Not So Fixed Effects: Correlated Structural Breaks in Panel Data

Hugo Kruiniger*
Department of Economics,
Queen Mary, University of London

This version December 2008

Abstract

In this paper we generalize conventional panel data models by allowing for, possibly multiple, individual effects that are subject to random structural breaks. We assume that the data have been collected for a large number (N) of cross-sectional units ('individuals') and a fixed number (T) of time periods. The pattern of the breaks, i.e., the timing, the number and the magnitude(s) of the breaks, can be different across the individuals. The breaks, that is, the changes in the individual effects may be correlated over time, e.g., they may have a factor structure with correlated heterogeneous factor loadings or follow MA processes, possibly of different orders. Furthermore, just like the individual effects may be correlated with some of the regressors, the breaks may be correlated with changes in the regressors.

The LSDV estimator is not consistent for a static panel data model with breaks in the individual effects that are correlated with the regressors. However, such models with correlated breaks can be consistently estimated by the IV method or GMM.

We distinguish between individual effects with *common* correlated breaks and those with *individual* correlated breaks. The former can be modelled through an identifiable factor structure. We consider estimation of models that contain both time-varying individual effects with an identifiable factor structure and individual effects with individual correlated breaks. To deal with such models we generalize an estimation algorithm for factor models proposed in Kruiniger (2008). We also discuss several tests for the absence of various kinds of correlated breaks.

Finally, we consider GMM estimation of dynamic panel data models with breaks in the individual effects.

*Mile End Road, London E1 4NS, United Kingdom; E-mail: h.kruiniger@qmul.ac.uk.

Keywords: panel data, structural breaks, correlated breaks, individual breaks, common breaks, latent variables, correlated individual effects, multiple time-varying individual effects, factor models, GMM, dynamic panel data models.

JEL Classification: C11, C14, C23.

1. Introduction

In the last few decades empirical research in especially microeconomics but also in other areas has greatly benefitted from the increasing availability of panel data sets. One of the attractive features of panel data is that in regression analysis they allow one to take unobserved 'individual' effects into account that are correlated with (some of) the observed explanatory variables. This is usually done either by including an individual specific constant term in the model (fixed effects) or by adding a time-invariant individual specific random term and some functions of the regressors, e.g. time series averages of some of the regressors, to the model ((correlated) random effects).[1]

In this paper we will argue that one can in fact consistently estimate more general panel regression models that allow for random changes in the latent individual effects that are correlated with changes in the regressors. That is, one can estimate models with correlated (structural) breaks in the individual effects.[2] The timing, the number and the magnitude(s) of the changes in the effects may differ across the individuals, are not known or specified, and essentially unrestricted.[3]

[1] Since the seminal paper by Mundlak (1961), numerous papers have proposed econometric methods for estimating models for panel data, e.g. Wallace and Hussain (1969), Mundlak (1978), Hausman and Taylor (1981), and Chamberlain (1982) for static linear panel data models, Balestra and Nerlove (1966), Anderson and Hsiao (1981, 1982), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991), and Ahn and Schmidt (1995) for dynamic panel data models, Chamberlain (1980) for panel models with binary data, Honoré (1992) for truncated and censored regression models for panel data, Nijman and Verbeek (1992) for panel data models with selectivity, and Chamberlain (1992) and Wooldridge (1999) for other nonlinear panel data models. Horowitz and Markatou (1996) and Li and Stengos (1996) considered semiparametric estimation of panel data models.

[2] Strictly speaking, referring to the changes in the individual effects as (random) structural breaks constitutes an abuse of terminology since we will view the individual effects as random (latent) variables rather than as parameters; hence it would be better to say that the individual effects are time-varying. However, in general the individual effects change infrequently and resemble parameters that are subject to random structural breaks. Because of this analogy it is convenient to use the structural breaks terminology.

[3] However, we do (need to) make a number of assumptions about the variances of the breaks, the autocovariances of the breaks and the covariances of the breaks with the instruments (e.g. lagged values of the regressors), such as the existence of the variances of the breaks and the lack

One can even have a correlated break in the effects in every period. We assume that only the individual effects may exhibit breaks and that the breaks are independent across the 'individuals'. The asymptotic properties of the estimators for the common parameters are derived by assuming that $N \rightarrow \infty$ and T is fixed. We do not estimate the break dates or the number of breaks or their magnitudes.

The individual effects with correlated breaks that are introduced in this paper are different from the time-varying correlated individual effects that have been considered by Kiefer (1980), Holtz-Eakin et al. (1988), and more recently by Ahn, Lee and Schmidt (2001, 2006), Bai (2005) and Pesaran (2006). In these papers time-invariant individual effects, g_i, are multiplied by common time effects, θ_t, leading to $g_i\theta_t$. Such factor structures are sometimes used to model strong cross-sectional dependence, cf. Andrews (2005). However, they can also be used to model (possibly multiple) *common* correlated breaks, i.e., synchronous breaks in the individual effects (possibly at multiple points in time) that are correlated with changes in the regressors. Such breaks (changes) in the effects can be uncorrelated, imperfectly correlated or perfectly correlated over time, i.e., uncorrelated / (im-) perfectly correlated with common correlated breaks that occur in other periods. Common breaks at different points in time are perfectly correlated if for each 'individual' $i \leq N$ the breaks affect the same factor loading, say the $p-th$ factor loading $g_{i,p}$; they are imperfectly correlated over time if the the breaks affect different factor loadings, say $g_{i,p}$ and $g_{i,q}$, with $Cov(g_{i,p}, g_{i,q}) \neq 0$, $p \neq q$ and $i \leq N$. Note that not all correlated breaks are or even can be modelled as common correlated breaks, that is, as a factor structure that is identifiable.[4] Therefore we will sometimes use the phrase '*individual* correlated breaks' for correlated breaks that are not (or cannot be) modelled through an (identifiable) factor structure.

From an empirical point of view the generalization of the traditional panel data model with time-invariant individual effects to panel data models that include individual effects with correlated breaks is useful. Consider the following examples. When estimating production functions with panel data, one typically includes individual effects to capture unobserved technology and/or management effects. These effects tend to be correlated with the inputs. Both (the quality of) the technology and the management of a firm (or farm) may well change during the sample period and any such changes are likely to be correlated with changes in the inputs. When estimating earnings, labour supply or consumption models, the individual effects may represent unobserved personal attributes and/or socio-

of correlation between the breaks and the instruments. The assumptions we make are natural and often rather weak or the weakest that can be made.

[4]For example, if a factor structure is used to model correlated breaks in the effects that occur in every period in the sample and are uncorrelated or imperfectly correlated over time, then it is unidentified.

economic conditions like ability, skills, training, relevant work experience, health, 'marital status', family composition, (local) economic conditions, social environment et cetera. These attributes/circumstances may well change over time and are often correlated with changes in observed explanatory variables like age, education, (family) income et cetera.

The presence of correlated breaks in the individual effects has implications for the choice of the estimation strategy. Conventional estimators for the slope parameters in static panel regression models with correlated time-invariant individual effects, such as the LSDV (or Within) estimator and the Correlated Random Effects GLS estimator, will no longer be consistent when the individual effects are time-varying and the changes in the effects are correlated with changes in (some of) the regressors. However, in that case it may still be possible to consistently estimate the model in first-differences by the IV method or GMM using lagged levels or lagged first-differences of the regressors as instruments even without knowing the number of breaks or when and for which 'individuals' they occur.

The traditional panel regression models with time-invariant individual effects can be regarded as special cases of models with time-varying correlated individual effects such as effects with correlated breaks. The extra structure of the traditional models gives rise to the availability of additional moment conditions. Therefore one can test for the absence of correlated breaks in the individual effects by employing a Hausman-type test or a Sargan-Difference test.

In the simplest version of a panel regression model with random structural breaks in the individual effects, the changes in the individual effects are serially uncorrelated. However, if the 'structural changes' take place in a gradual manner, then it may be appropriate to model them as moving average (MA) processes. A model may even contain multiple individual effects that differ in terms of the orders of the MA representations of their first-differences. If certain conditions are met, then it is possible to consistently estimate the (maximum) number of individual effects with individual correlated breaks, which is equal to the maximum of the orders of the MA models for the correlated breaks plus one.

The panel models with correlated breaks that have been introduced above are special cases of models that contain both correlated individual effects that have a factor structure and individual effects that exhibit individual correlated breaks. We consider estimation of models that include both kinds of time-varying individual effects. Among other things we describe an estimation algorithm based on Kruiniger (2008) for models that include a factor structure. We also propose a test of H_{0a} : no common correlated breaks that are uncorrelated over time and no individual correlated breaks versus H_{1a} : not H_{0a}, and a sequential test of H_{0b} : no individual correlated breaks versus H_{1b} : not H_{0b}. Both tests allow for common correlated breaks (also under their null hypotheses) that are correlated over time.

However, the sequential test also allows for an increasing but limited number of common correlated breaks that are uncorrelated over time.

The usual IV and GMM estimators for the panel AR(1) model that employ lagged differences of the data as instruments for the model in first-differences are still consistent when the changes in the individual effects are correlated with the changes in (some of) the regressors other than the lagged dependent variable; even lagged levels of the data could serve as instruments as long as the changes in the individual effects are not correlated with those lagged levels.

The plan of the paper is as follows. In section 2 we consider estimation of static panel data models with individual correlated breaks that are serially uncorrelated. In section 3 we generalize the static panel data model by allowing for (possibly multiple) individual effects with breaks that follow MA processes of different orders, we discuss estimation of these more general models and we describe a method for determining the (maximum) number of different individual effects. In section 4 we consider estimation of even more general static models that also include (possibly multiple) time-varying correlated individual effects that have a factor structure. We also discuss several tests for the absence of various types of correlated breaks. Section 5 considers estimation of the panel AR(1) model with random structural breaks. In section 6 we present an empirical example where it is important to allow for time-varying correlated individual effects. In section 7 we conduct some simulation experiments. Section 8 concludes.

2. Estimation of static panel data models with random structural breaks

In this section we will consider estimation of the following static linear panel data model with random structural breaks in the individual effects:

$$
\begin{aligned}
y_{i,t} &= \beta' x_{i,t} + \varepsilon_{i,t}, \quad i = 1, ..., N, \; t = 1, ..., T, \\
\varepsilon_{i,t} &= f_{i,t} + u_{i,t}, \\
\triangle f_{i,t} &= f_{i,t} - f_{i,t-1} = v_{i,t}, \quad \text{for } t \geq 2.
\end{aligned} \tag{2.1}
$$

The panel covers N cross-sectional units (individuals) and T time periods. We assume that N is large and T is small so that asymptotic properties of estimators will be derived assuming that $N \to \infty$ while T remains fixed. The dependent variable is $y_{i,t}$ and $x_{i,t}$ is a $K-$vector of time-varying regressors. The idiosyncratic errors are denoted by $u_{i,t}$ and the unobserved individual effects are denoted by $f_{i,t}$. For convenience, we assume here that the cross-sectional averages of the data are zero. Alternatively, we could include additive time dummies in the model, see section 4.

Let $X_{i,s}^t = (x'_{i,s} \; ... \; x'_{i,t})$, $X_i^t = X_{i,1}^t$, $X_i = X_i^T$, $\Delta X_{i,s}^t = (\Delta x'_{i,s} \; ... \; \Delta x'_{i,t})$, $\Delta X_i^t = \Delta X_{i,2}^t$, $F_i^t = (f_{i,1} \; ... \; f_{i,t})$, and $F_i = F_i^T$. Let $\Sigma = E[(X_i \; F_i)'(X_i \; F_i)] = \begin{bmatrix} \Sigma_{XX} & \Sigma'_{FX} \\ \Sigma_{FX} & \Sigma_{FF} \end{bmatrix}$. In addition, let $u_i^t = (u_{i,1} \; ... \; u_{i,t})$ and $u_i = u_i^T$.

We make the following Standard Assumptions [SA]:

SA.1:	$W_i = (X_i \; F_i \; u_i)$ is i.i.d. over i, and $E(W_i) = 0$,	(2.2)
SA.2:	$0 < E(u_{i,t}^2) = \sigma_t^2 < \infty$, for $t \geq 1$,	(2.3)
SA.3:	Σ is finite, $\det(\Sigma_{XX}) > 0$,	(2.4)
	$\det[E(\Delta x_{i,t-1} \Delta x'_{i,t})] \equiv d_t \neq 0$, for $t > 2$, and $\sum_{t=3}^T d_t \neq 0$,	
SA.4:	$(X_i \; F_i)$ is uncorrelated with u_i,	(2.5)
SA.5:	$E(v_{i,t} \Delta x_{i,t}) \neq 0$, for $t \geq 2$,	(2.6)
SA.6:	$E(v_{i,t} \Delta X_i^{t-1}) = 0$, for $t > 2$,	(2.7)
SA.7:	$E(v_{i,t} \Delta X_{i,t+1}^T) = 0$, for $2 \leq t \leq T-1$.	(2.8)

The individual effects may or may not change over time. Regardless $E(v_{i,t}) = 0$, $t \geq 2$, cf. assumption (2.2). [5] The break dates of the individual effects and the identities of the individuals for which the effects change in a particular period, i.e., the values of i and t for which $f_{i,t} \neq f_{i,t-1}$, are assumed to be unknown. However, according to assumption (2.6) the changes in the $f_{i,t}$, the $v_{i,t}$, are contemporaneously correlated with the changes in all the elements of the $x_{i,t}$. [6] This assumption implies that $\Pr(f_{i,t} \neq f_{i,t-1}) > 0$. If the $v_{i,t}$ and the $\Delta x_{i,s}$ would be uncorrelated $\forall s, t \in \{2, ..., T\}$ and $\forall i \in \{1, ..., N\}$, the vector of slope coefficients, β, could be consistently estimated by a conventional panel data estimator, e.g. the Least Squares Dummy Variable estimator. [7]

The i.i.d assumption about W_i in (2.2) is standard in the traditional panel data context where N is large relative to T. One can allow for heterogeneous distributions by strengthening the assumptions on the moments of the $W_i = [X_i \; F_i \; u_i]$. One can also relax the independence assumption, for instance, by allowing for common factors in the X_i and the F_i with possibly heterogeneous factor loadings

[5] Given $E(u_i) = 0$, the assumption that the cross-sectional averages of the data are equal to zero is equivalent to the assumption that both $E(X_i) = 0$ and $E(F_i) = 0$. When the model includes additive time dummies, it is not only possible to allow $E(X_i) \neq 0$ but it is also possible to recenter the individual effects in every period such that the assumption $E(F_i) = 0$ still holds even when the cross-sectional averages of the data are different from zero.

[6] One can replace (2.6) by the more general assumption that $E(v_{i,t} \Delta x_{k,i,t}) \neq 0$, for some $k \in \{1, ..., K\}$ and for some $t \in \{T_0, ..., T_1\}$ with $2 \leq T_0 \leq T_1 \leq T$. We use (2.6) instead in order to keep the exposition as simple as possible.

[7] Nevertheless an optimal GMM estimator for β is in general asymptotically more efficient than the LSDV estimator.

(cf. papers cited in the introduction and see section 4 below) or by assuming that $\{u_i\}_{i=1}^N$ and/or $\{v_i\}_{i=1}^N$ are/is strong mixing (cf. Conley, 1999). The strict exogeneity assumption (2.5) is also standard but stronger than necessary for the consistency of various estimators for (2.1): in a number of cases assumption (2.5) can be replaced by the weaker condition that the $x_{i,s}$ and the $f_{i,s}$ are predetermined with respect to the $u_{i,t}$:

$$\text{SA.4}': \quad [X_i^t \; F_i^t] \text{ is uncorrelated with } u_{i,t}, \text{ for } t \geq 1. \tag{2.9}$$

For the simple model above identification and estimation of β can be based on assumptions (2.4), (2.5) or (2.9), (2.7) and possibly (2.8).

Assumptions (2.6)-(2.8) are high level conditions, which are implied by more primitive conditions. Suppose that the Regressors, the $x_{i,t}$, obey the following autoregressive model:

$$x_{i,t} - \delta_t f_{i,t} = \Gamma_t(x_{i,t-1} - \delta_{t-1} f_{i,t-1}) + \xi_{i,t}, \quad i = 1, ..., N, \; t = 2, ..., T, \tag{2.10}$$

which includes the following Assumptions [RA]:

$$\begin{array}{ll} \text{RA.1:} & x_{i,1} = \delta_1 f_{i,1} + \xi_{i,1}, \\ \text{RA.2:} & \xi_i = (\xi'_{i,1} \; ... \; \xi'_{i,t}) \text{ is uncorrelated with } F_i. \end{array} \tag{2.11}$$

This model allows for correlation between the regressors and the individual effects. It follows from (2.10) that $\Delta x_{i,t} = (\Gamma_t - I)(x_{i,t-1} - \delta_{t-1} f_{i,t-1}) + \xi_{i,t} + \delta_t v_{i,t} + (\delta_t - \delta_{t-1}) f_{i,t-1}$. If $\Gamma_t = \Gamma_{t-1} = 0$, then $\Delta x_{i,t} = \Delta \xi_{i,t} + \delta_t v_{i,t} + (\delta_t - \delta_{t-1}) f_{i,t-1}$. Let $\delta_1 = \delta_2 = ... = \delta_T \neq 0$. Then the following Primitive Assumptions [PA]:

$$\text{PA.1:} \quad \Pr(v_{i,t} \neq 0) > 0, \text{ for } t \geq 2, \tag{2.12}$$

and

$$\text{PA.2:} \quad E(v_{i,t} v_{i,t-s}) = 0, \text{ for } s = 1, ..., t-2, \; t = 3, ..., T, \tag{2.13}$$

are (necessary and) sufficient for assumptions (2.6) and (2.7) (and (2.8)), respectively, to hold. [8]

Assumption (2.7) is a very weak lack of correlation assumption concerning the changes in the individual effects and the regressors: it only assumes absence of correlation between the $v_{i,t}$ and the *lagged first-differences* of the regressors. [9] A stronger assumption than (2.7) would be that the $v_{i,t}$ are uncorrelated with the *lagged levels* of the regressors, i.e.,

$$\text{SA.6}': \quad E(v_{i,t} X_i^{t-1}) = 0, \text{ for } t \geq 2. \tag{2.14}$$

[8] Conditions (2.12) and (2.13) are still (necessary and) sufficient for assumptions (2.6) and (2.7) (and (2.8)) to hold if the $(x_{i,t} - \delta_t f_{i,t})$ follow more general ARMA processes.

[9] Of course, an even weaker assumption is that $E(v_{i,t} \Delta x_{i,2}) = 0$ for $t > 2$.

If the levels of the regressors satisfy (2.10)-(2.11), then assumption (2.14) is implied by (2.13) and

$$\text{PA.3:} \qquad E(v_{i,t}f_{i,1}) = 0, \text{ for } t \geq 2, \tag{2.15}$$

whatever the values of δ_1, δ_2, ... , δ_T.

In the subsection below we consider additional restrictions on the correlations between the regressors and the individual effects. In the next section we discuss variations of assumption (2.7).

Finally we will show that under the SA, (2.2)-(2.8), the LSDV (Within or Covariance) estimator is generally inconsistent for β in model (2.1). The LSDV estimator is the traditional estimator for static panel data models with time-invariant correlated effects, see Hsiao (1986). Let us first assume that $T = 2$. Then $\widehat{\beta}_{LSDV} = \left(\sum_{i=1}^{N} \Delta x_{i,2}\Delta x'_{i,2}\right)^{-1} \sum_{i=1}^{N}(\Delta x_{i,2}\Delta y_{i,2})$ and

$$\begin{aligned}
\text{plim}_{N\to\infty}\widehat{\beta}_{LSDV} &= \beta + \text{plim}_{N\to\infty}\left(\sum_{i=1}^{N} \Delta x_{i,2}\Delta x'_{i,2}\right)^{-1} \sum_{i=1}^{N}(\Delta x_{i,2}\Delta \varepsilon_{i,2}) \\
&= \beta + \text{plim}_{N\to\infty}\left(\sum_{i=1}^{N} \Delta x_{i,2}\Delta x'_{i,2}\right)^{-1} \sum_{i=1}^{N}(v_{i,2}\Delta x_{i,2}) \\
&\neq \beta, \qquad (2.16)
\end{aligned}$$

because $E(\Delta x_{i,2}v_{i,2}) \neq 0$. [10] When $T > 2$,

$$\text{plim}_{N\to\infty}\widehat{\beta}_{LSDV} - \beta = T^{-1}\sum_{t=2}^{T}(T-t+1)(t-1)\eta_t, \tag{2.17}$$

where

$$\begin{aligned}
\eta_t &= \text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N} \eta_{i,t} \\
\eta_{i,t} &= (N^{-1}\sum_{i=1}^{N} \widetilde{X}'_i(I_T - T^{-1}\iota\iota')\widetilde{X}_i)^{-1}(v_{i,t}\Delta x_{i,t}),
\end{aligned}$$

$\widetilde{X}_i = (x_{i,1}\ ...\ x_{i,T})'$ and ι is a T-vector of ones. In general, the size of the inconsistency of the LSDV estimator, $\text{plim}_{N\to\infty}\widehat{\beta}_{LSDV} - \beta$, depends on

- the cross-sectional averages of the relative sizes of the breaks, the η_t, and
- the timing of the structural breaks.

Note that even if $\eta_s = \eta_t \neq 0$ for some $s \neq t$, the effects of η_s and η_t on the value of $\text{plim}_{N\to\infty}\widehat{\beta}_{LSDV} - \beta$ are different. When at least two of the η_t have different signs, it is still possible that $\text{plim}_{N\to\infty}\widehat{\beta}_{LSDV} = \beta$.

[10] The Correlated Random Effects GLS estimator of Mundlak (1978) is also inconsistent.

2.1. Instrumental Variable and GMM estimation

When estimating panel models, a common way of dealing with correlated individual effects is to transform the model. First-differencing of the model in (2.1) yields

$$\begin{aligned} \Delta y_{i,t} &= \beta' \Delta x_{i,t} + \Delta \varepsilon_{i,t} \\ &= \beta' \Delta x_{i,t} + v_{i,t} + \Delta u_{i,t}. \end{aligned} \tag{2.18}$$

Given the assumptions in (2.4), (2.9) (or (2.5)), and (2.7) (and (2.8)), all lags (and leads) of the first-differences of the regressors are valid instruments for estimating the model in first-differences whether or not a structural break occurs. If we would know that no break occurs in period t, i.e., $v_{i,t} = 0$, then $x_{i,t-1}$ (and $x_{i,t}$) would also be (a) valid instrument(s) for (2.18). However, in many applications exact knowledge about all the individual break dates is not available.

Notice that even though the individual effects change over time, differencing of the model is still important: the current and lagged differences of the regressors are not valid instruments for the equations in levels because $E(f_{i,t}\Delta x_{i,s}) = E([f_{i,1} + \sum_{k=2}^{t} v_{i,k}]\Delta x_{i,s}) = E([f_{i,1} + \sum_{k=2}^{s} v_{i,k}]\Delta x_{i,s}) \neq 0$ if $s \leq t$.

A simple consistent IV estimator for (2.1) is given by

$$\widehat{\beta}_{IV} = \left(\sum\nolimits_{i=1}^{N} \sum\nolimits_{t=3}^{T} (\Delta x_{i,t-1} \Delta x'_{i,t})\right)^{-1} \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=3}^{T} (\Delta x_{i,t-1} \Delta y_{i,t}). \tag{2.19}$$

A GMM estimator that is asymptotically more efficient than $\widehat{\beta}_{IV}$ when $T > 3$ optimally exploits the following moment conditions:

$$E[\Delta x_{i,s}(\Delta y_{i,t} - \beta' \Delta x_{i,t})] = 0, \quad s = 2, ..., t-1, \; t = 3, ..., T. \tag{2.20}$$

The optimal weight matrix for the moment conditions in (2.20) depends on the covariance matrices of the $D\varepsilon_i$, where $\varepsilon_i = (\varepsilon_{i,1} \; ... \; \varepsilon_{i,T})'$ and D is a $(T-2) \times T$ first difference matrix with $d_{k,k+1} = -1$, $d_{k,k+2} = 1$, $k = 1, ..., T-2$, and $d_{k,l} = 0$ elsewhere. Given a preliminary consistent estimator $\widehat{\beta}_1$ for β, for instance $\widehat{\beta}_{IV}$, one can consistently estimate the blocks of the inverse of the optimal weight matrix, $\Omega = [\omega_{p,q}]$, by

$$\widehat{\omega}_{0.5(l-1)(l-2)+k,\; 0.5(t-1)(t-2)+s} = N^{-1} \sum_{i=1}^{N} \Delta x_{i,k} \Delta x'_{i,s} (\Delta y_{i,l} - \widehat{\beta}'_1 \Delta x_{i,l})(\Delta y_{i,t} - \widehat{\beta}'_1 \Delta x_{i,t}),$$

$$k = 2, ..., l-1, \; l = 3, ..., T, \; s = 2, ..., t-1, \; t = 3, ..., T, \tag{2.21}$$

if the fourth-order moments of ΔW_i exist. This estimator for the optimal weight matrix allows for serial correlation in the $\{u_{i,t}\}$, random structural breaks in the

$\{f_{i,t}\}$, and serial (and cross-sectional) heteroskedasticity of the $\{u_{i,t}\}$ and the $\{v_{i,t}\}$.

If the regressors are strictly exogenous as in (2.5) and if assumption (2.8) also holds, then we have the following additional moment conditions for (2.1):

$$E[\Delta x_{i,s}(\Delta y_{i,t} - \beta' \Delta x_{i,t})] = 0, \;\; s = t+1, ..., T, \;\; t = 2, ..., T-1. \tag{2.22}$$

Replacing assumption (2.7) by the stronger assumption (2.14) leads to the following additional moment conditions:

$$E[x_{i,1}(\Delta y_{i,t} - \beta' \Delta x_{i,t})] = 0, \quad t = 2, ..., T. \tag{2.23}$$

When the autocorrelation of $\{\Delta x_{i,t}\}$ is weak, a GMM estimator based on (2.20) and (2.22) may suffer from a weak instruments problem. If that is the case, then the availability of lagged levels of the regressors as valid instruments may be exploited to mitigate the problem. Notice that current and future levels of the regressors are ruled out as instruments for the equations in first-differences because $E(x_{i,s}v_{i,t}) \neq 0$ if $s \geq t$.

Now consider the assumption that

$$E(f_{i,1}\Delta x_{i,t}) = 0, \;\; \text{for } t \geq 2. \tag{2.24}$$

This assumption is valid when the levels of the regressors satisfy (2.10)-(2.11) with $\delta_1 = \delta_2 = ... = \delta_T$, and $\{f_{i,t}\}$ satisfies (2.15).

Assumption (2.24) in combination with (2.5) implies the following moment conditions:

$$E[\Delta x_{i,t}(y_{i,1} - \beta' x_{i,1})] = 0, \;\; t = 2, ..., T. \tag{2.25}$$

These moment conditions may allow one to identify and estimate the coefficients of time-invariant regressors. Assumption (2.24) is very similar to an identification assumption first considered by Breusch, Mizon, and Schmidt (1989) in the context of model with time-invariant regressors and correlated time-invariant individual effects. Notice that current and lagged differences of the regressors are still ruled out as instruments for the equations in levels because $E(f_{i,t}\Delta x_{i,s}) \neq 0$ if $s \leq t$.

The moment conditions in (2.22) (and (2.25)) use future differences of the regressors as instruments, whereas the moment conditions in (2.20) (and (2.23)) use only lagged values of the regressors as instruments. Therefore, none of the moment conditions in (2.20), ((2.23), (2.25)) and (2.22) are redundant.

The moment conditions in (2.20) (and (2.22)) are valid whether or not structural breaks occur. They do not presuppose any knowledge about the timing of the structural breaks but only that the breaks are uncorrelated over time. Indeed these moment conditions are even valid if a structural break occurs in every period, as would be the case if $\{f_{i,t}\}$ were, for instance, a random walk process.

If no structural break occurs additional moment conditions are available and hence more efficient GMM estimators for β can be constructed. Therefore, one can formulate various tests for the absence of structural breaks, such as a Hausman test and a Sargan-Difference test. Even if knowledge about the presence of structural breaks in the individual effects is immaterial from an economic theory point of view, these tests may still be useful as diagnostic tests.

When the regressors are strictly exogenous and no structural break occurs, an efficient GMM estimator optimally exploits $E(m_{1,i,s,t}(\beta)) = E(x_{i,s}(\Delta y_{i,t} - \beta'\Delta x_{i,t})) = 0,\ 1 \leq s \leq T$ and $2 \leq t \leq T$. Let us denote this estimator by $\widehat{\beta}_{GMM,NB}$. A related GMM estimator that is still consistent under random structural breaks that are uncorrelated over time optimally exploits the moment conditions in (2.20) and (2.22), i.e., $E(m_{2,i,s,t}(\beta)) = E(\Delta x_{i,s}(\Delta y_{i,t} - \beta'\Delta x_{i,t})) = 0$, $s \neq t$ with $2 \leq s, t \leq T$. We will denote this estimator by $\widehat{\beta}_{GMM,RB}$. Then we can construct the following Hausman-type test-statistic:

$$\begin{aligned} H_{NB} &= (\widehat{\beta}_{GMM,NB} - \widehat{\beta}_{GMM,RB})'[Var(\widehat{\beta}_{GMM,RB}) - Var(\widehat{\beta}_{GMM,NB})]^{-1} \times \\ &\qquad (\widehat{\beta}_{GMM,NB} - \widehat{\beta}_{GMM,RB}). \end{aligned} \tag{2.26}$$

Under the null hypothesis of no structural breaks, $H_{NB} \sim \chi^2(K)$.

An alternative test is based on the Sargan-Difference test-statistic:

$$SD_{NB} = \sum_i m'_{1,i}[\sum_i(\widetilde{m}_{1,i}\widetilde{m}'_{1,i})]^{-1}\sum_i m_{1,i} - \sum_i m'_{2,i}[\sum_i(\widetilde{m}_{2,i}\widetilde{m}'_{2,i})]^{-1}\sum_i m_{2,i}, \tag{2.27}$$

where $m_{1,i} = m_{1,i}(\widehat{\beta}_{GMM,NB})$, $m_{2,i} = m_{2,i}(\widehat{\beta}_{GMM,RB})$, $\widetilde{m}_{1,i} = m_{1,i}(\widehat{\beta}_{1,NB})$ and $\widetilde{m}_{2,i} = m_{2,i}(\widehat{\beta}_{1,RB})$ with $m_{1,i}(\beta)$ and $m_{2,i}(\beta)$ vectors comprising the functions $m_{1,i,s,t}(\beta)$, $1 \leq s \leq T$ and $2 \leq t \leq T$, and $m_{2,i,s,t}(\beta)$, $s \neq t$ with $2 \leq s, t \leq T$, respectively, and $\widehat{\beta}_{1,NB}$ and $\widehat{\beta}_{1,RB}$ preliminary consistent estimators for β. Under the null hypothesis of no structural breaks $SD_{NB} \sim \chi^2(2(T-1)K)$. The test with more degrees of freedom will have less power than the alternative test.

3. A model with multiple correlated individual effects

Often structural changes take more than one period to complete. For instance, the adoption of a new technology by a firm or a change in the organisation of the firm are usually slow processes that involve learning and/or adjustment costs. Also a change in the international economic or political environment may only gradually reach its full impact on a national economy. Even if a structural change occurs rapidly, it may still affect more than one observation if the change starts near the end of the time interval between two observations. In many of these cases it seems appropriate to model the changes in the individual effects as moving average processes and to replace assumption (2.7) by

$$E(v_{i,t}\Delta X_i^{t-1-q+r}) = 0, \text{ for } t > 2+q-r, \text{ for some } q \geq 0 \text{ and } r, \tag{3.1}$$

where q is the maximum order of autocorrelation of the individual processes $\{v_{i,t}\}$. Assumption (3.1) allows for MA processes for the correlated breaks, the $v_{i,t}$, that are different across the individuals and/or time. Furthermore, it allows for the possibility that there is a lag period ($r > 0$) or a lead period ($r < 0$) between the start of the breaks and the start of the correlated changes in the values of the regressors. [11]

Under assumptions (2.9) and (3.1), the following moment conditions in (2.20) for model (2.1) remain valid:

$$E[\Delta x_{i,s}(\Delta y_{i,t} - \beta'\Delta x_{i,t})] = 0, \;\; s = 2, ..., \min(t-1-q+r, T), \;\; t = 2, ..., T. \tag{3.2}$$

Suppose that the structural change starts in period t, i.e., $v_{i,t} \neq 0$ whereas $v_{i,t-1} = 0$, then $E[\Delta x_{i,s}(\Delta y_{i,l} - \beta'\Delta x_{i,l})] = 0$, $s = l-q+r, ..., t-1+r$, $l = t, ..., t+q-1$ would also be valid. However, in general we do not have precise information about the start of the structural changes and hence we cannot exploit these additional moment conditions.

If the correlated breaks follow MA(q) processes, i.e., $v_{i,t} = \sum_{k=0}^{q} \alpha_{i,t,k} w_{i,t-k}$ with $\alpha_{i,t,0} = 1$ and $q > 0$, then it is quite possible that some of the conditions in (2.8) and (2.24) are not met. For instance, suppose that $v_{i,t} \neq 0$ while $v_{i,t-1} = 0$,

[11] Suppose that $v_{i,t} \sim MA(q)$ if $v_{i,t} \neq 0$, i.e. $v_{i,t} = \sum_{k=0}^{q} \alpha_{i,t,k} w_{i,t-k}$ with $\alpha_{i,t,0} = 1$, and suppose that the structural change begins in period $t-q$, so that $w_{i,t-q} \neq 0$. [If the structural change begins in period s with $t-q < s \leq t$, then $w_{i,t-k} = 0$ for $k = t-s+1, ..., q$.] Now the lag period r could be larger than 0 either because $E(w_{i,t-q+p} x_{i,l}) = 0$ for $p = 0, ..., r-1$ and $l = 1, ..., t-q+r-1$ while $E(w_{i,t-q+r} x_{i,t-q+r}) \neq 0$, or because $E(w_{i,t-q} x_{i,l}) = 0$ for $l = 1, ..., t-q+r-1$ while $E(w_{i,t-q} x_{i,t-q+r}) \neq 0$, or because of some combination of these reasons. Only in the former case the break $v_{i,t}$ can be written as $v_{i,t} = v_{1,i,t} + v_{2,i,t}$ where $v_{1,i,t} \sim MA(q-r)$ is a correlated break and $v_{2,i,t} \sim MA(q)$ is an uncorrelated break. The difference between the two cases is reflected in the number of 'leads' of the first-differences of the regressors that can serve as instruments when assumption (2.5) holds as well.

then the covariances between $v_{i,t}$ $(= w_{i,t})$ and each of $x_{i,t+r}, ..., x_{i,t+q+r}$ may be different. However, as long as assumption (2.5) holds and

$$
\begin{aligned}
&E(v_{i,t}\Delta x_{i,s}) = 0, \text{ for } s = t+1+q+r, ..., T, \\
&\qquad\qquad\qquad\qquad t = 2, ..., \min(T-q-r-1, T), \text{ and} \\
&E(f_{i,1}\Delta x_{i,t}) = 0, \text{ for } t = 2+q+r, ..., T,
\end{aligned}
\tag{3.3}
$$

respectively, then we still have the following additional moment conditions: [12]

$$
\begin{aligned}
&E[(\Delta y_{i,t} - \beta'\Delta x_{i,t})\Delta x_{i,s}] = 0, \;\; s = t+1+q+r, ..., T, \\
&\qquad\qquad\qquad\qquad t = 2, ..., \min(T-q-r-1, T), \text{ and} \\
&E[(y_{i,1} - \beta' x_{i,1})\Delta x_{i,s}] = 0, \;\; s = 2+q+r, ..., T,
\end{aligned}
\tag{3.4}
$$

respectively.

One can test joint hypotheses about the maximum order q of the MA processes for the correlated breaks and the lag/lead period r by means of a Hausman test or a Sargan Difference test.

In the sequel we will assume that $r = 0$ in order to keep the exposition as simple as possible.

So far we have assumed that all the regressors for an individual are correlated with one and the same individual effect process $\{f_{i,t}\}$ and that the changes in all these regressors are correlated with $\{\Delta f_{i,t}\}$. However, the error term may include several individual effect processes with correlated breaks that differ from each other in terms of the maximum order of autocorrelation exhibited by their first-differences, i.e., the value of q. The error term may, of course, also contain a conventional individual effect that is constant over time. Thus the model may have the following composite error term:

$$
\begin{aligned}
&\varepsilon_{i,t} = \mu_i + \textstyle\sum_{q=0}^{Q} f_{q,i,t} + u_{i,t}, \text{ where} \qquad (3.5)\\
&f_{q,i,t} = f_{q,i,t-1} + v_{q,i,t}, \;\; q = 0, ..., Q, \\
&(X_i \;\mu_i \;F_i \;u_i) \text{ is i.i.d. over } i, \text{ and } E(\mu_i \;F_i \;u_i) = 0 \text{ with} \\
&\qquad f_{i,t} = (f_{0,i,t}, ..., f_{Q,i,t}) \text{ and } F_i = (f_{i,1} \;...\; f_{i,T}), \\
&u_i \text{ is uncorrelated with } (X_i \;\mu_i \;F_i), \\
&E(X_i \mu_i) \neq 0, \\
&E(v_{q,i,t}\Delta x_{i,t-q}) \neq 0, \;\text{ for } t \geq q+2, \;\; q = 0, ..., Q, \\
&E(v_{q,i,t}\Delta X_i^{t-1-q}) = 0, \;\text{ for } t > q+2, \;\; q = 0, ..., Q.
\end{aligned}
$$

[12] If the adjustment of the regressors takes longer than the adjustment of the correlated individual effects, then some of the constant covariance conditions in (3.3) will not hold. In that case some of the moment conditions in (3.4) are not valid.

Each individual effect in $\varepsilon_{i,t}$ may have a different interpretation. For instance, when estimating farm production functions, μ_i could be a location effect (soil, climate) and $\{f_{3,i,t}\}$ could be a 'new technology' effect.

We say that a regressor $x_{k,i,t}$ is correlated with the individual effect $\{f_{q,i,s}\}$ if for some $t > q+1$, $\Delta x_{k,i,t-q} = x_{k,i,t-q} - x_{k,i,t-1-q}$ is correlated with $v_{q,i,t} = \Delta f_{q,i,t}$, i.e., if $E(v_{q,i,t}\Delta x_{k,i,t-q}) \neq 0$ for some $t > q+1$. A regressor may be correlated with more than one individual effect. However, for the selection of the lags of a particular regressor that can serve as instruments for the equations in first-differences, one only needs to know the maximum of the maximum orders of autocorrelation, i.e., the maximum of the values of the q's, of the changes (first-differences) of the individual effects that are correlated with the changes in that regressor. If a regressor is only correlated with the time-invariant individual effect μ_i, all its leads and lags and contemporaneous values are valid instruments for the equations in first-differences. Of course, it is also possible that a regressor is not correlated at all with the composite error term. In that case all leads and lags and the contemporaneous values of the regressor are valid instruments for the equations in levels. On the other hand when $Q > T-2$, there is at least one regressor that cannot be used as instrument.

When the number of regressors is fixed, one can consistently select (all the) valid moment conditions (or instruments) that only involve $Y_i = (y_{i,1} \ldots y_{i,T})$ and X_i by using a sequential testing procedure that is based on the J statistic for testing overidentifying restrictions, or by using a procedure that is based on minimizing a criterion function that consists of the J statistic and a bonus term that rewards the use of more moment conditions, see Andrews (1999) and Andrews and Lu (2001) for details. The bonus term can be based on the Schwarz (BIC) or the Hannan-Quinn (HQIC) criterion, see the text after Propositions 4.2 and 4.3 below for examples. When $Q \leq T-2$, one can also use such procedures to consistently estimate the value of Q, i.e., the maximum of the orders of the MA models for the correlated breaks, which is equal to the maximum number of individual effects with correlated breaks minus one.

Let $Z_i^s = \bigcup_{t=2}^{T} Z_{i,t}^s$ be a set of instruments for the system of $T-1$ equations of the model in first-differences where $Z_{i,t}^s$ is a subset that contains the instruments for the equation corresponding to $\Delta y_{i,t}$. The number of sets of instruments considered in the aforementioned procedures should be chosen as small as possible otherwise the estimators for Q (and possibly β) may have poor finite sample properties. Therefore it may be useful to only consider sets that consist of blocks of instruments that correspond to a minimum value s for the lags of the instruments, e.g. Z_i^s, $s = 0, 1, 2, \ldots, (T-2)$ with $Z_{i,t}^s = \Delta X_i^{t-s}$, $s = 0, 1, 2, \ldots, (t-2)$, $t = 2, \ldots, T$.

4. A model with common and individual correlated breaks

Individual effects that exhibit individual correlated breaks are different from correlated individual effects that have a common factor structure. We will now consider a general model whose composite error term includes multiple time-varying individual effects of both types. Unlike Ahn et al. (2006), we also allow for additive time effects in the model. Let $\zeta = (\zeta_1 \ ... \ \zeta_T)'$ be a parameter vector comprising T (fixed) time effects, let $\Theta = [\theta_{t,p}]$ be a $T \times P$ (with $P < T$) constant matrix whose columns are P common factors, $\Theta_1, ..., \Theta_P$, and let $g_i = (g_{i,1} \ ... \ g_{i,P})$ be a P-vector of factor loadings. Then the general model reads as

$$y_{i,t} = \beta' x_{i,t} + \zeta_t + \varepsilon_{i,t}, \quad i = 1, ..., N, \ t = 1, ..., T, \tag{4.1}$$

$$\varepsilon_{i,t} = \sum_{p=1}^{P} g_{i,p}\theta_{t,p} + \sum_{q=0}^{Q} f_{q,i,t} + u_{i,t}, \tag{4.2}$$

$$f_{q,i,t} = f_{q,i,t-1} + v_{q,i,t}, \ q = 0, ..., Q.$$

Note that $\sum_{p=1}^{P} g_{i,p}\Theta_p = \Theta g_i'$. We sometimes use the subscript 0 when we need to distinguish the true value of a parameter from an arbitrary or hypothetical value of the parameter. For instance, P_0 and Θ_0 denote the true values of P and Θ. Recalling that $f_{i,t} = (f_{0,i,t}, ..., f_{Q,i,t})$ and $F_i = (f_{i,1} \ ... \ f_{i,T})$, and letting $\widetilde{X} = (\widetilde{X}_1' \ ... \ \widetilde{X}_N')'$ and $\Sigma = E[(X_i - \mu_X \quad F_i)'(X_i - \mu_X \quad F_i)]$ with $\mu_X = E(X_i)$, we make the following Standard Assumptions about the General model [GSA]:

GSA.1:	$(X_i \ g_i \ F_i \ u_i)$ is i.i.d. over i, and $E(g_i \ F_i \ u_i) = 0$,	(4.3)
GSA.2:	$0 < E(u_{i,t}^2) = \sigma_t^2 < \infty$, for $t \geq 1$,	(4.4)
GSA.3:	$0 < E(g_{i,p}^2) = \tau_p^2 < \infty$, for $p = 1, ..., P_0$,	(4.5)
GSA.4:	Σ is finite, $\det(\Sigma_{XX}) > 0$,	(4.6)
GSA.5:	u_i is uncorrelated with $[X_i \ g_i \ F_i]$,	(4.7)
GSA.6:	$\text{rank}(\Theta_0) = \text{rank}E(X_i' g_i) = P_0$,	(4.8)
GSA.7:	$(\widetilde{X} \quad I_N \otimes \Theta_0)$ is of full column rank,	(4.9)
GSA.8:	$E(v_{q,i,t}\Delta x_{i,t-q}) \neq 0$, for $t \geq q+2$, $q = 0, ..., Q$,	(4.10)
GSA.9:	$E(v_{q,i,t}\Delta X_i^{t-1-q}) = 0$, for $t > q+2$, $q = 0, ..., Q$.	(4.11)

Below we will consider estimation of this general model and various tests of common versus individual correlated breaks.

Note that if g_i and Θ satisfy GSA (with Θ_0 replaced by Θ and $P = P_0$), then so do $\widetilde{g}_i = g_i R'$ and $\widetilde{\Theta} = \Theta R^{-1}$, where R is an arbitrary nonsingular $P \times P$ matrix. Moreover $\Theta g_i' = \widetilde{\Theta}\widetilde{g}_i'$. Therefore, apart from the conditions in (4.8) and (4.9), which imply that $\dim(g_i) = P_0$ and $\text{rank}E(g_i' g_i) = \dim(g_i)$, some additional restrictions on the g_i and the parameters in Θ are required in order to identify the elements of Θ. One possible parametrization of Θ is $\Theta = [I_P\ \Theta'_{**}]'$ where Θ_{**} is an unrestricted $(T-P)\times P$ matrix, cf. Ahn et al. (2006). However, this parametrization is overly restrictive as it rules out that $\Theta_{p,p}=0$ for some or all $p \in \{1,...,P\}$. Another possible parametrization is $E(g_i'g_i) = I_P$ and

$$\begin{aligned} \Theta &= [\Theta'_*\ \Theta'_{**}]' \quad \text{where} \\ &\quad \Theta_* \text{ is a lower triangular } P\times P \text{ matrix and} \\ &\quad \Theta_{**} \text{ is an unrestricted } (T-P)\times P \text{ matrix.} \end{aligned} \tag{4.12}$$

The latter parametrization of Θ allows any row in Θ to equal zero. However, this parametrization is not practical for estimation as it relies on imposing the restrictions $E(g_i'g_i) = I_P$, which is not straightforward. A solution to this problem is to replace the $0.5P(P+1)$ restrictions on $E(g_i'g_i)$ by $0.5P(P+1)$ orthonormality restrictions on Θ. This leads to our preferred parametrization: $\text{rank}E(g_i'g_i) = P$ and $\Theta = [\Theta'_*\ \Theta'_{**}]'$ with

$$\Theta'\Theta = I_P. \tag{4.13}$$

These restrictions uniquely determine Θ up to the signs of the columns. However, knowledge about the signs of the factors is immaterial.

Most of the conditions in GSA are similar to those in SA, which have already been discussed in section 2. The second assumption in (4.6) rules out time-invariant regressors and regressors that are constant across individuals. The coefficients of such regressors are not identified when the model contains additive time effects, ζ, and/or constant correlated individual effects. We now discuss the assumptions that are specific to the individual effects with a common factor structure. The assumption $E(g_i) = 0$ in (4.3) enables identification of ζ. The assumptions in (4.8) and (4.9) are also identification conditions. The assumption $\text{rank}(\Theta_0) = P_0$ in (4.8) stipulates that the common factors in Θ_0 are linearly independent. The assumptions $\text{rank}(\Theta_0) = P_0$ and $\text{rank}E(g_i'g_i) = P_0$ taken together imply that there exist no $\widetilde{g}_i$ and $\widetilde{\Theta}$ such that $\widetilde{\Theta}\widetilde{g}_i' = \Theta_0 g_i'$ and $\dim(\widetilde{g}_i) < P_0$. On the other hand there exist infinitely many factor structures $\widetilde{\Theta}\widetilde{g}_i'$ with $\text{rank}\widetilde{\Theta} = \dim(\widetilde{g}_i) > P_0$ such that $\widetilde{\Theta}\widetilde{g}_i' = \Theta_0 g_i'$. That is, there are infinitely many $(T\times P)$ matrices $\widetilde{\Theta}$ with $\text{rank}\widetilde{\Theta} = P > P_0$ for each of which there exists a matrix R such that $\widetilde{\Theta}R = \Theta_0$. Putting $\widetilde{g}_i' = Rg_i'$, it immediately follows that $\widetilde{\Theta}\widetilde{g}_i' = \Theta_0 g_i'$. Observe that although $\text{rank}\widetilde{\Theta} = \dim(\widetilde{g}_i) > P_0$, $\text{rank}E(\widetilde{g}_i'\widetilde{g}_i) = P_0$. Thus using more than P_0 linearly independent factors leads to a singular covariance matrix for the corresponding factor loadings. Note also that amongst all the equivalent factor structures that satisfy the restrictions $\text{rank}E(\widetilde{g}_i'\widetilde{g}_i) = P_0$ and $\widetilde{\Theta} = [\Theta'_*\ \Theta'_{**}]'$ with $\widetilde{\Theta}'\widetilde{\Theta} = I_{\dim(\widetilde{g}_i)}$, $\Theta_0 g_i'$ is the only factor structure with P_0 factors, i.e., the smallest number of linearly independent factors. The assumption $\text{rank}E(X_i'g_i) = P_0$ is another identification condition that requires that the loadings of all the factors in Θ_0 are correlated with the regressors. If it were the case that $E(X_i' g_i) < P_0$, then some of the columns of Θ would not be identified unless additional restrictions had been imposed on the covariance matrix of the u_i, $Var(u_i)$. Note that a possible constant correlated individual effect would be included in $\Theta_0 g_i'$. Finally, assumption (4.9) rules out certain types of perfect multicollinearity, see Ahn et al. (2006).

We will now discuss GMM estimation of β, ζ and Θ in the model given by (4.1), (4.2) and the GSA in (4.3)-(4.11) with $f_{q,i,t} = 0$ for all possible values of q, i and t, ('$Q = -1$') i.e., the model without individual effects that are subject to correlated breaks. We will also impose the restrictions on Θ given in (4.12) and (4.13). We will assume that P_0 is known and that $P = P_0$. In reality P_0 is unknown. Below we discuss how P_0 can be consistently estimated. Our GMM estimator uses (lagged and possibly leading values of) the regressors as instruments. Since the factor loadings are correlated with the regressors by assumption (4.8), we need to transform the model to obtain an error term that is uncorrelated with the instruments. Define the following $(T \times T)$ matrix

$$H(\theta) = I_T - \Theta\Theta', \tag{4.14}$$

where $\theta = vec(\Theta)$ is a vector comprising the $TP - 0.5(P-1)P$ elements of $\Theta = [\Theta_*'\ \Theta_{**}']'$. Note that under (4.13) $H(\theta)$ is an idempotent matrix with $H(\theta)\Theta = 0$ and rank $T - P$. Let $\overline{H}(\theta)$ be a $((T-P) \times T)$ matrix comprising $T - P$ rows of $H(\theta)$ such that rank$\overline{H}(\theta) = T - P$. The matrix $\overline{H}(\theta)$ also depends on all the elements of θ and $\overline{H}(\theta)\Theta = 0$.

Let $y_i = (y_{i,1} \ldots y_{i,T})$. Assumption (4.3) implies that $E(\Theta_0 g_i' + u_i') = 0$. After adding the assumption that the regressors are strictly exogenous, we also have $E(X_i \otimes (\overline{H}(\theta_0)(\Theta_0 g_i' + u_i'))) = 0$. This suggests that estimation of $\delta = (\beta'\ \zeta'\ \theta')'$ may be based on the following moment conditions

$$\begin{aligned} E(m_{1,i}(\delta)) &= E(Z_i \otimes (\overline{H}(\theta)(y_i' - \widetilde{X}_i'\beta - \zeta))) = 0, && (4.15) \\ E(m_{2,i}(\delta)) &= E(y_i' - \widetilde{X}_i'\beta - \zeta) = 0, && (4.16) \end{aligned}$$

and the restrictions on Θ given in (4.12) and (4.13), where Z_i is a subvector of X_i or the vector X_i itself. If $(T-P) \times \dim(Z_i) \geq \dim(\theta) - 0.5P(P+1) + K$ (which equals $(T-P)P + K$), then the restrictions in (4.12) and (4.13), $P = P_0$ and the assumptions (4.8) and (4.9) guarantee that β_0, ζ_0 and θ_0 are the unique solutions of (4.15) and (4.16). Note that $(T-P) \times \dim(X_i) = (T-P)TK \geq (T-P)P + K$ so that in principle there are enough moment conditions available for identification of δ_0. In case ζ_0 is in the span of Θ_0 so that $\overline{H}(\theta)\zeta = 0$, the moment conditions in (4.15) can be used to estimate β and those in (4.16) can be used to estimate ζ.

Direct GMM estimation of δ based on (4.15), (4.16), (4.12) and (4.13) may be cumbersome due to the fact that both (4.15) and (4.13) are nonlinear in the elements of θ. We will now describe an algorithm that will be an important ingredient of an indirect but computationally easier estimation method for δ that we discuss subsequently and is based on Kruiniger (2008). We call a $((T-P)\times T)$ matrix $\Pi=[\pi_{s,t}]$ uppertriangular if $\forall s\ \exists t_s$ such that $\pi_{u,t_s}=0\ \forall u>s$ and $\pi_{u,t}=0$ $\forall u\geq s$ and $\forall t<t_s$. This definition allows for $\pi_{s,t_s}\neq 0$ and $\pi_{s,t_s}=0$. When $\pi_{s,t_s}\neq 0\ \forall s$, Π has full rank. It is easily verified that for any $T\times P$ matrix Θ that satisfies the restrictions in (4.12) and (4.13) there exists an uppertriangular $((T-P)\times T)$ matrix of full rank, say Π, that satisfies $\Pi\Theta=0$ and has rows that contain up to $P+1$ nonzero elements and satisfy the normalisation restrictions

$$\begin{aligned} \pi_s'\pi_s &= 1\ \forall s, \qquad \text{and} \\ \forall s\ \exists t_s \text{ such that } \pi_{s,t_s} &> 0 \text{ and } \pi_{s,t}=0\ \forall t<t_s, \end{aligned} \tag{4.17}$$

where π_s' is the s-th row of Π.[13] On the other hand, any uppertriangular $((T-P)\times T)$ matrix Π of full rank whose rows satisfy (4.17) will uniquely determine (up to their signs) the columns of a $T\times P$ matrix Θ that satisfies (4.12) and (4.13) via the restrictions $\Pi\Theta=0$. A useful algorithm for finding an uppertriangular $((T-P)\times T)$ matrix Π of full rank that satisfies (4.17) and $\Pi\Theta_0=0$ (for a given matrix Θ_0) is as follows. Step 1: Initially set $\pi_{s,t}=0$ when $s>t$ or $s<t-P$. In this case Π is a band-diagonal matrix with up to $P+1$ free, i.e., possibly nonzero elements in each of its $T-P$ rows. Step 2: Solve $\Pi\Theta_0=0$ subject to (4.17). If rank$(\Pi)=T-P$, then a unique solution, Π^0, has been found. However, it is also possible (depending on Θ_0) that rank$(\Pi^0)<T-P$ and that some rows of Π are not uniquely determined. These problems will occur when a matrix that consists of (at least) P consecutive rows of Θ_0 bar the first and last row of Θ_0 has rank less than P. For example, suppose that a matrix that consists of rows $s,...,s+P-1$ of Θ_0 has rank less than P with $1<s\leq T-P$. Then it is possible that $\pi_{s-1}^0=\pi_s^0$. In this case we have $\pi_{s-1,s-1}^0=0$. It is also possible that π_s is not uniquely determined by $\Pi\Theta_0=0$ and (4.17). In order to obtain a unique row π_s^0 select π_s with the lowest value t_s for which $\pi_{s,t_s}\neq 0$ (in fact $\pi_{s,t_s}>0$). If π_s is still not uniquely determined, then select a solution π_s for which the remaining elements are equal to zero or, if such a solution does not exist, for which the next nonzero element has the lowest column index. Continue this selection process until a unique π_s^0 has been obtained. Even when unique rows have been obtained, some

[13] Since Θ is a $T\times P$ matrix of full rank, there exists an $((T-P)\times T)$ matrix Π^* of full rank that satisfies $\Pi^*\Theta=0$. By performing linear operations on Π^* one can obtain an uppertriangular matrix Π^{**} of full rank $(T-P)$ that satisfies $\Pi^{**}\Theta=0$ and has rows that contain up to $P+1$ nonzero elements (because rank$(\Pi^*)=T-P$, one can find $T-P$ linear combinations of the rows of Π^* that each have at least $T-P-1$ zero elements). There exists a unique nonsingular diagonal matrix, say $D_{\pi^{**}}$, such that $\Pi=D_{\pi^{**}}\Pi^{**}$ satisfies (4.17). Furthermore, $\Pi=D_{\pi^{**}}\Pi^{**}$ is full rank and has rows that contain up to $P+1$ nonzero elements and $\Pi\Theta_0=D_{\pi^{**}}\Pi^{**}\Theta=0$.

rows of Π may be linearly dependent. In order to obtain a solution of full rank the linearly dependent rows need to be deleted and replaced by different rows. Thus when $\text{rank}(\Pi^0) < T - P$, the algorithm continues with additional steps. Step 3: Delete a row of Π^0, say π_s^0, and fill the gap (temporarily) with a row (with the same row index s) of zeros, whenever $\pi_{s,s}^0 = 0$. Let P_{del}^{0a} be the number of deleted rows and call the resulting matrix Π^{0a}. Step 4: Add $\min(P_{del}^{0a}, P)$ new rows to $\Pi^{0a} : \pi_s,\ s = T - P + 1, ..., T - P + \min(P_{del}^{0a}, P)$, with $\pi_{s,t} = 0$ when $t < s$. The other elements of these new rows are free. To find the values of these new rows solve $\Pi\Theta_0 = 0$ subject to (4.17) for the nonzero rows. Delete a new row of Π^{0a}, say π_s^{0a}, whenever $\pi_{s,s}^{0a} = 0$. Let P_{del}^{0b} be the number of deleted new rows, let $P_{del}^1 = P_{del}^{0a} - (\min(P_{del}^{0a}, P) - P_{del}^{0b})$ and call the resulting matrix Π^1. Let $m = 1$ and $n = 0$. Step 5: Include P_{del}^m new rows in Π^m by replacing the (first) P_{del}^m rows that were deleted from Π^{m-1} with new rows as follows. Suppose that π_s^{m-1} was deleted (because $\pi_{s,s}^{m-1} = 0$). Then replace this old π_s^{m-1} by a new π_s where $\pi_{s,t},\ t = s, ..., s + P + m$, are free (i.e., possibly nonzero) elements apart from $\pi_{s,t_{s,k}} = 0,\ k = 1, ..., m$, where $t_{s,k}$ is the smallest value for t such that $\pi_{s,t}^{k-1} > 0$; the other elements of π_s are zero. We (can) impose $\pi_{s,t_{s,m}} = 0$ because rows $t_{s,m}$,...,$s + P + m - 1$ of Θ_0 are linearly dependent (since $\pi_{s,s}^{m-1} = 0$, $\pi_{s,t}^{m-1} = 0$, $t < t_{s,m}$ and $\pi_{s,t_{s,k}} = 0,\ k = 1, ..., m-1$) and $\pi_{s,t_{s,m}}^{m-1} > 0$. Note that we try to find a row s, π_s, that satisfies $\pi_s'\Theta_0 = 0$ with $\pi_{s,s} \neq 0$ so that π_s is linearly independent from the other rows of Π. Imposing the restriction $\pi_{s,t_{s,m}} = 0$ does not make it more difficult to find such a π_s since row $t_{s,m}$ of Θ_0 is a linear combination of other rows of Θ_0 whose corresponding elements in π_s are allowed to be different from zero. Also note that the new π_s will again have $P + 1$ free elements. To find the values of the new rows solve $\Pi\Theta_0 = 0$ subject to (4.17) for the nonzero rows. Delete a new row of Π^m, say π_s^m, and fill the gap temporarily with a row of zeros, whenever $\pi_{s,s}^m = 0$. Let P_{del}^{m+1} be the number of deleted new rows and let Π^{m+1} be the resulting matrix. Step 6: Set $m := m + 1$ and repeat step 5 as long as $\text{rank}(\Pi^m) < T - K$. Whenever $\pi_{s,s}^{m-1} = 0$ and $s + P + m > T$, set $n := n + 1$, replace the old π_s^{m-1} by a row of zeros, and replace the $n-th$ of the last $P_{del}^{0a} - P_{del}^1$ rows that were deleted from Π^0 and temporarily replaced by rows of zeros, say row s_2, by a new row according to the method described in step 5 upon resetting $m = 1$ for this row, i.e., replace the original row s_2, $\pi_{s_2}^0$, by a new π_{s_2} where $\pi_{s_2,t}$, $t = s_2, ..., s_2 + P + 1$, are free elements apart from $\pi_{s_2,t_{s_2,1}} = 0$ (where $t_{s_2,1}$ is the smallest value for t such that $\pi_{s_2,t}^0 > 0$); the other elements of π_{s_2} are zero. When $\text{rank}(\Pi^m) = T - K$, drop the rows comprising only zeros. Call the resulting matrix Π_0.

The following example shows an application of the algorithm when $P = 2$:

$$\Theta_0' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ a & a & a & b & b & d & d \end{pmatrix}, \tag{4.18}$$

$$\Pi_0 = \frac{1}{2}\begin{pmatrix} \sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & c(b-d) & 0 & c(d-a) & c(a-b) & 0 \\ 0 & 0 & 0 & \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{2} & -\sqrt{2} \end{pmatrix},$$

with $b \neq d$ and $c^{-1} = \frac{1}{2} sign(b-d)\sqrt{(b-d)^2 + (d-a)^2 + (a-b)^2}$.

We have not imposed the restrictions in (4.12) and (4.13) on Θ_0. Nevertheless, because $b \neq d$, rank$(\Theta_0) = 2$. When also $a \neq b$, the factor structure represents correlated individual effects that are subject to two common correlated breaks (through the second factor), i.e., breaks that affect the correlated individual effects synchronously.[14] Note that as the two common breaks affect the *same* correlated individual effects, i.e., the second factor loadings, the $g_{i,2}$, they are also *perfectly* correlated over time. Since $P = 2$, Π_0 is a 5×7 matrix. The first, second and fourth rows of Π_0 are determined in step 2 of the algorithm. Its fifth row is found in step 4, while its third row is found in step 5 of the algorithm.

Remarks. 1. The algorithm produces a unique $((T-P) \times T)$ matrix Π_0 of full rank that solves $\Pi\Theta_0 = 0$ subject to (4.17) for the following reasons. Firstly, as mentioned above, there exists an uppertriangular $((T-P) \times T)$ matrix Π of full rank that solves $\Pi\Theta_0 = 0$ and satisfies (4.17) and has rows that contain up to $P+1$ nonzero elements. Secondly, if necessary the algorithm will consider all elements of a set of possible solutions that consists of (excluding the rows of zeros) uppertriangular $((T-P) \times T)$ matrices Π that satisfy $\Pi\Theta_0 = 0$ and (4.17) and have rows that contain up to $P+1$ nonzero elements, and includes at least one such matrix whose rank is equal to $T-P$. Note that although the restrictions the algorithm imposes on the rows of the uppertriangular matrices, i.e., restrictions such as $\pi_{s,t_{s,k}} = 0$, $k = 1, ..., m$, may exclude some of the matrices that satisfy $\Pi\Theta_0 = 0$ and (4.17), have rows that contain up to $P+1$ nonzero elements and have rank $T-K$ from the set of possible solutions, these restrictions will not exclude all of them. For example, when a particular matrix Π^e that satisfies $\Pi\Theta_0 = 0$ and (4.17) and has rows that contain up to $P+1$ nonzero elements with $\pi_{s,s} \neq 0$ is not included in the set of possible solutions only because its s-th row does not satisfy

[14] The two essential properties of common correlated breaks, which distinguish them from individual correlated breaks, are (i) that the common break(s) affect(s) the individual effects at the same point(s) in time and (ii) that the/all common break(s) can be modelled (jointly) through a common factor structure that is identifiable. The latter property rules out the existence of common correlated breaks in every period that are uncorrelated or imperfectly correlated over time. We call such breaks individual correlated breaks. Note that a common break may affect only a subgroup of 'individuals' in the sample rather than all of them. In this case some individual effects (factor loadings) are equal to zero.

$\pi_{s,t_{s,k}} = 0$, $k = 1, ..., m$, then the set includes at least one matrix with $\pi_{s,s} \neq 0$ whose s-th row does satisfy $\pi_{s,t_{s,k}} = 0$, $k = 1, ..., m$, and whose other rows are equal to those of Π^e.

2. After imposing the restrictions in (4.17), Π has the same number of free elements as Θ after imposing the restrictions in (4.12) and (4.13), namely $(T-P)P$.

3. Premultiplication of $\Theta_0 g_i' + u_i' = \varepsilon_i' = y_i' - \widetilde{X}_i'\beta - \zeta$ by Π_0 yields a system of $T - P$ linearly independent equations that do not involve $\Theta_0 g_i'$. Each row of Π_0 gives rise to an equation that is a combination of at most $P + 1$ equations from the system $y_i' - \widetilde{X}_i'\beta - \zeta = u_i'$.

Given the relationship between Θ_0 and Π_0 an adapted version of the algorithm can be used in an estimation procedure to obtain unique estimates of Π and δ when the aforementioned identification conditions are met. An (indirect) GMM estimation procedure (algorithm) for $\pi = vec(\Pi)$, β, ζ and θ can be based on

$$E(X_i \otimes \Pi(y_i' - \widetilde{X}_i'\beta - \zeta)) = 0, \tag{4.19}$$

where Π is an uppertriangular $((T - P) \times T)$ matrix whose structure is determined in each step of the algorithm, on the moment conditions in (4.16) and the restrictions in (4.17) as well as on the restrictions given by (4.12) and (4.13) and the restrictions

$$\Pi\Theta = 0. \tag{4.20}$$

The estimation procedure starts with a band-diagonal matrix Π as defined in step 1 of the original algorithm described above. In subsequent steps the estimation procedure modifies the structure of Π in a similar way as the original algorithm but uses and evaluates estimates of Π^0, Π^{0a} and Π^m, $m = 1, 2, ...$ rather than their true values in this process. In particular, at each step the procedure tests whether $\hat{\pi}_{s,t} = 0$ for some s and t, or whether $\text{rank}(\widehat{\Pi}) < T - P$. To test whether $\text{rank}(\widehat{\Pi}) < T - P$ one can, for instance, use the methods of Cragg and Donald (1997), e.g., by testing $H_0 : \text{rank}(\widehat{\Pi}) = T - P - 1$ against $H_A : \text{rank}(\widehat{\Pi}) = T - P$.

Note that the indirect estimation procedure replaces the moment conditions in (4.15) by the moment conditions in (4.19). The latter are linear in the elements of Π. As long as the algorithm has not converged or if one is not interested in the value of θ_0 (Θ_0) but only in the values of β_0 and ζ_0, then the estimation will be based on (4.16), (4.19) and (4.17) alone. When $\text{rank}(\widehat{\Pi}) = T - P$ at the final stage of the estimation algorithm, the restrictions in (4.12), (4.13) and (4.20) can be used to calculate an estimate of θ. In order to obtain standard errors for an estimate of θ one can apply the direct estimation method based on (4.15) and (4.16) using the estimates of β, ζ and θ obtained in the indirect estimation procedure as starting values.

Remarks. 1. The estimation algorithm generalizes the quasi-differencing approach of Chamberlain (1983) and Holtz-Eakin et al. (1988) for models with

$P = 1$.[15] The approach of Holtz-Eakin et al. uses a band-diagonal $(T-1) \times T$ matrix Π with $\pi_{s,t} = 0$ when $s > t$ or $s < t-1$. Moreover $\pi_{t,t} = 1 \ \forall t$. This parametrization of Π is overly restrictive. Even if we replace $\pi_{t,t} = 1 \ \forall t$ by the more flexible normalization (4.17), the two potentially nonzero (so-called free) elements in each row of Π are adjacent. When (an) element(s) of Θ_0 is/are equal to zero, this will lead to selecting the same equation(s) twice and therefore to fewer moment conditions than our approach. In our approach, which can be used when $P \geq 1$, the $P+1$ free elements may be separated by zeros in some rows, as in the example above.

2. The moment conditions included in (4.19) can be based on a subset of equations from the system $\widehat{\Pi}(y_i' - \widetilde{X}_i'\beta - \zeta) = \widehat{\Pi}u_i'$.

3. When $P = P_0$, the procedure will yield consistent estimators for π, β, ζ and θ. However, in finite samples it is possible that the algorithm will produce an estimate of Π, $\widehat{\Pi}$, with $\text{rank}(\widehat{\Pi}) < T - P$ rather than $\text{rank}(\widehat{\Pi}) = T - P$. In that case the estimation procedure will exploit those moment conditions in (4.19) that correspond to a subset of linearly independent equations from the system $\widehat{\Pi}(y_i' - \widetilde{X}_i'\beta - \zeta) = \widehat{\Pi}u_i'$ and it will not be possible to use the procedure to obtain an estimate of θ.

4. When $P < P_0$, the model is misspecified and the procedure will not yield consistent estimators for π, β, ζ and θ. In finite samples the algorithm may produce an estimate of Π, $\widehat{\Pi}$, with $\text{rank}(\widehat{\Pi}) < T-P$ rather than $\text{rank}(\widehat{\Pi}) = T-P$. In that case the estimation procedure will exploit those moment conditions in (4.19) that correspond to a subset of linearly independent equations from the system $\widehat{\Pi}(y_i' - \widetilde{X}_i'\beta - \zeta) = \widehat{\Pi}u_i'$. Irrespective of whether $\text{rank}(\widehat{\Pi}) = T - P$ or not, the $J-$test of overidentifying restrictions will reject their validity w.p.1 when $N \to \infty$ (provided there are overidentifying restrictions).

5. The factor structure can be used to model one or several common correlated breaks in the individual effects as illustrated by the example given above. In general common correlated breaks may be uncorrelated, imperfectly correlated or perfectly correlated over time. When there are only a few (relative to T) common correlated breaks that are uncorrelated over time and no additional individual correlated breaks in the other periods of the panel, modelling such common breaks through a factor structure may be preferable over using the methods described in section 2 that allow for individual correlated breaks in every period. Unlike the 'individual break methods' described in the previous sections, the estimation approach based on the factor model determines the number and the date(s) of the common break(s) and does not allow for breaks at other dates, that is, at dates when they do not occur. This leads to the availability of moment conditions

[15] We note that Holtz-Eakin et al. allow for slope coefficients that vary over time whereas we assume that $\beta_t = \beta \ \forall t$. Furthermore their paper considers a dynamic rather than a static model.

in the estimation algorithm for the factor model that are not exploited by the 'individual break methods'. They correspond to periods in which no breaks occur. On the other hand, the estimation algorithm eliminates the factors completely and allows for at most $T-2$ correlated breaks that are not perfectly correlated over time (provided that there are no other factors unrelated to breaks), whereas the 'individual break methods' allow for correlated breaks in every period, i.e., for (up to) $T-1$ correlated breaks that are uncorrelated over time. The latter methods only take differences between equations for two consecutive periods and do not remove the changes in the individual effects. This leads to the use of extra moment conditions by the 'individual break methods' as compared to the estimation approach based on the factor model. These moment conditions can be added to those that are already exploited by the estimation algorithm for the factor model once the dates of the common breaks have been determined. Thus both estimation approaches for models with correlated breaks use different sets of moment conditions that overlap.

When the regressors are strictly exogenous, the 'factor model GMM estimators' for the slope parameters, β, in a static panel data model with individual effects and $P-1$ common correlated breaks (in the effects) that are not perfectly correlated over time (but without other common factors and without individual correlated breaks), exploit $(T-P)(TK-P)$ moment conditions after ζ and either π or θ have been eliminated. When the breaks are uncorrelated over time, the 'individual break method' for estimating β (in the same model) exploits possibly $(T-1)^2K$ moment conditions, namely (2.20), (2.22) and possibly (2.23). Thus when there are breaks, i.e., when $P \geq 2$, the 'factor GMM estimator' uses not only different but also less moment conditions than the 'individual break method'. Consider the case with only one common break, i.e., with $P = 2$, and no individual breaks. Suppose that the break occurs in period t_b with $2 \leq t_b \leq T$ and assume that $E(u_i|X_i) = 0$ and that the u_i are conditionally homoskedastic and serially uncorrelated, i.e., $E(u_i u_i'|X_i) = \sigma^2 I_T$. Then the optimal 'factor GMM estimator' for β is asymptotically equivalent to an LSDV-type estimator that leaves out the first-differences of the data that correspond to periods t_b and $t_b - 1$. Once the timing of the break, t_b, has been determined, these 'factor GMM estimators' can be improved upon by also exploiting the moment conditions $E[\Delta x_{i,s}(\Delta y_{i,t_b} - \beta'\Delta x_{i,t_b})] = 0$, $2 \leq s \leq T$ with $s \neq t_b$, and $E[x_{i,1}(\Delta y_{i,t_b} - \beta'\Delta x_{i,t_b})] = 0$.

6. It is conceivable, both in micro- and macroeconomic panels, that the dependent variables (and some of the regressors) of the 'individuals' are affected by common (correlated) shocks. These shocks can be modelled through a factor structure, cf. Andrews (2005). When (some of) these shocks affect an arbitrary individual differently at different points in time, the model will contain several

factors that mainly consist of zeros. In such cases it is likely that Θ_0 will contain P consecutive rows that form a matrix of rank less than P and that Π_0 will be different from a band-diagonal matrix with $P+1$ consecutive nonzero diagonals. Moreover, the parametrization of Θ of Ahn et al. (2006) will be too restrictive.

Consistent estimation of P_0 requires that P_0 is uniquely identified. Let P be the number of factors allowed for when estimating the model. The GSA guarantee that if $P < P_0$, then there exist no β, ζ and θ (or π) such that (4.16) and (4.15) (or (4.19)) and the restrictions on θ (or π) are satisfied; if $P = P_0$, then β_0, ζ_0 and θ_0 (or some true π) satisfy the moment conditions and the parameter restrictions; and finally if $P > P_0$, then there exist many combinations of β, ζ and θ (or π) that satisfy the moment conditions and the parameter restrictions. In particular, when $P > P_0$, θ (π) is not uniquely identified. It follows that P_0 is the smallest value of P such that there exist values for β, ζ and θ (or π) that satisfy the moment conditions and the parameter restrictions. Following Ahn et al. (2006), we will now describe two GMM procedures by which both P_0 and δ_0 can be consistently estimated. Both procedures are based on Cragg and Donald (1997). Similar procedures can be described for the estimation of P_0, β_0, ζ_0 and some true π. Let $m_i(\delta) = (m'_{1,i}(\delta)\ m'_{2,i}(\delta))'$. Then the optimal GMM estimator for δ based on (4.15) and (4.16) is given by $\widehat{\delta}(P_0)$ where

$$\widehat{\delta}(P) = \arg\min_{\delta} J_N(\delta, P) \text{ subject to (4.12) and (4.13), and} \tag{4.21}$$

$$J_N(\delta, P) = N\left(N^{-1}\sum_{i=1}^{N} m_i(\delta)\right)'\left(N^{-1}\sum_{i=1}^{N} m_i(\widehat{\delta}_I)m_i'(\widehat{\delta}_I)\right)\left(N^{-1}\sum_{i=1}^{N} m_i(\delta)\right), \tag{4.22}$$

and $\widehat{\delta}_I$ is an initial consistent estimator for δ. The corresponding J-test of overidentifying restrictions is given by $J_N(\widehat{\delta}(P_0), P_0)$. We have the following results on $J_N(\widehat{\delta}(P), P)$:

Proposition 4.1. *Suppose that the model in (4.1) and (4.2) satisfies GSA in (4.3)-(4.11) and that $Q = -1$. Then as $N \to \infty$,*

(i) for $P = P_0$, $J_N(\widehat{\delta}(P_0), P_0) \xrightarrow{d} \chi^2((T-P)(\dim(Z_i) - P) - K)$

(ii) for $P < P_0$, $J_N(\widehat{\delta}(P), P) \xrightarrow{p} \infty$.

The results in proposition 1 allow one to consistently estimate P_0 (and δ_0) by using a (downward) sequential testing procedure based on $J_N(\widehat{\delta}(P), P)$: test $H_0 : P_0 = p$ against $H_A : P_0 > p$ for $p = 0, 1, 2, ...$ until one can no longer reject H_0. Thus the estimator $\widehat{P}_{seq}$ for P_0 is the smallest value for p for which one cannot reject H_0. Consistency of $\widehat{P}_{seq}$ requires that the significance level tends to zero as $N \to \infty$ but no too fast (cf. Cragg and Donald, 1997):

Proposition 4.2. *Let b_N be the significance level used at each step of the sequential testing procedure. If $b_N \to 0$ and -$\ln(b_N)/N \to 0$ as $N \to \infty$, then $\widehat{P}_{seq}$ is a consistent estimator for P_0.*

The choice of the significance level is subjective. Therefore one may prefer an alternative estimation method based on minimizing a criterion function that consists of the J statistic and a bonus term that rewards the use of fewer parameters for a given number of moment conditions or the use of more moment conditions for a given number of parameters. Let $S_N(P) = J_N(\widehat{\delta}(P), P) - \kappa_N((T-P)(\dim(Z_i) - P) - K)$ where $\{\kappa_N\}$ is a sequence of positive constants and $(T-P)(\dim(Z_i) - P) - K$ is the number of overidentifying restrictions. Let $\widehat{P} = \arg\min_P S_N(P)$ We have the following result for $\widehat{P}$:

Proposition 4.3. *Suppose that the model in (4.1) and (4.2) satisfies GSA in (4.3)-(4.11) and that $Q = -1$. Then $\widehat{P}$ is a consistent estimator for P_0 if $\{\kappa_N\}$ is such that $\lim_{N\to\infty} \kappa_N = \infty$ and $\kappa_N = o(N)$.*

The choice of $\{\kappa_N\}$ is subjective. One possibility is $\kappa_N = c\ln(N)$ for some constant $c > 0$. This corresponds to the Schwarz (BIC) criterion. Another possibility is $\kappa_N = c\ln\ln(N)$. This corresponds to the Hannan-Quinn (HQIC) criterion. Given the same value of c and the same number of instruments ($\dim(Z_i)$), the BIC penalizes the use of extra parameters (factors) more than the HQIC. Nevertheless both criteria yields consistent estimators for P_0 (and δ_0).

The downward sequential testing procedure and the penalized criterion function method can also be used to estimate the general model in (4.1) and (4.2) with both common factors and $Q \geq 0$ after applying a first-difference transformation to the model.

4.1. Tests for the absence of individual and common correlated breaks that are uncorrelated over time

When individual correlated breaks are uncorrelated over time and occur in every period in the sample, β can be estimated by GMM based on moment conditions such as those in (2.20). In this case modelling the breaks through a factor structure would require T common factors leaving the model unidentified. On the other hand when the breaks occur in only one or just a few unknown period(s) (relative to T), then modelling the break(s) through a factor structure can lead to a more efficient estimator for β than a GMM estimator based only on moment conditions like (2.20). Thus it is useful to distinguish between individual and common correlated breaks that are uncorrelated over time.

In the general model (4.1) and (4.2) one can test for the absence of individual and possibly common correlated breaks that are uncorrelated over time as follows:

1. Apply the downward sequential testing procedure or the penalized criterion function method to the moment conditions in (4.16) and to the model in first-differences using suitable lags (and possibly leads) of the regressors (if necessary in first-differences) as instruments. This yields estimators $\widehat{P}_U$ and $\widehat{\delta}_U$ for P and δ. They are consistent in the presence of individual and common correlated breaks.

2. Re-estimate the model using (4.16) and (4.15) (or (4.19)) with $P = \widehat{P}_U$ and all lags and current values (and possibly leads) of the regressors as instruments. If there are no individual correlated breaks and no common correlated breaks that are uncorrelated over time, this yields a more efficient estimator $\widehat{\delta}_R$ for δ.

3. Next one can test the hypothesis that there are no individual correlated breaks and no common breaks that are uncorrelated over time either by conducting a Hausman test based on $\widehat{\delta}_R - \widehat{\delta}_U$ or $\widehat{\beta}_R - \widehat{\beta}_U$, or by conducting a Sargan-Difference test based on $J_N(\widehat{\delta}_R(\widehat{P}_U), \widehat{P}_U) - J_N(\widehat{\delta}_U(\widehat{P}_U), \widehat{P}_U)$. If this hypothesis is rejected, then one can sequentially test for the absence of individual correlated breaks whilst allowing for one or several common correlated breaks that are uncorrelated over time as follows:

   Re-estimate the model using (4.16) and (4.15) (or (4.19)) with $P = \widehat{P}_U + p$, where $p = 1, 2, 3, ...$ and all lags and current values (and possibly leads) of the regressors as instruments, and carry out a $J - test$ until the overidentifying restrictions are not rejected or p exceeds an upperbound $\bar{p}$. If the latter happens, then one concludes that the model has individual correlated breaks and/or more than $\bar{p}$ common correlated breaks that are uncorrelated over time.

5. Dynamic panel data models with random structural breaks

In this section we will study consistent estimation of the panel AR(1) model with random structural breaks in the individual effects: [16]

$$\begin{aligned} y_{i,t} &= \rho y_{i,t-1} + \varepsilon_{i,t} \quad (i = 1, ..., N,\ t = 2, ..., T), \\ \varepsilon_{i,t} &= f_{i,t} + u_{i,t}, \\ f_{i,t} &= f_{i,t-1} + v_{i,t}, \qquad \text{for } t \geq 2. \end{aligned} \tag{5.1}$$

We assume that the observed process, $\{y_{i,t}\}$, and the latent individual effect process, $\{f_{i,t}\}$, start in the same period, i.e., at $t = 1$. In addition, $-1 < \rho \leq 1$.

[16] We could easily allow for a (non-zero) common mean as in $y_{i,t} = (1-\rho)c + \rho y_{i,t-1} + \varepsilon_{i,t}$ or for common (additive) time effects by introducing time dummies.

Let $Y_i^t = (y_{i,1} \; ... \; y_{i,t})$, $\Delta Y_i^t = (\Delta y_{i,2} \; ... \; \Delta y_{i,t})$, $u_i^t = (u_{i,2} \; ... \; u_{i,t})$ and $u_i = u_i^T$. Furthermore, let $\Sigma^D = E[(y_{i,1} \; F_i)'(y_{i,1} \; F_i)] = \begin{bmatrix} \sigma_{yy} & \Sigma'_{Fy} \\ \Sigma_{Fy} & \Sigma^D_{FF} \end{bmatrix}$.

We make the following Standard Assumptions with respect to the above Dynamic panel data model [DSA]:

$$\begin{aligned}
&\text{DSA.1:} && (y_{i,1} \; F_i \; u_i) \text{ is i.i.d. over } i, \text{ and } E(F_i \; u_i) = 0, && (5.2)\\
&\text{DSA.2:} && 0 < E(u_{i,t}^2) = \sigma_t^2 < \infty, \text{ for } t \geq 2, && (5.3)\\
&\text{DSA.3:} && \Sigma^D \text{ is finite}, && (5.4)\\
&\text{DSA.4:} && E(u_{i,t} Y_i^{t-1}) = 0, \; E(u_{i,t} F_i^{t-1}) = 0, \text{ for } t \geq 2, && (5.5)\\
&\text{DSA.5:} && \Pr(f_{i,t} \neq f_{i,t-1}) > 0, \text{ for } t \geq 2, && (5.6)\\
&\text{DSA.6:} && E(v_{i,t} \Delta Y_i^{t-2}) = 0, \text{ for } t > 3. && (5.7)
\end{aligned}$$

Given that we are interested in the consequences of changes in the individual effects for estimation, it is practical to make assumption (5.6). Note that assumption (5.2) implies that $E(v_{i,t}) = 0$, $\forall t \geq 2$.

The above Standard Assumptions for the Dynamic model in (5.1) [DSA] can be compared with those listed by Ahn and Schmidt (1995) for the panel AR(1) model with time-invariant individual effects. The assumptions concerning the idiosyncratic errors, the $u_{i,t}$, are similar: in particular, assumptions (5.2) and (5.5) imply that $E(u_{i,t}) = 0$, $\forall t \geq 2$, $E(u_{i,s} u_{i,t}) = 0$, $\forall s, t \geq 2$ and $E(y_{i,1} u_{i,t}) = 0$, $\forall t \geq 2$. However, the assumption of Ahn and Schmidt that $E(f_{i,1} u_{i,t}) = 0$, (and $v_{i,t} = 0$) $\forall t \geq 2$, has been replaced by $E(u_{i,t} F_i^{t-1}) = 0$, $\forall t \geq 2$. Thus the individual effects are assumed to be predetermined. A stronger assumption than (5.5) would be that $E(u_{i,t} Y_i^{t-1}) = 0$ and $E(u_{i,t} F_i) = 0$, for $t \geq 2$, so that the individual effects are strictly exogenous. Such an assumption would imply that $E(u_{i,s} v_{i,t}) = 0$, $\forall s, t \geq 2$ and in combination with assumption (5.7) that $E(v_{i,s} v_{i,t}) = 0$, $\forall s, t$ with $t > s + 1 > 3$. Assumption (5.7) is very similar to assumption (2.7) for the static model although it does not include the assumption that $E(v_{i,t} \Delta y_{i,t-1}) = 0$. A somewhat stronger assumption than (5.7) is that $E(v_{i,t} \Delta y_{i,2}) = 0$, $E(v_{i,t} v_{i,s}) = 0$, $s = 3, ..., t-2$, and $E(v_{i,t} \Delta u_{i,s}) = 0$, $s = 3, ..., t-2$, for $t > 3$. [17] Such an assumption would be implied by the assumptions (5.7) and $E(u_{i,t} F_i) = 0$, for $t \geq 2$.

A simple IV estimator for (5.1) is given by

$$\widehat{\rho}_{IVdif} = \left(\sum_{i=1}^N \sum_{t=4}^T (\Delta y_{i,t-2} \Delta y_{i,t-1}) \right)^{-1} \sum_{i=1}^N \sum_{t=4}^T (\Delta y_{i,t-2} \Delta y_{i,t}).$$

Consistency of $\widehat{\rho}_{IVdif}$ follows immediately from the assumptions (5.5) and (5.7).

[17] If $\{v_{i,t}\}$ follows an MA(q) process, we replace assumption (5.7) by $E(v_{i,t} \Delta Y_i^{t-q-2}) = 0$, for $t > q + 3$.

A GMM estimator that is consistent and asymptotically more efficient than $\widehat{\rho}_{IVdif}$ when $T > 4$, optimally exploits the following moment conditions:

$$E[\Delta y_{i,s}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})] = 0, \quad s = 2, ..., t-2, \ t = 4, ..., T. \tag{5.8}$$

DSA do not rule out that $E(v_{i,t} y_{i,s}) \neq 0$ for some $t \in \{3, ..., T\}$ and $s \in \{1, ..., t-2\}$. To guarantee that the GMM estimator of Arellano and Bond (1991) for the panel AR(1) model, which exploits

$$E[y_{i,s}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})] = 0, \quad s = 1, ..., t-2, \ t = 3, ..., T, \tag{5.9}$$

will be consistent, we need to replace (5.7) in DSA by a stronger assumption, such as

$$E(v_{i,t} Y_i^{t-2}) = 0, \quad \text{for} \quad t \geq 3, \tag{5.10}$$

or $E(v_{i,t} y_{i,1}) = 0$, $E(v_{i,t} f_{i,s}) = 0$, $s = 2, ..., t-2$, and $E(v_{i,t} u_{i,s})$, $s = 2, ..., t-2$, for $t \geq 3$.

The following non-linear moment conditions due to Ahn and Schmidt (1995)

$$E[(y_{i,T} - \rho y_{i,T-1}))(\Delta y_{i,t} - \rho \Delta y_{i,t-1})] = 0, \quad t = 3, ..., T-1, \tag{5.11}$$

are no longer valid when there are breaks in the individual effects between $t = 2$ and $t = T - 1$. In fact, provided that $T \geq 4$, they can be used individually or jointly to test whether there are breaks in the $\{f_{i,t}\}$ at any point in time between $t = 2$ and $t = T - 1$.

Consider the following assumptions which are much stronger than assumptions (5.5) and (5.7) in DSA:

$$\begin{aligned} \text{DSA}' : \quad & E(u_{i,t} Y_i^{t-1}) \text{ and } E(u_{i,t} F_i^T) = 0, \quad \text{for} \quad t \geq 2 & (5.12) \\ & E(v_{i,t} Y_i^{t-1}) \text{ and } E(v_{i,t}\, f_{i,1}) = 0, \quad \text{for} \quad t \geq 2, \quad \text{and} & (5.13) \\ & E(f_{i,1} \Delta y_{i,2}) = 0. & (5.14) \end{aligned}$$

Given the model in (5.1), it follows from the assumptions (5.12) and (5.13) that $E(v_{i,t} Y_i^{t-1}) = 0$ and $E(v_{i,t} F_i^{t-1}) = 0$, for $t \geq 2$, and from the assumptions (5.12)-(5.14) that $E(f_{i,1} \Delta y_{i,t}) = 0$, for $t \geq 2$. Nevertheless, under DSA$'$ and assumption (5.6) the following moment conditions

$$E[\Delta y_{i,s}(y_{i,t} - \rho y_{i,t-1})] = 0, \quad s = 2, ..., t-1, \ t = 3, ..., T, \tag{5.15}$$

which have first been proposed by Arellano and Bover (1995), are still invalid because $E[\Delta y_{i,s}(y_{i,t} - \rho y_{i,t-1})] = E[(\rho \Delta y_{i,s-1} + v_{i,s} + \Delta u_{i,s})(f_{i,t} + u_{i,t})] = E[(\sum_{q=0}^{s-3} \rho^q v_{i,s-q} + \rho^{s-2} \Delta y_{i,2})(f_{i,1} + \sum_{p=2}^{t} v_{i,p})] \neq 0$, for $s = 2, ..., t-1$, $t = 3, ..., T$.

[18] Thus in general the moment conditions in (5.15) can be invalid either because $E(f_{i,1}\Delta y_{i,2}) = 0$ fails to hold or because the individual effects are subject to random structural breaks. On the other hand, the moment conditions due to Arellano and Bond given in (5.9) may still hold when the individual effects are subject to random structural breaks. Therefore, neither the moment conditions given in (5.9) nor those given in (5.15) can be used to construct a Sargan-Difference test (with useful power) for the simple hypothesis that the individual effects are not subject to structural breaks.

Finally, the 'homoskedasticity moment conditions' due to Ahn and Schmidt (1995), i.e.,

$$E[(y_{i,t} - \rho y_{i,t-1})^2 - (y_{i,t-1} - \rho y_{i,t-2})^2] = 0, \quad t = 3, ..., T, \tag{5.16}$$

will also be invalid if $\Pr(f_{i,t} \neq f_{i,t-1}) > 0$, for $t \geq 3$. Thus when $T \geq 3$ one can use these moment conditions to test the joint hypothesis of serial homoskedasticity of the $\{u_{i,s}\}$ and no breaks in the $\{f_{i,s}\}$ after $t = 2$.

6. Empirical example

In progress...

7. Simulation experiments

In progress...

8. Concluding remarks

In this paper the traditional panel data models have been generalized by allowing for, possibly multiple, individual effects that exhibit random structural breaks. The timing, the number and the magnitude(s) of these breaks may or may not differ across the individuals. There may even be breaks in every period of the sample. The changes in the individual effects can be uncorrelated, imperfectly correlated or perfectly correlated over time. Breaks that are correlated over time are modelled either through a factor structure or as MA processes. Moreover, the breaks may be correlated with the changes in the regressors. The LSDV estimator is not consistent for generalized static panel data models where this is the case. However, upon transforming them, such models can be consistently estimated by

[18] When $s = 2$ and $t = 3$, $E[\Delta y_{i,s}(y_{i,t} - \rho y_{i,t-1})] = E[\Delta y_{i,2}(f_{i,3})] = E[\Delta y_{i,2}(f_{i,2})] = E[\Delta y_{i,2}(v_{i,2})] = E[((\rho - 1)y_{i,1} + u_{i,2} + f_{i,2})v_{i,2}] = E[f_{i,2}v_{i,2}] = E[(v_{i,2})^2] > 0$ unless $\Pr(v_{i,2} = 0) = 1$. However, assumption (5.6) implies that $\Pr(v_{i,2} = 0) < 1$.

the IV method or GMM. The paper describes an estimation algorithm that can be used when the model contains common factors.

A further generalization of these models is obtained when the individual effects are modelled as ARMA processes with homogeneous autoregressive parameters rather than as IMA processes. The autoregressive parameters can be allowed to vary over time. Such models can also be consistently estimated by the IV method or GMM after applying a quasi-difference transformation.

After taking first-differences, the problem of estimating a panel data model with correlated breaks in the individual effects that follow MA(1) processes, may seem to be rather similar to that of estimating a panel data model with regressors that are both correlated with the individual effects and subject to serially uncorrelated measurement errors. However, for the latter model extra moment conditions may be available: in particular, past levels of the regressors are in principle valid instruments for the equations in first-differences and if the regressors are strictly exogenous, future levels of the regressors are also valid instruments for the equations in first-differences.

There are two versions of the Arellano-Bond GMM estimator for the panel AR(1) model: one version uses lagged levels of the data as instruments and the other version uses lagged differences of the data as instruments. The latter version of the estimator is more robust than the former version as it may still be consistent when the changes in the individual effects are correlated with the initial conditions, i.e., the $y_{i,1}$, and it is also consistent for a broader class of time-invariant individual effects than the former version, i.e., it is the fixed effects version of this GMM estimator (see Kruiniger (2002)).

Finally, we have seen that in the context of the panel AR(1) model one cannot construct a Sargan-Difference test with useful power for the simple hypothesis that the individual effects do not exhibit any structural breaks during the *entire* sample period: however, the absence of structural breaks in the effects between $t = 1$ and $t = T$ can be tested jointly with some kind of stationarity assumption.

References

[1] Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, Journal of econometrics 68, 5-28.

[2] Ahn, S.C., Y.H. Lee, and P. Schmidt, 2001, GMM estimation of linear panel data models with time-varying individual effects, Journal of econometrics 101, 219-255.

[3] Ahn, S.C., Y.H. Lee, and P. Schmidt, 2006, Panel data models with multiple time-varying individual effects, unpublished manuscript, Arizona State University.

[4] Anderson, T.W., and C. Hsiao, 1981, Estimation of dynamic models with error components, Journal of the American statistical association 76, 598-606.

[5] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of econometrics 18, 47-82.

[6] Andrews, D.W.K., 1999, Consistent moment selection procedures for generalized method of moments estimation, Econometrica 67, 543-564.

[7] Andrews, D.W.K., and B. Lu, 2001, Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models, Journal of econometrics 101, 123-164.

[8] Andrews, D.W.K., 2005, Cross-section regression with common shocks, Econometrica 73.

[9] Arellano, M., and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of economic studies 58, 277-297.

[10] Bai, J., 2005, Panel data models with interactive fixed effects, unpublished manuscript, New York University.

[11] Balestra, P., and M. Nerlove, 1966, Pooling cross-section and time series data in the estimation of a dynamic model: the demand for natural gas, Econometrica 34, 585-612.

[12] Bhargava A., 1991, Identification and panel data models with endogenous regressors, Review of economic studies 58, 129-140.

[13] Breusch, T.S., G.E. Mizon, and P. Schmidt, 1989, Efficient estimation using panel data, Econometrica 57, 695-701.

[14] Chamberlain, G., 1980, Analysis of covariance with qualitative data, Review of economic studies 47, 225-238.

[15] Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of econometrics 18, 5-46.

[16] Chamberlain, G., 1983, Panel Data, Ch. 22 in Handbook of econometrics, Vol. 2, Z. Griliches and M. Intrilligator (eds.), North-Holland, Elsevier, Amsterdam.

[17] Chamberlain, G., 1992, Comment: sequential moment restrictions in panel data, Journal of business and economic statistics 10, 20-26.

[18] Conley, T.G., 1999, GMM estimation with cross sectional dependence, Journal of econometrics 92, 1-45.

[19] Cragg, J.G., and S.G. Donald, 1997, Inferring the rank of a matrix, Journal of econometrics 76, 223-250.

[20] Hausman, J.A., and W.E. Taylor, 1981, Panel data and unobservable individual effects, Econometrica, 49, 1377-1398.

[21] Holtz-Eakin, D., W.K. Newey, and H. Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371-1395.

[22] Honoré, B., Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects, Econometrica 60, 533-565.

[23] Horowitz, J.L., and M. Markatou, 1996, Semiparametric estimation of regression models for panel data. Review of economic studies 63, 145-168.

[24] Kiefer, N.M., 1980, A time series-cross section model with fixed effects with an intertemporal factor structure. unpublished manuscript, Cornell University.

[25] Kruiniger, H., 2002, On the estimation of panel regression models with fixed effects, Working Paper no. 450, Queen Mary, University of London, also available at http://www.qmw.ac.uk/~ugte185/

[26] Kruiniger, H., 2008, A new method for estimating factor models, unpublished manuscript.

[27] Li, Q., and T. Stengos, 1996, Semiparametric estimation of partially linear panel data models, Journal of econometrics 71, 389-397.

[28] Mankiw, N.G., D. Romer, and D. Weil, 1992, A contribution to the empirics of economic growth, Quarterly journal of economics 107, 407-437.

[29] Mundlak, Y., 1961, Empirical production function free of management bias, Journal of farm economics 43, 44-56.

[30] Mundlak, Y., 1978, On the pooling of time series and cross section data, Econometrica 46, 69-85.

[31] Nijman, T., and M. Verbeek, 1992, Nonresponse in panel data: the impact on estimates of a life cycle consumption function, Journal of applied econometrics 7, 243-257.

[32] Pesaran, M.H., 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure, Econometrica 74, 967-1012.

[33] Wallace, T., and A. Hussain, 1969, The use of error components models in combining cross section with time series data, Econometrica 37, 55-68.

[34] Wooldridge, J.M., 1999, Distribution-free estimation of some nonlinear panel data models, Journal of econometrics 90, 77-97.