2019	0.2034	0.0416	0.0416	0.0416
	J=2			0.2019	0.2065	0.2019	0.1975		0.0589	0.0547
T=6	J=0	0.2021	0.2024	0.2021	0.2024	0.2021	0.2024	0.0273	0.0273	0.0273
	J=4			0.2021	0.2086	0.2021	0.1981		0.0462	0.0425
T=8	J=4	0.2021	0.2058	0.2021	0.2042	0.2021	0.2002	0.0436	0.0304	0.0292
	J=5	0.2021	0.2081	0.2021	0.2060	0.2021	0.1995	0.0614	0.0335	0.0312
	J=6			0.2021	0.2055	0.2021	0.1983		0.0369	0.0346
T=10	J=4	0.2020	0.2036	0.2020	0.2026	0.2020	0.2009	0.0285	0.0226	0.0223
	J=5	0.2020	0.2028	0.2020	0.2023	0.2020	0.2002	0.0340	0.0239	0.0236
	J=6	0.2020	0.2032	0.2020	0.2026	0.2020	0.1995	0.0423	0.0256	0.0250
	J=7	0.2020	0.2048	0.2020	0.2037	0.2020	0.1991	0.0605	0.0280	0.0269
	J=8			0.2020	0.2037	0.2020	0.1987		0.0297	0.0286
T=15	J=9	0.2021	0.2051	0.2021	0.2031	0.2021	0.2022	0.0286	0.0178	0.0176
	J=10	0.2021	0.2051	0.2021	0.2030	0.2021	0.2018	0.0331	0.0188	0.0185
	J=11	0.2021	0.2071	0.2021	0.2035	0.2021	0.2017	0.0435	0.0199	0.0194
	J=12	0.2021	0.2093	0.2021	0.2040	0.2021	0.2015	0.0612	0.0208	0.0202
	J=13			0.2021	0.2039	0.2021	0.2015		0.0218	0.0213

Note: As in the baseline specification, $\alpha = 0.75$, $N = 500$, $\mu_\eta = 0$, $\sigma_\eta^2 = 1$, $\pi_0 = -1.25$, and the unbalancedness is at random. However, here $\pi_1 = 0.5$ so the initial condition of the process is correlated with η.