Department of Economics

Detection of Structural Breaks in Linear Dynamic Panel Data Models

Stefan de Wachter and Elias Tzavalis

Working Paper No. 505 February 2004 ISSN 1473-0278

Queen Mary
University of London

Detection of structural breaks in linear dynamic panel data models

Stefan De Wachter
Queen Mary, University of London
s.dewachter@qmul.ac.uk

Elias Tzavalis
Queen Mary, University of London
e.tzavalis@qmul.ac.uk

January 20, 2004

Abstract

This paper develops a break detection procedure for the well-known AR(p) linear panel data model with exogenous or pre-determined regressors. The test allows for a structural break in the slope parameters as well as in the fixed effects. Breaks in the latter are not constrained by any type of cross-sectional homogeneity and are allowed to be correlated with all past information.
Keywords: Panel data, structural break, break detection
JEL-classification: C23

1 Introduction

With the increasing availability of longitudinal datasets, panel data methods have become a popular tool in applied econometric research. By comparing individuals across both the cross-sectional and the intertemporal dimension, more powerful inferences can be made than with a single cross-section or a single time-series. The most commonly used linear regression models for panel data impose fairly stringent additional structure on the data by assuming that slope coefficients are constant across individuals and over time, and that unobserved variables are constant over time.

In this paper, we develop a method to test the validity of the latter two assumptions. More particularly, we construct a standard significance test for the null hypothesis that unobserved variables are constant across time and that slope coefficients are constant across both dimensions against the alternative that there is a single "structural break" at a possibly unknown point in time in either or both components. Our method can serve as a diagnostic procedure as well as a break detection device. The method is designed for use with the type of short panels often encountered in practical work.

Our motivation for studying breaks in slope parameters and individual intercepts is twofold. Firstly, a commonly cited virtue of panel data models is their ability to identify structural parameters that are difficult to estimate using a single cross-section because of the presence of unobserved heterogeneity. Identification of the parameters of interest using panel data is based on the assumption that the unobserved variables are constant over time and can therefore be "differenced away". In many practical applications this assumption is reasonable[1], at least in the sense that these variables' fluctuations over time are very small in comparison with other sources of variability. However, when the system is subject to a significant shock (as e.g. in the examples cited below), the situation may be less clear-cut.

Secondly, one can think of a panel dataset as a concatenation of time series. In the time series literature there is a large and expanding literature on detection of structural breaks, starting with the well-known tests of Chow (1960), Quandt (1960) and Brown, Durbin, and Evans (1975) through more recent contributions by Andrews (1993), Bai (1994), Bai (1997), Bai and Perron (1998) and Perron (1989), to name just a few. Although panel datasets typically span a much shorter time period than time series data there is no reason why the former could not be subject to a structural break, nor is the limited time horizon an impediment to consistent parameter estimation or detection of the breakpoint[2]. In the time-series literature, extensive evidence has emerged that allowing for intercept breaks can alter the conclusions of unit root tests and reduce the magnitude of estimates of persistence; the same phenomenon may be present in the panel data context.

Examples of the kind of effects we have in mind are not difficult to find. When considering a model of firm-level output or investment, trade liberalization may significantly affect unobservable characteristics such as management practices, strategy, marketing methods etc. Both the direction and magnitude of the effect will typically differ very significantly across firms (by market, national vs international competitiveness, reliance on protection etc). Financial reforms such as the UK consumer credit reform in 1997 may affect consumption / saving decisions of different agents in different ways (e.g. depending on their reliance on consumer credit and risk aversion). In a model explaining dividend payout ratios, changing dividend taxation legislation will have an impact on companies' decision making regarding the optimal mix between buy-backs, dividend payouts and internal investment. This impact may be dependent on how different types of shareholders are affected by the legislation and hence vary across companies.

An important theme in the above examples is the differential impact exogenous shocks can have across the cross-section[3]. In the light of important changes in their environment, agents may be forced to reoptimize very costly decisions, thereby changing some or all (to the econometrician) unobservable characteristics. Since these characteristics were not constant across units before the break, there is no

[1] For instance, when estimating a wage equation the unobserved variable could capture the effect of "ability", which can reasonably be assumed to be time-invariant.

[2] The limited time-horizon does, however, justify our decision to restrict our attention to the case in which only a single break occurs.

[3] It is important to note that such effects cannot be captured by the inclusion of a period–specific dummy variable.

reason to believe that they will change by the same amount in response to a major shock. Additionally, the new decisions are likely to be related to the pre-break situation and history[4].

Given the importance of structural changes, both in terms of fundamental assumptions regarding identification in panel data models and the attention they have received in the time-series literature, there is a comparative lack of methods for their detection in a panel data context. The closest related to our work in practical terms is Andrews and Lu (2001), who consider a model selection approach to the break detection problem based on an extention of Andrews (1999). They do not allow for the type of breaks in unobserved individual effects we focus on, but their approach can be extended to include this feature using the ideas presented in Section 3 below. Holtz-Eakin, Newey, and Rosen (1988) and Ahn, Lee, and Schmidt (2001) discuss estimation of a number of models in which individual effects fluctuate through time following a *common* temporal pattern. When following such an approach, multiple changes in fixed effects may occur without jeopardizing identification. However, in the context of the examples cited above, our assumption that fixed effects can change independently for different cross-sectional units is more appropriate. Additionally it is consistent with the standard interpretation of fixed effects as agents' decision variables or modifiable characteristics.

This paper is organized as follows. In Section 2 we define the problem precisely. Section 3 develops the break detection procedure based on the GMM estimator of Arellano and Bond (1991). Section 4 examines the finite-sample behaviour of our test in a Monte Carlo experiment. Section 5 contains an illustrative application and Section 6 concludes.

2 Problem description and assumptions

Our test aims to detect structural breaks in the familiar linear $AR(p)$ regression model with exogenous or pre-determined regressors for a panel with N cross-sectional units and T time-periods. That is, under the null hypothesis that no breaks occur we postulate the model

$$y_{i,t} = \eta_i + \sum_{l=1}^{p} \rho_l y_{i,t-l} + \beta' x_{i,t} + v_t + \varepsilon_{i,t}$$

The model contains k time-varying[5] regressors $x_{i,t}$. The error term has the usual error component structure consisting of an individual-specific "fixed or random[6] effect" η_i which is typically correlated with some or all regressors, a time-period-specific component v_t and a temporally and cross-sectionally varying noise term $\varepsilon_{i,t}$.

[4] One may argue that this also holds for regression parameters. However, in this paper we will not address this more difficult problem and consider the case where slope parameters can change over time without abandoning cross-sectional homogeneity. In doing so, we keep the structure of the "post-break model" identical to that of the "pre-break model".

[5] Coefficients on regressors that do not vary over time are not identifiable by the estimation procedure we will use.

[6] In what follows, we treat η as a random variable.

The latter is assumed to be uncorrelated across time, cross-sectionally independent with bounded fourth moments and uncorrelated with the other random variables in the model. The initial conditions $y_{i,1},...,y_{i,p}$ may be correlated with η_i but not with the idiosyncratic noise $\varepsilon_{i,t}$. These assumptions on the process and error structure are sometimes referred to as the "standard assumptions" - see e.g. Ahn and Schmidt (1995).

Under the alternative hypothesis, a single structural break may occur at some (possibly unknown) time-point τ, $p+2 \leq \tau \leq T$, from when onwards slope parameters and fixed effects come into place:

$$
\begin{aligned}
y_{i,t} &= \eta_i + \sum_{l=1}^{p} \rho_l y_{i,t-l} + \beta' x_{i,t} + v_t + \varepsilon_{i,t} \qquad t < \tau && (1) \\
y_{i,t} &= (\eta_i + \delta_{i,\tau}) + \sum_{l=1}^{p} (\rho_l + \omega_{\tau l})\, y_{i,t-l} + (\beta + \gamma_\tau)' x_{i,t} + v_t + \varepsilon_{i,t} \quad t \geq \tau && (2)
\end{aligned}
$$

The (scalar) parameters $\omega_{\tau l}$ $(l = 1,...,p)$ and the vector γ_τ denote the changes in the autocorrelation and slope parameters, respectively. The change $\delta_{i,\tau}$ in the fixed effect is not restricted to be identical across individuals and may be correlated with η_i and past idiosyncratic shocks $\varepsilon_{i,t}$ $(t < \tau)$. No assumptions are made on the random variables η and δ_τ apart from existence of second moments (including joint moments with the initial conditions). Any kind of dependence structure between δ_τ and past $\varepsilon_{i,t}$ and $x_{i,t}$ and initial conditions $y_{i,1},...,y_{i,p}$ is allowed. Summarizing, we have the following assumptions regarding the fixed effects break:

$$
\begin{aligned}
E(\delta_{i,\tau}\varepsilon_{i,s}) &= /\neq 0 \quad (s < \tau); \quad = 0 \quad (s \geq \tau) \\
E(\delta_{i,\tau}\eta_i) &= /\neq 0 \\
E(\delta_{i,\tau}y_{i,\pi}) &= /\neq 0 \quad (\pi = 1,...,p)
\end{aligned}
$$

As mentioned in the Introduction, these assumptions regarding the changes in the fixed effects differ from those of Holtz-Eakin, Newey, and Rosen (1988) and Ahn, Lee, and Schmidt (2001), who allow for changes in every time-period but impose a common temporal pattern using the parametrization $\eta_{i,t} = \theta_t \mu_i$.

Our aim is to detect the breakpoint τ and to derive consistent estimators for all slope parameters (including their changes). We do this in the well-known GMM framework used by Arellano and Bond (1991), amongs others. Asymptotics are in the N-direction with T considered fixed. This implies that the individual intercepts η_i and their changes $\delta_{i,\tau}$ can not be treated as consistently estimable parameters. The assumptions on the error structure and initial conditions made above are the generalization of those of Arellano and Bond (1991) to the specification (1) - (2) and are in that sense minimal. We wish to refrain from making additional assumptions on the structure of the model in order to avoid spuriously detecting breaks. For instance, Arellano and Bover (1995) show how imposing mean-stationarity leads to an improved estimator. In our context this is not a natural assumption since our aim is to detect a particular type of departure from stationarity.

We remark that the assumptions on the error structure are too general to allow identification without a restriction on the number of breakpoints. By assuming the presence of at most a single breakpoint we can circumvent the problem. The fact that the breakpoint is unknown is dealt with by an iterative procedure, as explained below.

3 Break detection

3.1 Outline

Break detection in the set-up of model (1) - (2) follows an analogous procedure to that used in some of the time-series literature (e.g. Andrews (1993), Zivot and Andrews (1992)) and in the panel data literature (Tzavalis (2002)). We construct a standard significance test where the null hypothesis contains no break and the alternative allows for a single break. This break-point is not known (but we also cover the case where it is known); hence one can look at the alternative hypothesis as a number of models of the form (1) - (2), each associated with a break at some point $\tau = p + 2, ..., T$.

For each potential break-point, the parameters of the model (1) - (2) is estimated using GMM. In order to obtain a consistent estimator for a model of the form (1) - (2) with given τ, the standard moment conditions will obviously need to be adapted, except in a few special cases to be discussed below. The difference in the value of the GMM objective functions under the null and each of the alternatives will be the basis for the test-statistic. The point τ for which this difference is the largest is the candidate for the break-point[7]. The asymptotic distribution of the largest of these $T - (p + 1)$ differences is not the familiar chi-square distribution that often features in GMM specification testing, but that of the *maximum* of $T - (p + 1)$ (correlated) chi-squares[8]. The resulting confidence intervals will be wider than those for the case of a known breakpoint, reflecting the search over a number of possibilities for the "worst evidence" against the null.

To avoid obscuring the ideas through notational clutter associated with the general setup in (1) - (2), we first explain the details of our test in the simplified setting of the dynamic $AR(1)$ panel data model without regressors:

$$
\begin{aligned}
y_{i,t} &= \eta_i + \rho y_{i,t-1} + \varepsilon_{i,t} \qquad t < \tau && (3)\\
y_{i,t} &= (\eta_i + \delta_{i,\tau}) + (\rho + \omega_\tau)\, y_{i,t-1} + \varepsilon_{i,t} \qquad t \geq \tau && (4)
\end{aligned}
$$

As before, a structural break may occur at (unknown) time τ $(3 \leq \tau \leq T)$ either in the individual-specific intercepts η_i or in the autoregressive parameter ρ (or both). Since one can always include time dummies, we assume that $E(\eta_i) = E(\delta_{i\tau}) = 0$ $\forall i, \tau$. The other features of the model remain as before.

[7] We do not claim that this criterion leads to the optimal test in any conventional sense of the word.

[8] Because we use cross-sectional asymptotics, this result is much easier to establish than in related work in the time-series literature.

3.2 Moment functions for the $AR(1)$ model

In order to construct the GMM estimator for a model of the form (3) - (4) with fixed τ, we examine how the moment functions for the Arellano-Bond estimator change in the presence of a structural break. These moment functions are derived from the insight that past levels are valid instruments for differenced equations if no break occurs, or more generally that

$$E\left[y_{i,s}\Delta\varepsilon_{i,t}(\theta)\right] = 0 \quad t \geq 3;\ s \leq t-2 \tag{5}$$

with $\theta = \rho$. Expression (5) is still valid in the presence of a structural break, but as we now show, the corresponding moment functions differ and θ may contain extra parameters related to the break.

It is useful to introduce some additional notation for the moment functions at this point. One can order the moment conditions (5) by t and then by s to get a $(T-2)(T-1)/2$ vector $m_2(\theta_2)$ with components $m_2^{(j)}(\theta_2)$, $j = 1, ..., (T-2)(T-1)/2$

$$m_2(\theta_2) = \begin{bmatrix} E\left[y_{i,1}\left(\Delta y_{i,3} - \rho\Delta y_{i,2}\right)\right] \\ E\left[y_{i,1}\left(\Delta y_{i,4} - \rho\Delta y_{i,3}\right)\right] \\ E\left[y_{i,2}\left(\Delta y_{i,4} - \rho\Delta y_{i,3}\right)\right] \\ \vdots \\ E\left[y_{i,T-2}\left(\Delta y_{i,T} - \rho\Delta y_{i,T-1}\right)\right] \end{bmatrix} \tag{6}$$

The subscript-2 of the moment functions and the parameter vector refers to "break at $t = 2$", which is equivalent to "no break" because both the autoregressive parameter and the fixed effects only "start existing" at $t = 2$. Since the only parameter to be estimated under the null is ρ, one has $\theta_2 = \rho$. When a second subscript is added, this refers to an observation on a particular individual i, e.g. $m_{2i}(\rho) = \left[y_{i,1}\left(\Delta y_{i,3} - \rho\Delta y_{i,2}\right) \cdots y_{i,T-2}\left(\Delta y_{i,T} - \rho\Delta y_{i,T-1}\right)\right]'$. So $m_2(\rho) = E\left(m_{2i}(\rho)\right)$. The moment covariance matrix is then defined as $\Phi(\rho) = E\left[m_{2i}(\rho)\, m_{2i}(\rho)'\right]$.

Using this notation, we now examine what $m_3(\theta_3)$ (the vector of moment functions when a break occurs at $\tau = 3$) will look like. As mentioned before, the statements (5) are still valid in these situations, but the actual moment functions derived from them are somewhat different. First consider the first component (the moment condition based on the differenced 3^{rd} equation instrumented by y_{i1}).

$$\begin{aligned} E\left[y_{i,1}\Delta\varepsilon_{i,3}\right] &= E\left[y_{i,1}\left(\left\{y_{i,3} - (\eta_i + \delta_{i,3}) - (\rho + \omega_3)\, y_{i,2}\right\} - \left\{y_{i,2} - \eta_i - \rho y_{i,1}\right\}\right)\right] \\ &= E\left[y_{i,1}\left(\Delta y_{i,3} - \rho\Delta y_{i,2} - \delta_{i,3} - \omega_3 y_{i,2}\right)\right] \\ &= E\left[y_{i,1}\Delta y_{i,3}\right] - \rho E\left[y_{i,1}\Delta y_{i,2}\right] - E\left[y_{i,1}\delta_{i,3}\right] - \omega_3 E\left[y_{i,1}y_{i,2}\right] \end{aligned}$$

The two final terms in this expression did not show up in the corresponding moment functions derived under H_0. If there is a break at $t = 3$, it is possible that $E\left[y_{i,1}\delta_{i,3}\right] \equiv \sigma_{y_1\delta_3} \neq 0$. Additionally, ω_3 need not be equal to zero. Hence the first moment condition is

$$m_3^{(1)}(\theta_3) = E\left[y_{i,1}\Delta y_{i,3}\right] - \rho E\left[y_{i,1}\Delta y_{i,2}\right] - \sigma_{y_1\delta_3} - \omega_3 E\left[y_{i,1}y_{i,2}\right] = 0 \tag{7}$$

where $\theta_3 = \left[\rho\ \omega_3\ \sigma_{y_1\delta_3}\right]'$. As a prelude to Remark 1 below, we notice that if a break occurs at $\tau = 3$ with the properties that either the slope coefficient changes or the individual effect undergoes a shift that is correlated with the initial conditions, the standard Arellano-Bond estimator will be inconsistent. However, if there is no slope-break and the intercept shifts are uncorrelated with the initial conditions, i.e. $\sigma_{y_1\delta_3} = \omega_3 = 0$, there is no inconsistency.

The other moment conditions $m_3^{(j)}$; $j \geq 2$[9] are similar to the familiar Arellano-Bond expressions, but with an autoregressive parameter $\rho + \omega_3$:

$$m_3^{(j)}(\theta_3) = E\left[y_{i,s}\Delta y_{i,t}\right] - (\rho + \omega_3)\, E\left[y_{i,s}\Delta y_{i,t-1}\right] = 0 \quad t \geq 4,\ \ s \leq t-2 \tag{8}$$

If the objective were to test for the presence of a break at $t = 3$, one could now simply construct a Likelihood Ratio (Distance Metric) type statistic based on estimation using the standard Arellano-Bond moment conditions on the one hand and those in expression (8) on the other, discarding the moment function (7). If one uses the efficient GMM weighting matrix to estimate the parameters under both hypotheses then this LR statistic has a $\chi^2(1)$ distribution under the null, as shown below.

When testing for a break at $t = \tau$, $\tau > 3$, one can follow the same procedure based on the following general expressions for the moment functions m_τ:

1. All functions $m_\tau^{(j)}(\theta_\tau)$ where $j \leq (\tau-3)(\tau-2)/2$ (that is, derived from equations *before* the breakpoint) are like the components of the vector in (6).

2. All functions $m_\tau^{(j)}(\theta_\tau)$ with $j > (\tau-2)(\tau-1)/2$ (derived from equations after the breakpoint) are of the form (8) (with appropriate timings).

3. Moment functions derived from the differenced equation for $t = \tau$ (i.e. $m_\tau^{(j)}(\theta_\tau)$ where $(\tau-3)(\tau-2)/2 < j \leq (\tau-2)(\tau-1)/2$) are of the form

$$\begin{aligned} m_\tau^{(j)}(\theta_\tau) &= E\left[y_{i,s}\Delta y_{i,\tau}\right] - \rho E\left[y_{i,s}\Delta y_{i,\tau-1}\right] - \begin{pmatrix} \frac{1-\rho^{s-1}}{1-\rho}\sigma_{\eta\delta_\tau} + \\ \rho^{s-1}\sigma_{y_1\delta_\tau} + \sum_{k=0}^{s-2}\rho^k\sigma_{\varepsilon_{s-k}\delta_\tau} \end{pmatrix} \\ & \quad -\omega_\tau E\left[y_{i,s}y_{i,\tau-1}\right] \end{aligned} \tag{9}$$

   where $\sigma_{\eta\delta_\tau} := E\left[\eta\delta_\tau\right]$, $\sigma_{y_1\delta_\tau} := E\left[y_1\delta_\tau\right]$, $\sigma_{\varepsilon_{s-k}\delta_\tau} := E\left[\varepsilon_{s-k}\delta_\tau\right]$.

Expression (9) is easily derived by realizing that for any $s < \tau$ one has $y_{is} = \eta_i\left(\sum_{k=0}^{s-2}\rho^k\right) + \rho^{s-1}y_{i1} + \sum_{k=0}^{s-2}\rho^k\varepsilon_{i,s-k}$ because there is no break before τ. Consequently

$$E\left[y_{i,s}\delta_{i,\tau}\right] = \left(\sum_{k=0}^{s-2}\rho^k\right) E\left[\eta_i\delta_{i,\tau}\right] + \rho^{s-1}E\left[y_{i,1}\delta_{i,\tau}\right] + \sum_{k=0}^{s-2}\rho^k E\left[\varepsilon_{s-k}\delta_\tau\right]$$

Plugging this expression in equation (5) yields (9).

Only the moment functions of the form (9) contain the parameters related to intercept breaks; the number of these parameters is greater than the number of moment conditions. Consequently, the moment functions derived from the differenced

[9] We do not make the relationship between j and t and s explicit; j is just an index that indicates the position of the moment function in the vector m_3 and is only used to distinguish between the vector m and its components.

equation for $t = \tau$ will equal zero in sample for any value of the slope parameters. When estimating the model under the hypothesis that a break occurs at τ, one can therefore disregard the moment conditions (9) and construct a GMM estimator using only the remaining moment functions.

Remark 1 *It is clear from the above derivation of the moment functions that the standard Arellano-Bond estimator will be inconsistent when a structural break is present.*
The only exception to this is the case where there is no slope break and the individual intercept breaks are uncorrelated with the initial fixed effects, initial conditions and past errors. Indeed, as can be seen from comparing the expressions for m_2 *and* m_τ $(\tau > 2)$*, the breaks in the fixed effects will essentially be detected by looking at the covariance of the break-size* δ_τ *with the original fixed effect* η*, the initial conditions* y_1 *and past idiosyncratic errors. No parameter capturing* $E(\delta_\tau^2)$ *is present. This means that breaks in the fixed effects that are not correlated with either* η*,* ε *or* y_1 *will not be detected.*
The intuition for this is as follows. The moment conditions we have used relate to differenced data. When there is a break, some or all of the fixed effects will follow a step function (of time). Their differences (looked at as a function of time) are zero except for a "spike" at the break-point. When the break is independent of all other variables and the model is estimated under the null of no break, this spike is indistinguishable from an increase in the variance of $\Delta\varepsilon_\tau$*. Since the Arellano-Bond moment conditions do not impose homoskedasticity, such an increase in variance is not at odds with the assumptions under the null.*

Remark 2 *As a corrolary to the previous remark, one sees that the standard Arellano-Bond estimator is robust to uncorrelated breaks in fixed effects, regardless of whether these occur more than once.*

Remark 3 *When correlated fixed effects breaks and/or slope breaks occur, the above procedure provides consistent parameter estimates when the breakpoint is known.*

Remark 4 *For certain* τ*, not all components of the vector* θ_τ *are identified under the alternative. For instance in the case of* $\tau = 3$*, it is clear from expressions (7) - (8) that* $\sigma_{y_1\delta_3}$ *and* ω_3 *are not jointly identified. Hence the "size" of the model to be estimated under the alternative that a break occurs at* τ *will depend on* τ *and the dimension of the parameter* θ_τ *is adjusted accordingly.*

3.3 Test-statistic

For any given candidate breakpoint τ, the slope and slope break parameters ρ and ω_τ can be estimated by GMM. To construct the estimator, the moment functions based on the differenced equation at the breakpoint are discarded. The same procedure can be followed to obtain estimates for ρ under the null that no break occurs. Given a consistent estimate of the efficient GMM weighting matrix and estimated sample moments vectors $\widehat{m}_\tau(\widehat{\theta}_\tau)$[10] evaluated at the estimated parameters $\widehat{\theta}_\tau$ for each τ, one can construct a vector of Likelihood Ratio statistics.

Some more notation is needed at this point. Let $Q_\tau(\theta_\tau) = m_\tau(\theta_\tau)'\Phi_\tau(\theta_\tau^o)^{-1}m_\tau(\theta_\tau)$ denote the objective function for the model with a break at time τ evaluated at parameter value θ_τ, where in this case the population moments and the population covariance matrix $\Phi_\tau(\theta_\tau^o) = E\left[m_{\tau i}(\theta_\tau^o)\, m_{\tau i}(\theta_\tau^o)'\right]$ evaluated at the true parameter value θ_τ^o are used. The sample counterpart is $\widehat{Q}_\tau(\widehat{\theta}_\tau) = \widehat{m}_\tau(\widehat{\theta}_\tau)'\widehat{\Phi}_\tau^{-1}\widehat{m}_\tau(\widehat{\theta}_\tau)$ where $\widehat{\Phi}_\tau = \frac{1}{n}\sum\left[m_{\tau i}(\widehat{\theta}_\tau)\, m_{\tau i}(\widehat{\theta}_\tau)'\right]$. Note that the dimension of this matrix is smaller under the alternative than under the null if the former contains intercept breaks. Note also that depending on the timing of the break, slope break parameters may not be identifiable[11].

Estimation of the model under the alternative for $\tau = 3, ..., T$ and under the null yields the $T-2$ vector $V(\widehat{\theta})$ of $LR-$statistics

$$V(\widehat{\theta}) = n\left[\widehat{Q}_2(\widehat{\theta}_2) - \widehat{Q}_3(\widehat{\theta}_3)\ \cdots\ \widehat{Q}_2(\widehat{\theta}_2) - \widehat{Q}_T(\widehat{\theta}_T)\right] \tag{10}$$

$\widehat{\theta}$ is the vector containing all parameter estimates. Denote its components by $V^{(k)}(\widehat{\theta})$. The test-statistic is the maximum of the scaled LR-statistics

$$V_{\max}(\widehat{\theta}) = \max_{k=1,\ldots,T-2} V^{(k)}(\widehat{\theta}).S(k) \tag{11}$$

and the candidate break-point is $\tau = k+2$ if the maximum is attained at component k. The scaling factor $S(k)$ is constructed such that each component $V^{(k)}(\widehat{\theta}).S(k)$ has the same marginal distribution.

Note that one could choose to not include all lagged y as instruments when estimating the model under the null. The above procedure still applies in this case as long as no instruments are used under the alternative that are not used under the null.

3.4 Asymptotic distribution

The derivation of the asymptotic distribution of both tests is straightforward and is therefore relegated to the Appendix. Here we present the results and a brief discussion. For the test with known break-point, the asymptotics are similar to the standard Sargan difference test. The only difference between our test and a Sargan difference test is that in our case the model with fewer moment conditions also contains more parameters. This does not affect asymptotic chi-squaredness of the test, however:

Theorem 1 *(known breakpoint) The test-statistic $n\left(\widehat{Q}_2(\widehat{\theta}_2) - \widehat{Q}_\tau(\widehat{\theta}_\tau)\right)$ is asymptotically chi-squared distributed with #(omitted moment conditions) + ($dim(\theta_\tau) - dim(\theta_2)$) degrees of freedom.*

[10] The "hat" on m indicates that the expectations are replaced by sample averages.
[11] In the basic AR(1) model we consider here, this is the case when $\tau = 3$ and $\tau = T$.

Proof. See Appendix. ■

The actual number of degrees of freedom depends on the model specification and how many moment conditions the user chooses to include for estimation of the model without a break.

The test with unknown breakpoint does not have a chi-square distribution but that of the maximum of correlated (scaled) chi-squared variables. The dependence structure is given by the matrices G_τ in Theorem 2. For precise formulae the reader is referred to the Appendix.

Theorem 2 *(unknown breakpoint) The statistic $V_{\max}(\widehat{\theta})$ is asymptotically distributed as $max(v'G_{\tau_{\min}}v, ..., v'G_T v)$ where $v \sim N(0, I_d)$ and d is the number of moment conditions used when estimating the parameters under the null of no break. The idempotent matrices G_τ depend on the parameters but can be consistently estimated.*

Proof. See Appendix. ■

In practice, quantiles of the asymptotic distribution need to be computed by simulation. Each replication involves the generation of $T - \tau_{\min} + 1$ normal variates and the computation of the quadratic forms $v'G_\tau v$.

To derive this asymptotic distribution, only standard arguments are needed: the weak convergence of the scaled vector $V(\widehat{\theta})$ in (10) to its limit is a simple case of multivariate convergence in distribution. This is a consequence of our decision to keep T fixed in the approximation procedure. In contrast, in related work in time-series analysis the convergence is to a process with index set the positive real line because asymptotics are in the time-direction. Ensuring weak converge then becomes much harder, and in particular requires establishing asymptotic equicontinuity of the normalized score process - see Andrews and Ploberger (1994). In our case the limiting stochastic process $V^{(k)}(\widehat{\theta}).S(k)$ (with index k) has a finite index set.

3.5 Generalization to the AR(p) model with exogenous regressors

The generalization of the procedure to the full model described in expressions (1) - (2) is straightforward and does not contain any new insights. In fact, the test procedure and the asymptotic distribution theory are identical. It only remains to derive the exact expressions for the Arellano-Bond moment functions in the presence of a structural break. As before, they are derived from the expressions

$$E\left[y_{i,s}\Delta\varepsilon_{i,t}\right] = 0 \quad t \geq p+2; \; s \leq t-2 \tag{12}$$

$$E\left[x_{i,s}\Delta\varepsilon_{i,t}\right] = 0 \quad t \geq p+2 \tag{13}$$

In (13), the feasible values of s are not explicitly stated: which lags and leads of the regressors can be used as instruments depends on their "exogeneity status". Their treatment is identical to the estimation procedure for the AR(p) panel data model without structural breaks - see Arellano and Bond (1991) for details. For clarity, we split the discussion in two parts.

Moments based on (13) The moments for $t < \tau$ and $t > \tau$ are standard (in the latter case with the post-break parameters). The only issue arises for the moments for the break-point equation:

$$\begin{aligned}
E(x_{i,s}\Delta\varepsilon_{i,\tau}) &= E\left[x_{is}\left(\begin{array}{c} \left(y_{i,\tau}-(\eta_i+\delta_{i\tau})-\sum_{l=1}^{p}(\rho_l+\omega_{\tau l})\,y_{i,\tau-l}-(\beta+\gamma_\tau)'\,x_{i,\tau}-\varepsilon_{i,\tau}\right)- \\ \left(y_{i,\tau-1}-\eta_i+\sum_{l=1}^{p}\rho_l y_{i,\tau-1-l}+\beta' x_{i,\tau}+\varepsilon_{i,\tau}\right)\end{array}\right)\right] \\
&= E\left(x_{i,s}\Delta y_{i,\tau}\right)-\sum_{l=1}^{p}\rho_l E\left(x_{i,s}\Delta y_{i,\tau-l}\right)-\beta' E\left(x_{i,s}\Delta x_{i,\tau}\right) \\
&\quad -E\left(x_{i,s}\delta_{i,\tau}\right)-\sum_{l=1}^{p}\omega_{\tau l}E\left(x_{i,s}\,y_{i,\tau-l}\right)-\gamma_\tau' E\left(x_{i,s}x_{i,\tau}\right) \qquad (14)
\end{aligned}$$

Multiplication of two vectors of the same dimension is interpreted as component-by-component multiplication.

The 3 last terms are related to the structural break. In particular, $E\left(x_{i,s}\delta_{i,\tau}\right)$ is an extra parameter, call it $\sigma_{x_s\delta_\tau}$. The user may specify that $E\left(x_{i,s}\delta_{i,\tau}\right)=0$, but in general this will not be the case; hence moment conditions (14) are not used for estimation under the alternative. This shows that regressors can contribute to break detection to the extent that they are related to the break.

Moments based on (12) These moments are again standard for $t < \tau$ and $t > \tau$. For the break-point equation we have as above

$$\begin{aligned}
E(y_{i,s}\Delta\varepsilon_{i,\tau}) &= E\left(y_{i,s}\Delta y_{i,\tau}\right)-\sum_{l=1}^{p}\rho_l E\left(y_{i,s}\Delta y_{i,\tau-l}\right)-\beta' E\left(y_{i,s}\Delta x_{i,\tau}\right) \\
&\quad -E\left(y_{i,s}\delta_{i,\tau}\right)-\sum_{l=1}^{p}\omega_{\tau l}E\left(y_{i,s}\,y_{i,\tau-1-l}\right)-\gamma_\tau' E\left(y_{i,s}x_{i,\tau}\right) \qquad (15)
\end{aligned}$$

Now the term $E\left(y_{i,s}\delta_{i,\tau}\right)$ is somewhat problematic as it needs to be expressed in terms of the initial conditions (as in the simple $AR(1)$ case); the only difference is that the recursion is not as simple as in the $AR(1)$ case. Transform the system in (1) into a first-order p-dimensional system (the so-called companion form). Define $\overline{y}_{i,t}=[y_{i,t}...y_{i,t-p+1}]$. (1) is then the first component in the system of equations

$$\begin{aligned}
\overline{y}_{i,t} &= e_1\eta_i+\begin{bmatrix}\rho_1 & \rho_2 & \cdots & \rho_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & \cdots & 1 & 0\end{bmatrix}\overline{y}_{i,t-1}+\begin{bmatrix}\beta_1 & \beta_2 & \cdots & \beta_k \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0\end{bmatrix}x_{i,t}+e_1\varepsilon_{i,t} \\
&= e_1\eta_i+\Pi\,\overline{y}_{i,t-1}+B\,x_{i,t}+e_1\varepsilon_{i,t}
\end{aligned}$$

where $e_1=(1\ 0\ ...\ 0)'$. Now recursions are easy again: one gets

$$\overline{y}_{i,t}=\left(\sum_{j=0}^{t-p+1}\Pi^j\right)e_1\eta_i+\Pi^{t-p}\overline{y}_{i,p}+\sum_{j=0}^{t-p+1}\Pi^j Bx_{i,t-j}+\sum_{j=0}^{t-p+1}\Pi^j e_1\varepsilon_{i,t-j}$$

This is not easily expressed in terms of the original parameters, but the term $E\left(y_{is}\delta_{i\tau}\right)$ can be written as

$$E\left(y_{i,s}\delta_{i,\tau}\right)=\left(\sum_{j=0}^{t-p+1}e_1'\Pi^j e_1\right)\sigma_{\eta\delta_\tau}+e_1'\Pi^{t-p}\sigma_{\overline{y}_p\delta_\tau}+\sum_{j=0}^{t-p+1}e_1'\Pi^j B\sigma_{x_{t-j}\delta_\tau}+\sum_{j=0}^{t-p+1}e_1'\Pi^j e_1\sigma_{\varepsilon_{t-j}\delta_\tau}$$

Here $\sigma_{\eta\delta_\tau}$ and $\sigma_{\varepsilon_{t-j}\delta_\tau}$ are scalar parameters, and $\sigma_{\overline{y}_p\delta_\tau}$ and $\sigma_{x_{t-j}\delta_\tau}$ are p and k vectors of parameters, respectively. As in the simple case, moment conditions (15) contain too many new parameters and are not used for estimation under the alternative. Hence the testing procedure works exactly as in the simple $AR(1)$ case.

4 Finite-sample properties

We now discuss the results of a small Monte Carlo experiment. The set-up is as follows. Data are generated from a DGP described by

$$\begin{aligned} y_{i,t} &= \eta_i^*(1-\rho)+\rho y_{i,t-1}+\varepsilon_{i,t} \qquad t<\tau && (16)\\ y_{i,t} &= \left(\eta_i^*+\delta_{i,\tau}^*\right)(1-\rho-\omega_\tau)+(\rho+\omega_\tau)\,y_{i,t-1}+\varepsilon_{i,t} \qquad t\geq\tau && (17)\end{aligned}$$

where ε, η_i^*, $\delta_{i\tau}^*$ and the initial conditions y_{i1} are jointly normally distributed. The difference with our notation in (3) - (4) is that the fixed effects are now denoted by $\eta_i^*(1-\rho)$ and $(\eta_i^*+\delta_{i\tau}^*)(1-\rho-\omega_\tau)$. η_i^* and $\eta_i^*+\delta_{i\tau}^*$ are interpreted as the mean of the stationary distribution of the relevant regime[12].

The initial conditions are normally distributed: $y_{i,1}=\eta_i^*+e_i$ with $e_i\sim N(0,\sigma_e^2/(1-\rho^2))$. Here σ_ε is the standard deviation of the idiosyncratic errors, which is fixed to 1 in all the experiments; the standard deviation σ_η of the pre-break fixed effects is set to 2 throughout. The ratio of these standard deviations affects the results significantly: when the noise in the data is large, the test loses power; when the noise is small compared to the variability of the fixed effects, the test behaves badly under the null but has larger power. However, these effects only change the conclusions below for rather unrealistic values (ratios >5 or <1/5).

The dependence of the fixed effects break $\delta_{i,\tau}^*$ on η_i^* and past ε was specified as follows: $Corr(\delta_{i,\tau}^*,\eta_i^*)=\rho_{\eta\delta}$, $Corr(\delta_{i,\tau}^*,\varepsilon_{i,t})=1_{\{t<\tau\}}\rho_{\eta\delta}.t^2/(\tau-1)^2$, i.e. the idiosyncratic shock the period before the break has the same correlation with the intercept break as the initial fixed effect; then the correlation decreases quadratically with the lag. The standard deviation of the fixed effects break is kept fixed throughout all experiments at $\sigma_\delta=0.4$, i.e. 20% of the pre-break fixed effect. To mimic common practice in applied work with the Arellano-Bond estimator, the maximum number of instruments for any estimating equation is fixed at 4. When $T=6$, this implies the full set of moment conditions.

In the experiments, we investigate the effect of the sample dimensions, the location of the breakpoint, the autocorrelation parameter and the size of its change, and of the correlation of the fixed effects break with the initial fixed effect and past idiosyncratic shocks. All computations are executed using programs written in the matrix language Ox version 3.30 (Doornik (1999)).

[12] This is a more easily interpretable quantity than the intercept itself.

		$\rho = 0.5$		$\rho = 0.8$	
		$\tau = 3$	$\tau = 5$	$\tau = 3$	$\tau = 5$
$N = 100; T = 6$	*full*	6/7	6	6/7	8
	red.	7/8	7	8/9	7
$N = 500; T = 6$	*full*	4/7	4	4/6	5
	red.	6/5	5	6/7	6
$N = 500; T = 12$	*full*	6/7	7	6/7	5
	red.	5/6	7	5/6	5

Table 1: actual size of the tests at the 5 percent level

4.1 Size

Rejection frequencies of the test for a range of experimental designs are reported in Table 1. The rows marked "full" contain results for the test for breaks in both fixed effects and slopes, whereas the one marked "red." (reduced) refers to the test for a break in the fixed effects only. Where the notation a/b is used, a is the rejection frequency for the test with known breakpoint and b that for unknown breakpoint. In some cells, the latter is suppressed to avoid duplication.

There seems to be only a very slight tendency to over-rejection. In a number of further experiments - not reported here - we did notice an increase in size as the ratio of the variability of the fixed effects to that of the idiosyncratic errors was increased furtherr.

Other studies on the GMM estimator for the AR1 panel data model have reported often substantial negative biases in the estimates of the autocorrelation parameter. When estimating the model under the alternative of a break, these biases do not disappear, and become more severe in some cases (details not reported here).

4.2 Power

We now study the power of the tests. Tables 2, 3 and 4 present the results of a series of experiments for 3 different sample sizes. Results are displayed in the form $a/b(c)$ where a represents the rejection frequency at the nominal 5% level of the test with known breakpoint, b that of the test with unknown breakpoint, and c the frequency of correct breakpoint detection[13].

For the small sample size ($N = 100$; $T = 6$) in Table 2, power at the nominal 5% level hardly exceeds the actual size for most cases. However, comparison with the results in Table 1 reveals that the test is not biased. Additionally, correct detection of the breakpoint happens in less than 50% where rejection occurs. The only exceptions are the cases with a positive (negative) slope break and positive (negative) fixed effect break correlation. Not allowing for a break in the autocorrelation parameter in the model (third and fourth lines from the bottom of the table) does not result in a noticeable increase in performance.

The low power of the test in small samples against alternatives without a slope break suggests that the Arellano-Bond estimator is relatively robust against insta-

[13] One has $c \leq b$ in general; if the test always finds the correct breakpoint when it rejects, $c = b$.

		$\rho = 0.5$		$\rho = 0.8$	
		$\tau = 3$	$\tau = 5$	$\tau = 3$	$\tau = 5$
$\omega = -0.1$	$\rho_{\eta\delta} = -0.5$	21/15(11)	12/11(6)	19/17(11)	9/11(5)
	$\rho_{\eta\delta} = 0.5$	7/7(2)	11/7(3)	9/12(4)	6/5(2)
$\omega = 0$	$\rho_{\eta\delta} = -0.5$	11/10(5)	8/6(2)	8/8(5)	8/6(2)
	$\rho_{\eta\delta} = 0.5$	9/7(1)	9/7(2)	5/7(2)	6/12(3)
$\omega = N\backslash A$	$\rho_{\eta\delta} = -0.5$	10/9(4)	8/10(5)	7/8(2)	7/9(3)
(no slope break)	$\rho_{\eta\delta} = 0.5$	8/6(2)	8/7(2)	6/8(2)	6/7(2)
$\omega = 0.1$	$\rho_{\eta\delta} = -0.5$	6/6(2)	6/7(2)	5/6(1)	8/9(1)
	$\rho_{\eta\delta} = 0.5$	14/10(4)	9/6(0)	5/7(1)	10/10(2)

Table 2: power for N=100; T=6

		$\rho = 0.5$		$\rho = 0.8$	
		$\tau = 3$	$\tau = 5$	$\tau = 3$	$\tau = 5$
$\omega = -0.1$	$\rho_{\eta\delta} = -0.5$	69/58(51)	52/40(33)	63/56(47)	28/22(18)
	$\rho_{\eta\delta} = 0.5$	18/11(5)	28/22(15)	20/17(11)	7/6(3)
$\omega = 0$	$\rho_{\eta\delta} = -0.5$	27/20(14)	17/20(12)	5/8(3)	5/8(4)
	$\rho_{\eta\delta} = 0.5$	36/21(12)	30/22(14)	9/6(2)	6/9(2)
$\omega = N\backslash A$	$\rho_{\eta\delta} = -0.5$	26/15(11)	23/21(15)	8/7(4)	6/8(3)
(no slope break)	$\rho_{\eta\delta} = 0.5$	40/22(14)	35/20(14)	8/7(1)	7/8(1)
$\omega = 0.1$	$\rho_{\eta\delta} = -0.5$	4/6(3)	16/9(4)	22/11(3)	9/11(2)
	$\rho_{\eta\delta} = 0.5$	62/42(29)	33/17(10)	28/15(3)	20/21(5)

Table 3: power for N=500; T=6

bility in the fixed effects, even if their changes are correlated with other variables in the model.

Power properties greatly improve with an increase in cross-sectional sample size, as illustrated in Table 3. When ω and $\rho_{\eta\delta}$ have the same sign, both tests perform well, and the test with unknown breakpoint functions as a good break detection device. Comparing these results with the case where $\omega = 0$, it seems that slope breaks are easier to detect than fixed effects breaks. This is true regardless of whether one incorporates knowledge that there is no slope break.

Generally, the results for $\rho = 0.8$ are worse than those for $\rho = 0.5$, especially in the absence of a slope break. This may be partly explained by the fact that the invalidity of the standard Arrelano-Bond moment conditions depends on the covariance of $\delta^*_{i\tau}(1 - \rho - \omega_\tau)$ with other variables in the model, and for a given magnitude of $\delta^*_{i\tau}$, the total fixed effect is smaller the larger ρ.

Increasing the time-series dimension of the data as well, the results improve further. Table 4 reports very large power and good detection accuracy in all cases except when ρ is large and no slope break is present.

		$\rho = 0.5$		$\rho = 0.8$	
		$\tau = 6$	$\tau = 10$	$\tau = 6$	$\tau = 10$
$\omega = -0.1$	$\rho_{\eta\delta} = -0.5$	90/84(66)	79/68(57)	74/64(47)	55/42(36)
	$\rho_{\eta\delta} = 0.5$	49/31(18)	76/56(39)	14/13(4)	6/9(1)
$\omega = 0$	$\rho_{\eta\delta} = -0.5$	41/28(17)	52/31(21)	7/9(2)	8/8(3)
	$\rho_{\eta\delta} = 0.5$	63/33(21)	73/46(32)	7/7(1)	6/10(0)
$\omega = N\backslash A$	$\rho_{\eta\delta} = -0.5$	44/26(15)	50/33/(24)	8/9(2)	11/9(3)
(no slope break)	$\rho_{\eta\delta} = 0.5$	70/42(28)	73/49(35)	8/7(2)	12/10(1)
$\omega = 0.1$	$\rho_{\eta\delta} = -0.5$	25/17(7)	32/21(11)	27/15(3)	11/7(1)
	$\rho_{\eta\delta} = 0.5$	90/71(44)	76/47(38)	40/20(6)	34/35(6)

Table 4: power for N=500; T=12

variable	parameter estimate	t-ratio
EXPEND(-1)	0.404	12.000
REV(-1)	0.034	1.040
GRANTS(-1)	0.068	0.633
Sargan test: 62.93, p-value = 0.004		

Table 5: Results for standard 2-step Arellano-Bond estimation

5 An application

As an illustration of how break detection can affect parameter estimates in practice, we now apply our method to the data taken from Dahlberg and Johansson (2000). This study examines causal links between expenditures and revenues of local governments based on a yearly sample of 265 Swedish municipalities over the period 1979-1987. Since our aim is purely illustrative, we will not replicate the entire original study, but rather re-examine the results presented in Table VI of Dahlberg and Johansson (2000, p. 414). In particular, we will re-estimate the equation

$$EXPEND_{i,t} = \eta_i + \rho EXPEND_{i,t-1} + \beta_1 REV_{i,t-1} + \beta_2 GRANTS_{i,t-1} + v_t + \varepsilon_{i,t} \quad (18)$$

where $EXPEND$ denotes local government expenditures and REV and $GRANTS$ represent the two components of revenues (own-source revenues and special grants from central government). The parameters β_1 and β_2 do not have a structural interpretation; rather, the aim of the original paper is to examine whether the revenue components Granger-cause expenditures. This is simply done by examining the statistical significance of the parameter estimates.

We do not wish to question the issues addressed in Dahlberg and Johansson (2000, p. 414)[14]; instead we illustrate what kinds of effects can arise when a structural break is accounted for. First applying the standard Arellano-Bond estimator for the model (18) yields the results in Table 5. We only use the second and third lags of each variable as instruments (resulting in a total of $(3 + 6 \times (9 - 3)) + 7 = 46$

[14]The original paper is concerned with the reliability of standard GMM asymptotics and examines whether bootstrap confidence intervals yield different inferences. The authors find that this is indeed the case, both in a Monte Carlo experiment and in the actual data.

	test-statistic	p-value
unknown breakpoint	19.7 (attained for break in 1983)	0.02
known BP (break in 1983)	19.7	0.003

Table 6: Results for unknown and known breakpoint tests. Tests do not allow for the presence of slope breaks (but these turn out to be not important).

moment conditions) because including further lags results in numerical singularity of the moment covariance matrix[15]. The estimates in Table 5 suggest that there is a fair amount of persistence in expenditures and that own-source revenues and especially grants do not Granger-cause expenditures. The joint validity of all 36 overidentifying moment restrictions is rejected by the Sargan test. These conclusions change somewhat when additional lags are included: in particular, the coefficient on REV does become statistically significant when all Arellano-Bond moment conditions are used[16].

Table 6 presents the results of applying our test for both the known and unknown breakpoint cases. Our break detector clearly reveals the presence of a structural break at $\tau = 1983$. We remark that test-statistics are numerically identically since 1983 saw the maximal value of the test-statistic - the test for known breakpoint was set up with the benefit of hindsight. Note that the p-value for the UBP test is larger than that for the KBP test since it reflects the fact that some searching for the largest break has taken place. Variations in the choice of instruments do not alter the conclusions of the test.

We have not been able to associate the break we discover with any kind of direct intervention. Reforms of local government finances happened before and after the sample period, but not in 1983 or 1982. We remark that an actual government reform is not the exclusive potential explanation for the observed break in the data: announcements of potential future reforms or even simple redefinitions in public accounting conventions may also result in the observations we make. Since the researcher may not always be aware of potential anomalies in the data, our break detector can act as a useful diagnostic, even when no direct interventions are a priori suspected to be present.

It is interesting to visually inspect the expenditure series as plotted in Figure 1 for the first 100 observations in the sample[17]. A level-shift is visible in 1983 for most expenditure series. Since it is negative for almost all municipalities, part of the effect will be picked up by the time-dummy. However, in line with the reasoning presented

[15] This is not really a problem since one can use generalized inverses - results are not substantially affected by doing so.

[16] The estimates in Table 5 differ from those in Dahlberg and Johansson (2000) because we use a different instrument set and a different first-step covariance matrix. When using exactly the same set-up as the original paper (all appropriate available lags of the dependent variable only as instruments and the 1st step covariance matrix of Holtz-Eakin, Newey, and Rosen (1988)), the results below do not change qualitatively (but they become somewhat stronger).

[17] The order of the observations was kept identical to that of the original file downloaded from the Journal of Applied Econometrics data archive. Plots of observations 100 to 256 result in the same overall impression but are suppressed to save space.

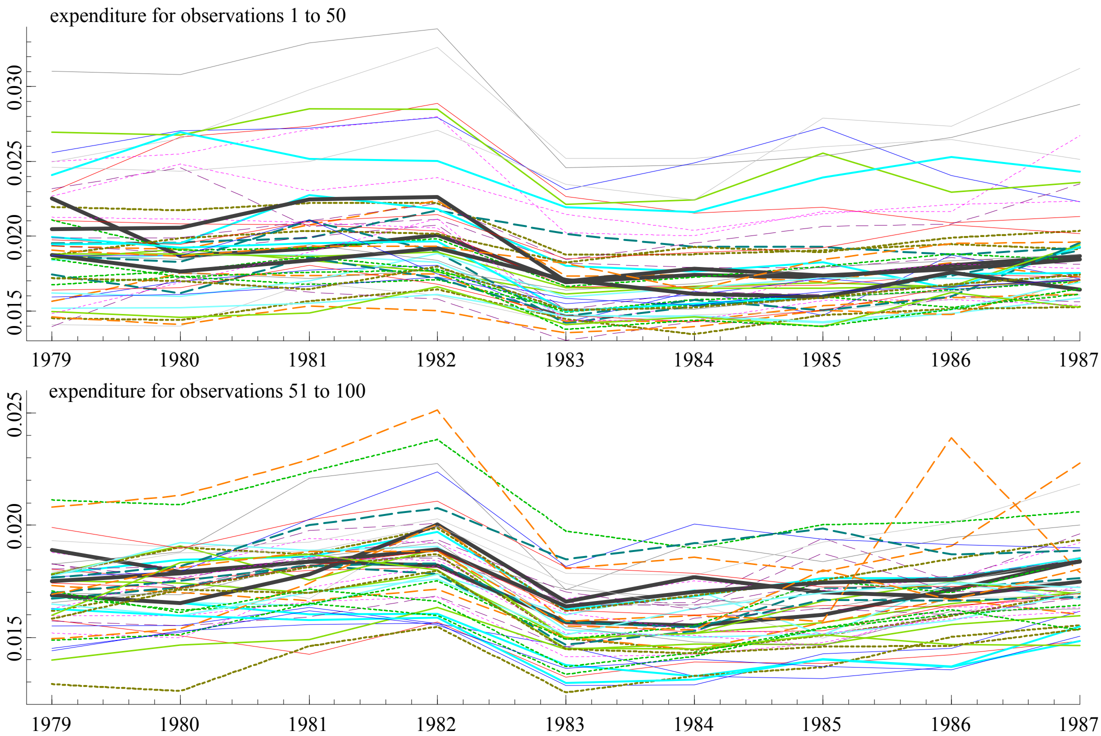

Figure 1: Plot of the expenditure time series by municipality for the first 100 municipalities in the sample.

in the introduction, there is no reason why an external intervention should affect all municipalities in the same manner and to the same extent. This differential impact, to the extent that it is correlated with the initial municipality-specific effect, leads to invalidity of some moment restrictions and allows us to detect a break.

We remark that, although the graphical evidence confirms the test results in this case, *purely* graphical methods are much less useful in short panel datasets than they are in typical time-series applications where the data consist of a single long time-series. Indeed, on the basis of Figure 1 alone, without knowing the information provided by our break detector, it would have been very difficult to convincingly argue for the presence of a break, especially given that any "apparent overall shifts in the data" may be explained by time dummies rather than a structural break.

As a counterpart to Table 5, Table 7 presents the results from estimation under the alternative that a break occurs at $\tau = 1983$. Clearly, the parameter estimates are drastically different. Most strikingly, the persistence parameter is much smaller and statistically far less significant than before. When one does not allow for the presence of a break, the persistence parameter mistakenly picks up the part of the level shift in the data that cannot be accounted for by the time-dummy. Once a break is accounted for, this effect disappears. Further, because the break in the present dataset is rather large, it represents a significant amount of the variability. This variability essentially disappears when a break with the correct timing is accounted for, explaining the drastically different estimation results for the parameters on

variable	parameter estimate	t-ratio
EXPEND(-1)	0.100	2.657
REV(-1)	-0.109	-2.825
GRANTS(-1)	0.398	2.183
Sargan test: 36.39, p-value = 0.20		

Table 7: Estimation results under the alternative that a break occurs at 1983. The estimator is a standard GMM estimator based on only the Arellano-Bond moment conditions that remain valid in the presence of a break at 1983.

lagged revenues. Perhaps most interestingly, the Sargan test for the model under the alternative does not reject the validity of the overidentifying restrictions anymore. This is additional evidence that the structural break led to the apparent inadequacy of model (18) for this dataset.

6 Conclusion

This paper has developed a break detection procedure for the standard $AR(p)$ panel regression model. The set-up allows for a general dependence structure between changes in fixed effects at a single unknown point in time and other sources of variability in the model as well as for breaks in the slope parameters. Other assumptions on the error structure are the weakest in the literature so as to avoid spurious detection of breaks.

Although no closed-form solution is available for the distribution of the test statistic when the breakpoint is unknown, implementation of the procedure is easy. Its performance was found to be satisfactory in a Monte Carlo study.

In a small illustrative application of our test to data on spending behaviour of Swedish municipalities, it was shown how the type of intercept breaks that are the main focus of the paper can affect estimation results. In particular estimates of persistence in the data can be misleading when structural breaks are ignored.

7 Appendix

In this Appendix, we prove the results given in the main text. We first derive the asymptotic distribution under the null hypothesis of a single component of $V(\widehat{\theta})$. Analogously to that of tests for overidentifying restrictions, this distribution will be chi-squared if the GMM weighting matrix is a consistent estimate of the inverse of the covariance matrix of the moment functions. The proof of Theorem (2) then follows by some reorganizing.

For notational simplicity we consider the case of the $AR(1)$ model with regressors, i.e. model (1) - (2) with $p = 1$. The parameter vector θ contains, in the following order, the autoregressive coefficients, the regressor coefficients, dummy coefficients, breaks in the AR and regressor coefficients (where applicable), and the dummy coefficient that is not identified under the alternative (where applicable). Wherever the symbol 0 or I is used in a matrix without further specifications, the dimensions should be clear from the context.

Proof of Theorem 1 Because of the assumption of cross-sectional independence and boundedness of fourth moments of ε, the multivariate Lindeberg-Levy CLT implies that

$$\sqrt{n}\widehat{m}_\tau(\theta_\tau^o) = \sqrt{n}\widehat{m}_\tau(\theta_\tau^o) \overset{a}{\sim} N(0, \Phi_\tau(\theta_\tau^o)) \qquad \forall \tau, t \tag{19}$$

Now let D_τ denote the Jacobian of the mapping $\theta_\tau \rightarrow m_\tau(\theta_\tau)$ evaluated at θ_τ^o. A consistent estimate of this matrix can be obtained by replacing expectations by sample averages and the true parameter value by a consistent estimate. The first order approximation to the sample moment vectors evaluated at the estimated values of the parameters reads (see e.g. Newey and McFadden (1994))

$$\sqrt{n}\widehat{m}_\tau(\widehat{\theta}_\tau) = \left(I - D_\tau \left(D_\tau' \Phi_\tau^{-1} D_\tau\right)^{-1} D_\tau' \Phi^{-1}\right) \sqrt{n}\widehat{m}_\tau(\theta_\tau^o) + o_p(1)$$

Reorganizing this expression, one obtains

$$\begin{aligned} \sqrt{n}\Phi^{-1/2}\widehat{m}_\tau(\widehat{\theta}_\tau) &= \left(I - \Phi^{-1/2} D_\tau \left(D_\tau' \Phi_\tau^{-1} D_\tau\right)^{-1} D_\tau' \Phi_\tau^{-1/2}\right) \sqrt{n}\Phi^{-1/2}\widehat{m}_\tau(\theta_\tau^o) + o_p(1) \\ &\equiv M_{\Phi_\tau^{-1/2} D_\tau} \left(\sqrt{n}\Phi_\tau^{-1/2}\widehat{m}_\tau(\theta_\tau^o)\right) + o_p(1) \end{aligned} \tag{20}$$

where the last equality is a definition: $M_{\Phi_\tau^{-1/2} D_\tau}$ is the idempotent matrix defining a projection off the space spanned by the columns of the matrix $\Phi_\tau^{-1/2} D_\tau$.

In what follows, fix τ and reorder the vector m_2 such that the moment conditions that are not included for estimation under the alternative of a break at τ are grouped at the bottom. The generic component of the vector (10) is written as

$$V^{(\tau-2)}(\widehat{\theta}) = n\left(\widehat{Q}_2(\widehat{\theta}_2) - \widehat{Q}_\tau(\widehat{\theta}_\tau)\right) = n\left(\widehat{m}_2(\widehat{\theta}_2)'\widehat{\Phi}_2^{-1}\widehat{m}_2(\widehat{\theta}_2) - \widehat{m}_\tau(\widehat{\theta}_\tau)'\widehat{\Phi}_\tau^{-1}\widehat{m}_\tau(\widehat{\theta}_\tau)\right)$$

Using (20) and ignoring $o_p(1)$ quantities this becomes

$$\begin{aligned} V^{(\tau-2)}(\widehat{\theta}) \overset{a}{=} \quad & \sqrt{n}\widehat{m}_2(\theta_2^o)'\Phi_2^{-1/2} M_{\Phi_2^{-1/2} D_2} \sqrt{n}\Phi_2^{-1/2}\widehat{m}_2(\theta_2^o) \\ & -\sqrt{n}\widehat{m}_\tau(\theta_\tau^o)'\Phi_\tau^{-1/2} M_{\Phi_\tau^{-1/2} D_\tau} \sqrt{n}\Phi_\tau^{-1/2}\widehat{m}_\tau(\theta_\tau^o) \end{aligned} \tag{21}$$

Note that $\theta_\tau^o = [\rho^o \ \beta^{o\prime} \ \nu_{-\tau}' \ 0]'$ where $\nu_{-\tau}$ refers to a full set of dummy coefficients excluding the one for the breakpoint and 0 is a $k+1$ vector of zeros (de values of the slope break parameters under the null). We use the convention that the omitted dummy is located at the bottom of the vector θ_2^o. All that is needed to show the chi-squaredness of expression (21) is to prove that $M_{\Phi_2^{-1/2} D_2} - \begin{bmatrix} M_{\Phi_\tau^{-1/2} D_\tau} & 0 \\ 0 & 0 \end{bmatrix}$ is idempotent. To do so, we need to deal with the fact that D_2 has more rows and potentially fewer columns than D_τ and that $\Phi_2^{-1/2}$ is not block-diagonal (and hence does not have a block equal to $\Phi_\tau^{-1/2}$.

First introduce a transformation that orthogonalizes the components of m_2 relating to the moment conditions that do not feature in m_τ to those that do. Writing $\Phi_2 = \begin{bmatrix} \Phi_{2(11)} & \Phi_{2(12)} \\ \Phi'_{2(12)} & \Phi_{2(22)} \end{bmatrix}$, this transformation is denoted by $A_\tau = \begin{bmatrix} I & 0 \\ -\Phi'_{2(12)}\Phi^{-1}_{2(11)} & I \end{bmatrix}$. Denote the transformed version of $\widehat{m}_2(\theta_2^o)$ by $\widetilde{m}_2(\theta_2^o)$, i.e.

$$\widetilde{m}_2(\theta_2^o) = A_\tau \widehat{m}_2(\theta_2^o)$$

and the counterpart of Φ_2 by $\widetilde{\Phi}_2$, which is block-diagonal. The Jacobian of the transformed moment functions is written as $\widetilde{D}_2$. Since $\widehat{m}_2(\widehat{\theta}_2)'\Phi_2^{-1}\widehat{m}_2(\widehat{\theta}_2) = \widetilde{m}_2(\widehat{\theta}_2)'\widetilde{\Phi}_2^{-1}\widetilde{m}_2(\widehat{\theta}_2)$ one has

$$\sqrt{n}\widehat{m}_2(\theta_2^o)'\Phi_2^{-1/2}M_{\Phi_2^{-1/2}D_2}\sqrt{n}\Phi_2^{-1/2}\widehat{m}_2(\theta_2^o) = \sqrt{n}\widetilde{m}_2(\theta_2^o)'\widetilde{\Phi}_2^{-1/2}M_{\widetilde{\Phi}_2^{-1/2}\widetilde{D}_2}\sqrt{n}\widetilde{\Phi}_2^{-1/2}\widetilde{m}_2(\theta_2^o)$$

Finally, let SL_τ denote the matrix selecting the columns corresponding to the moments that are still valid under the alternative from $\widetilde{m}_2'$, i.e. $SL_\tau\widetilde{m}_2(\theta_2^o) = SL_\tau\widehat{m}_2(\theta_2^o)$ $= \widehat{m}_\tau(\theta_\tau^o)$ (because $\theta_\tau^o = [\rho^o\ \beta^{o\prime}\ 0\ \nu'_{-\tau}]'$). Using this notation we now rewrite expression (21) as

$$\begin{aligned} V^{(\tau-2)}(\widehat{\theta}) \quad &\overset{a}{=} \sqrt{n}\widehat{m}_2(\theta_2^o)'\Phi_2^{-1/2}(\Phi_2^{1/2}A'_\tau\widetilde{\Phi}_2^{-1/2} \\ &\{M_{\widetilde{\Phi}_2^{-1/2}\widetilde{D}_2} - SL'_\tau M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau\}\widetilde{\Phi}_2^{-1/2}A_\tau\Phi_2^{1/2})\sqrt{n}\Phi_2^{-1/2}\widehat{m}_2(\theta_2^o) \end{aligned} \tag{22}$$

At this point some extra notation is needed. Define $\widetilde{D}_2 := \begin{bmatrix} \widetilde{D}_{2(1)} \\ \widetilde{D}_{2(2)} \end{bmatrix}$, and $\widetilde{D}_{2(1)} = \begin{bmatrix} \widetilde{D}_{2(I)} & 0 \end{bmatrix}$ with 0 a column vector of zeros, and $\widetilde{D}_{2(2)} = \begin{bmatrix} \widetilde{D}_{2(II)} & d_\tau \end{bmatrix}$ where the column vector d_τ is the observed derivative of the moment conditions that are obsolete under the alternative w.r.t. the dummy for the breakpoint equation. Using these conventions, note that $\widetilde{D}_{2(I)}$ is equal to the first block of columns of D_τ (denoted correspondingly by $D_{\tau(I)}$). Also note that the upper left block of $\widetilde{\Phi}_2^{-1/2}$, $\widetilde{\Phi}_{2(11)}^{-1/2}$ is equal to $\Phi_\tau^{-1/2}$.

It only remains to be proven that the matrix between {} is idempotent. Denote $M_{\widetilde{\Phi}_2^{-1/2}\widetilde{D}_2}$ as $\widetilde{M} = I - \widetilde{P}$, we need to show (since $SL'_\tau M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau = \begin{bmatrix} M_{\Phi_\tau^{-1/2}D_\tau} & 0 \\ 0 & 0 \end{bmatrix}$ is idempotent) that

$$\widetilde{M}.SL'_\tau M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau = SL'_\tau M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau.$$

Indexing the blocks of $\widetilde{M}$ according to the partitioning of the moment functions, we therefore need to establish that

$$\widetilde{P}_{21} = \widetilde{P}_{21}P_{\Phi_\tau^{-1/2}D_\tau} \tag{23}$$

$$\widetilde{P}_{11} = \widetilde{P}_{11}P_{\Phi_\tau^{-1/2}D_\tau} \tag{24}$$

The RHS of condition (23) is written out as

$$\widetilde{\Phi}_{2(22)}^{-1/2}\widetilde{D}_{2(2)}\left(\widetilde{D}'\widetilde{\Phi}_2^{-1}\widetilde{D}\right)^{-1}\widetilde{D}'_{2(1)}\widetilde{\Phi}_{2(11)}^{-1/2}\Phi_\tau^{-1/2}D_\tau\left(D'_\tau\Phi_\tau^{-1}D_\tau\right)^{-1}D'_\tau\Phi_\tau^{-1/2}$$

Now define $D_\tau = \left[\begin{array}{cc} D_{\tau(I)} & D_{\tau(II)} \end{array}\right]$ and note that $D_{\tau(I)} = \widetilde{D}_{2(I)}$. Define the transformation

$$B_\tau = \left[\begin{array}{cc} 1 & -\left(D'_{\tau(I)}\Phi_\tau^{-1}D_{\tau(I)}\right)^{-1}\left(D'_{\tau(I)}\Phi_\tau^{-1}D_{\tau(II)}\right) \\ 0 & I \end{array}\right]$$

which orthogonalizes the second colomn of the matrix $\Phi_\tau^{-1/2}D_\tau$ w.r.t. the first (by postmultiplication), i.e. $\Phi_\tau^{-1/2}D_\tau B_\tau$ has orthogonal blocks of columns. The first block of columns is not affected by the transformation. We can rewrite

$$\begin{aligned} P_{\Phi_\tau^{-1/2}D_\tau} &= \Phi_\tau^{-1/2}D_\tau\left(D'_\tau\Phi_\tau^{-1}D_\tau\right)^{-1}D'_\tau\Phi_\tau^{-1/2} \\ &= \Phi_\tau^{-1/2}D_\tau B_\tau\left(B'_\tau D'_\tau\Phi_\tau^{-1}D_\tau B_\tau\right)^{-1}B'_\tau D'_\tau\Phi_\tau^{-1/2}. \end{aligned}$$

For the following, recall that the first block of columns of $\Phi_\tau^{-1/2}D_\tau B_\tau$ equals $\widetilde{\Phi}_{2(11)}^{-1/2}\widetilde{D}_{2(I)}$ and that $(B'_\tau D'_\tau\Phi_\tau^{-1}D_\tau B_\tau)^{-1}$ is block-diagonal by construction. We can now write

$$\begin{aligned} & \widetilde{D}'_{2(I)}\widetilde{\Phi}_{2(11)}^{-1/2}\Phi_\tau^{-1/2}D_\tau B_\tau\left(B'_\tau D'_\tau\Phi_\tau^{-1}D_\tau B_\tau\right)^{-1}B'_\tau D'_\tau\Phi_\tau^{-1/2} \\ = \; & \widetilde{D}'_{2(I)}\widetilde{\Phi}_{2(11)}^{-1/2}\left[\Phi_\tau^{-1/2}D_{\tau(I)} \;\; (\Phi_\tau^{-1/2}D_\tau B_\tau)_{(II)}\right]\left(B'_\tau D'_\tau\Phi_\tau^{-1}D_\tau B_\tau\right)^{-1}B'_\tau D'_\tau\Phi_\tau^{-1/2} \\ = \; & \left[D'_{\tau(I)}\Phi_\tau^{-1/2}\Phi_\tau^{-1/2}D_{\tau(I)} \;\; D'_{\tau(I)}\Phi_\tau^{-1/2}(\Phi_\tau^{-1/2}D_\tau B_\tau)_{(II)}\right]\left(B'_\tau D'_\tau\Phi_\tau^{-1}D_\tau B_\tau\right)^{-1}B'_\tau D'_\tau\Phi_\tau^{-1/2} \\ = \; & \left[D'_{\tau(I)}\Phi_\tau^{-1}D_{\tau(I)} \;\; 0\right]\left[\begin{array}{cc} \left(D'_{\tau(I)}\Phi_\tau^{-1}D_{\tau(I)}\right)^{-1} & 0 \\ 0 & \left(\begin{array}{c} (B'_\tau D'_\tau\Phi_\tau^{-1/2})_{(II)}\times \\ (\Phi_\tau^{-1/2}D_\tau B_\tau)_{(II)} \end{array}\right)^{-1} \end{array}\right]B'_\tau D'_\tau\Phi_\tau^{-1/2} \\ = \; & \left[\begin{array}{cc} I & 0 \end{array}\right]\left[\begin{array}{c} D'_{\tau(I)}\Phi_\tau^{-1/2} \\ B'_\tau D'_{\tau(II)}\Phi_\tau^{-1/2} \end{array}\right] = D'_{\tau(I)}\Phi_\tau^{-1/2} = \widetilde{D}'_{2(I)}\widetilde{\Phi}_{2(11)}^{-1/2} \end{aligned}$$

Finally, recall that $\widetilde{D}_{2(1)} = \left[\begin{array}{cc} \widetilde{D}_{2(I)} & 0 \end{array}\right]$. We now have

$$\widetilde{D}'_{2(1)}\widetilde{\Phi}_{2(11)}^{-1/2}\Phi_\tau^{-1/2}D_\tau\left(D'_\tau\Phi_\tau^{-1}D_\tau\right)^{-1}D'_\tau\Phi_\tau^{-1/2} = \widetilde{D}'_{2(1)}\widetilde{\Phi}_{2(11)}^{-1/2}$$

proving the validity of (23). The argument proving statement (24) is similar.

The conclusion is that $M_{\widetilde{\Phi}_2^{-1/2}\widetilde{D}_2} - SL'_\tau M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau$ in (22) is idempotent with rank #rows($\widehat{m}_2$) - #rows($\widehat{m}_\tau$) + dim(θ_τ) - dim(θ_2). This proves the theorem.

Proof of Theorem 2 So far we have assumed that the obsolete moment conditions under the alternative are grouped at the bottom of $\widehat{m}_2(\theta_2^o)$. To make the notation more general, we introduce the row-reordering matrix R_τ. Note that $R_\tau^{-1} = R'_\tau$. Where equation (22) assumed the correct ordering, we want to make the different ordering explicit. If A_τ and SL_τ already assume the obsolete moment conditions to be at the bottom of the vector $\widehat{m}_2(\theta_2^o)$ but other quantities do not, we get

$$\begin{aligned} V^{(\tau-2)}(\widehat{\theta}) \overset{a}{=} \; & \sqrt{n}\widehat{m}_2(\theta_2^o)'R'_\tau R_\tau\Phi_2^{-1/2}R'_\tau(R_\tau\Phi_2^{1/2}R'_\tau A'_\tau R_\tau\widetilde{\Phi}_2^{-1/2}R'_\tau\{M_{R_\tau\widetilde{\Phi}_2^{-1/2}\widetilde{D}_2} \\ & -SL'_\tau M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau\}R_\tau\widetilde{\Phi}_2^{-1/2}R'_\tau A_\tau R_\tau\Phi_2^{1/2}R'_\tau)\sqrt{n}R_\tau\Phi_2^{-1/2}R'_\tau R_\tau\widehat{m}_2(\theta_2^o) \\ = \; & v(\Phi_2^{1/2}R'_\tau A'_\tau\left(R_\tau\widetilde{\Phi}_2^{-1/2}R'_\tau\right)\{M_{R_\tau\widetilde{\Phi}_2^{-1/2}\widetilde{D}_2} \\ & -SL'_\tau M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau\}\left(R_\tau\widetilde{\Phi}_2^{-1/2}R'_\tau\right)A_\tau R_\tau\Phi_2^{1/2})v \end{aligned}$$

where we have used the fact that $\sqrt{n}\widehat{m}_2(\theta_2^o)'\Phi_2^{-1/2}$ has asymptotically a multivariate standard normal distribution. Construction of the matrix $R_\tau\widetilde{\Phi}_2^{-1/2}R_\tau'$ is most easily done by first reordering Φ_2 and then applying the orthogonalization.

Denoting

$$\begin{aligned} G_\tau \quad : \quad &= \Phi_2^{1/2}R_\tau'A_\tau'\left(R_\tau\widetilde{\Phi}_2^{-1/2}R_\tau'\right)\left\{M_{R_\tau\widetilde{\Phi}_2^{-1/2}\widetilde{D}_2} - SL_\tau'M_{\Phi_\tau^{-1/2}D_\tau}SL_\tau\right\}\times \\ &\left(R_\tau\widetilde{\Phi}_2^{-1/2}R_\tau'\right)A_\tau R_\tau\Phi_2^{1/2} \end{aligned}$$

we have that the joint distribution of V is

$$V(\widehat{\theta}) \overset{a}{=} \left(v'\left[G_3 \ \cdots \ G_T\right]\left(I_{T-2}\otimes v\right)\right)' \tag{25}$$

The components of V are asymptotically marginally χ^2 distributed, but jointly are distributed as in (25), which is a multivariate chi-squared distribution with a correlation structure given by the matrices G_τ. The asymptotic distribution of the maximum of the scaled components of $V(\widehat{\theta})$ equals that of the maximum of the scaled right-hand side of (25) because the maximum component of a vector is a continuous function of its components. One can therefore apply the continuous mapping theorem.

References

Ahn, S., and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–28.

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.

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Andrews, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67 (3), 543–564.

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.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Ecometrica*, 62, 1383–1414.

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.

ARELLANO, M., AND O. BOVER (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

BAI, J. (1994): "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–472.

——— (1997): "Estimation of a Change Point in Multiple Regression Models," *The Review of Economics and Statistics*, 79, 551–563.

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

BROWN, R. L., J. DURBIN, AND J. M. EVANS (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society Series B*, 37, 149–192.

CHOW, G. C. (1960): "Tests for Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605.

DAHLBERG, M., AND E. JOHANSSON (2000): "An examination of the dynamic behaviour of local governments using GMM bootstrapping methods," *Journal of Applied Econometrics*, 15, 401–416.

DOORNIK, J. (1999): *Object-oriented matrix programming using Ox*. Timberlake Consultants Press and www.nuff.ox.ac.uk/Users/Doornik, 3rd edn.

HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56(6), 1371–1395.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics Volume IV*, ed. by R. F. Engle, and D. L. McFadden, chap. 36, pp. 2111–2245. Elsevier Science.

PERRON, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.

QUANDT, R. E. (1960): "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–330.

TZAVALIS, E. (2002): "Structural Breaks and Unit Root Tests for Short Panels," *mimeo, Queen Mary, University of London*, p. 30.

ZIVOT, E., AND D. W. K. ANDREWS (1992): "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmul.ac.uk
Website: www.econ.qmul.ac.uk/papers/wp.htm