A New Structural Break Test for Panels with Common Factors

Huanjun Zhu*[1], Vasilis Sarafidis†[2], and Mervyn J. Silvapulle ‡ [2]

[1] *Department of Statistics, School of Economics, Xiamen University, China*

[2] *Department of Econometrics and Business Statistics, Monash University, Australia*

Abstract

This paper develops new tests against structural breaks in panel data models with common factors when T is fixed, where T denotes the number of observations over time. For this class of models, the available tests against a structural break are valid only under the assumption that T is 'large'. However, this may be a stringent requirement; more commonly so in datasets with annual time frequency, in which case the sample may cover a relatively long period even if T is not large. The proposed approach builds upon the existing

*This author would like to thank Monash University for support through a Postgraduate Scholarship, and also the Fundamental Research Funds for the Central Universities for support under grant number 20720171071.

†Corresponding Author. Email: vasilis.sarafidis@monash.edu. This author gratefully acknowledges financial support from the Australian Research Council, under research grant number DP-170103135.

‡This author gratefully acknowledges support from the Australian Research Council, under research grant number DP150100210. All three authors would like to express their warm appreciation to Jiti Gao for helpful comments and discussions. Further, all three authors thank two anonymous referees and the editor, Professor Dennis Kristensen, for comments and suggestions that led to significant improvements.

GMM methodology and develops distance-type and LM-type tests for detecting a structural break, both when the breakpoint is known as well as when it is unknown. The proposed methodology permits weak exogeneity and/or endogeneity of the regressors. In a simulation study, the method performed well, both in terms of size and power, as well as in terms of successfully locating the time of a structural break. The method is illustrated by testing the so-called 'Gibrat's Law', using a dataset from 4,128 financial institutions, each one observed for the period 2002-2014.

Key words: Method of moments, unobserved heterogeneity, break-point detection, fixed T asymptotics.
JEL Classification: C12, C23, C26.

1 Introduction

Methods of testing for the presence of structural breaks are important in econometrics and statistics. Failure to incorporate such breaks in the model, if they exist, may lead to unreliable inferences and forecasts; on the other hand, incorporating a break when it does not exist, would unnecessarily complicate econometric analyses and/or lead to loss of efficiency. There is a vast time-series literature on testing for structural breaks in the mean (e.g. Harchaoui and Lévy-Leduc 2010), in the variance (e.g. Chen and Gupta 1997), in the covariance structure (e.g. Aue et al. 2009), and in regression models (Andrews 1993, Bai and Perron 1998, Perron and Qu 2006, and Qu and Perron 2007). More recently, there has been a growing literature on structural-break tests for panel data models.

A family of models for panel data that has been a topic of much recent research is the *common factor approach*. This paper develops a new test to detect a *single* structural break in panels with common factors when the number of cross-sectional units (N) is large and the number of observations over time (T) is small; in the asymptotic results, this is represented by $N \to \infty$ and T is fixed. This is an empirically relevant scenario in panel data analysis, especially when the time frequency of the data is annual, in which case the sample may cover a relatively long period even if T is much smaller than N. As an example, the application considered in Section 4 studies the relationship between the size of a firm and its growth rate, and investigates the well-known 'Law of Proportionate Effect'. The data set consists of over 4,000 banks, each one observed annually over the 13 years, 2002 to 2014. As such, T is relatively small. Testing for a structural break is important, particularly because the sample of the study spans the GFC. Indeed, the results of our analysis indicate the existence of a structural break in the model, following the establishment of the 'Basel III' capital regulatory framework in 2011.

Currently a suitable method is not available for testing against a structural break in the general context studied in this paper. That is, existing literature on structural break testing in panels with common factors is designed specifically for models with $T \to \infty$. Some notable contributions include Chan et al. (2008), which extends the time series test statistic of Andrews (2003) to heterogeneous panel data models; Baltagi et al. (2016), which studies estimation of static heterogenous panels with a common break using the common correlated effects estimator of Pesaran (2006); Qian and Su (2016) and Li et al. (2016), which consider estimation and inference of possibly multiple common breaks in panel data models, allowing for common factors and cross-sectional dependence. Therefore, the present paper makes a significant methodological contribution to this currently active area of research.

The proposed methodology builds upon and extends the GMM approach in Robertson and Sarafidis (2015) by developing distance-type and LM-type tests for detecting a structural break. In particular, Robertson and Sarafidis (2015) considered point estimation of an unknown parameter; by contrast, the present paper focuses on testing. Thus, these two papers study different topics in inference, and the main results in our paper cannot be deduced from those in Robertson and Sarafidis (2015). We consider both the case where the breakpoint is known as well as when it is unknown. For the case when the breakpoint is unknown, we apply the Union Intersection Principle to develop suitable tests. The proposed approach remains valid under weak exogeneity and/or endogeneity of regressors. An extensive simulation study demonstrates that the method performs well, both in terms of size and power, as well as in terms of successfully locating the break point.

The rest of the paper is organized as follows. Section 2 describes the model and develops the theory for the test against a structural break. Section 3 examines the performance of the method in finite samples using simulated data. Section 4 presents the aforementioned empirical illustration. Section 5 provides a discussion of the results and Section 6 concludes. Proofs of the main results are provided in the Appendix, and the remaining ones are available in Supplementary Materials to this paper.

2 A New Structural Break Test

We start with a description of the model and a discussion of the required assumptions. The next subsection describes the moment conditions employed, and provides an illustrative example. The final part provides the asymptotic results of the paper.

2.1 Model Specification

We study a linear panel data model with regressors and a multi-factor error structure. Our aim is to detect a possible break in the structural parameters of the model, i.e. the slope coefficients. Consider the model,

$$y_{it} = \begin{cases} \boldsymbol{x}_{it}'\boldsymbol{\beta}_1^0 + \boldsymbol{\lambda}_i'\boldsymbol{f}_t^0 + \varepsilon_{it}, & t = 1, \ldots, \tau - 1; \\ \boldsymbol{x}_{it}'\boldsymbol{\beta}_\tau^0 + \boldsymbol{\lambda}_i'\boldsymbol{f}_t^0 + \varepsilon_{it}, & t = \tau, \ldots, T, \end{cases} \tag{1}$$

where $\boldsymbol{x}_{it}$ is a $K \times 1$ vector of regressors, $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t^0$ denote $r \times 1$ vectors of factor loadings and factors, respectively, and ε_{it} is a purely idiosyncratic error term ($t = 1, \ldots, T$). Notice that $\boldsymbol{\beta}_\tau^0$ replaces $\boldsymbol{\beta}_1^0$ starting from the break at time τ ($\tau \geq 2$). The foregoing model for structural break assumes that the break may be in any component of $\boldsymbol{\beta}$. The derivations presented in this paper need only minor modifications if the possible break is limited to a subvector of $\boldsymbol{\beta}$.

Our stochastic framework (1) is robust under a structural break in the unobserved factor loadings. To illustrate this, consider a single-factor model with a common break point in λ_i at period τ, expressed as $u_{it} = (\lambda_i f_t + \varepsilon_{it})\mathbb{I}(1 \leq t < \tau) + (\lambda_{i,\tau} f_t + \varepsilon_{it})\mathbb{I}(\tau \leq t \leq T)$, where $\mathbb{I}(\cdot)$ is the indicator function. Empirically, this structure can be captured by the following two-factor model: $u_{it} = \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \varepsilon_{it}$, for $t = 1, ..., T$ with $\lambda_{1i} = \lambda_i$, $\lambda_{2i} = \lambda_{i,\tau}$, $f_{1t} = d_{t,\tau} f_t$ and $f_{2t} = (1 - d_{t,\tau}) f_t$, where $d_{t,\tau}$ is a dummy variable that takes the value of unity when $t = 1, ..., \tau - 1$, and zero otherwise. Therefore, the foregoing model with a structural break in factor loadings is of the form (1).

The model (1) can be expressed in vector form as

$$\boldsymbol{y}_i = \boldsymbol{X}_i^{(1)} \boldsymbol{\beta}_1^0 + \boldsymbol{X}_i^{(\tau)} \boldsymbol{\beta}_\tau^0 + (\boldsymbol{I}_T \otimes \boldsymbol{\lambda}_i') \boldsymbol{f}^0 + \boldsymbol{\varepsilon}_i, \tag{2}$$

where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'_{T\times 1}$, $\boldsymbol{X}_i^{(1)} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{i,\tau-1}, \boldsymbol{0}_{K\times 1}, \ldots, \boldsymbol{0}_{K\times 1})'_{T\times K}$, $\boldsymbol{X}_i^{(\tau)} = (\boldsymbol{0}_{k\times 1}, \ldots, \boldsymbol{0}_{K\times 1}, \boldsymbol{x}_{i\tau}, \ldots, \boldsymbol{x}_{iT})'_{T\times K}$, $\boldsymbol{f}^0 = \text{vec}\left[(\boldsymbol{F}^0)'\right]$, $\boldsymbol{F}^0 = (\boldsymbol{f}_1^0, \boldsymbol{f}_2^0, \ldots, \boldsymbol{f}_T^0)'_{T\times r}$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'_{T\times 1}$; the superscripts '(1)' and '(τ)' indicate that the vectors correspond to the periods before τ and from τ onwards respectively, irrespective of whether or not a break has occurred. Next, let us introduce the following three hypotheses:

H_0 : There are no structural breaks;

$H_{(\tau)}$: There is a structural break at time τ, where τ is known;

H_1 : There is a structural break at time τ, where τ is unknown.

If H_1 is true then we denote the true point in time where the break occurs, by τ_0. In the rest of this section, we develop a Distance-type test and an LM-type test, based on the Method of Moments, for H_0 against H_1 as well as for H_0 against H_τ.

2.2 Moment Conditions

We assume that there exists a $d \times 1$ vector of potential instruments, $\boldsymbol{w}_i$; these instruments may correspond to the variables of the model or be extraneous variables. In each period t, ζ_t instruments are available, expressed in vector form as $\boldsymbol{z}_{it} = \boldsymbol{S}_t \boldsymbol{w}_i$, where $\boldsymbol{S}_t$ is a $\zeta_t \times d$ selector matrix of 0's and 1's that picks up from $\boldsymbol{w}_i$ the variables at period t that are uncorrelated with ε_{it}, i.e. for which $E(\boldsymbol{z}_{it}\varepsilon_{it}) = \mathbf{0}$ holds true. The total number of moment conditions is $\zeta \equiv \sum_{t=1}^{T} \zeta_t$. The following assumption is employed throughout the paper.

Assumption 1: (i) $(\boldsymbol{x}_{it}, \boldsymbol{w}_i, \boldsymbol{\lambda}_i, \varepsilon_{it})$ *are independently and identically distributed [i.i.d.] for* $i = 1, \ldots, N$, *with each component having finite fourth moment.* (ii) $\boldsymbol{f}^0$ *is non-stochastic and there exists* $c < \infty$ *such that* $\|\boldsymbol{f}^0\| \leq c$. (iii) $E(\varepsilon_{it} | \boldsymbol{\lambda}_i, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{ih}) = 0$, *for* $t = 1, \ldots, T$, $i = 1, \ldots, N$, *and some positive integer* h.

Assumption 1 is often employed in the literature of fixed-T panels with endogenous regressors; for example, see Assumptions BA.1-BA.4 in Ahn et al. (2013), and Assumption 2 in Robertson and Sarafidis (2015).

The independence assumption over $i = 1, \ldots, N$, in the first part of Assumption 1, can be weakened so long as additional boundedness restrictions on higher moments of the data generating process are imposed. The requirement that the observations are identically distributed can also be relaxed. For instance, ε_{it} could be heterogeneously distributed across both i and t. As with a large body of the fixed-T panel data literature, we do not consider such generalizations to avoid unnecessary notational complexity.

Assumption 1(ii) treats $\boldsymbol{f}^0$ as fixed constants in asymptotic analysis, which is typical in the literature. Alternatively, all probabilistic statements can be formulated conditionally on $\boldsymbol{F}$ without qualitatively changing the main results of this paper; see Kuersteiner and Prucha (2013).

The value of h in Assumption 1 (iii) characterises the exogeneity properties of the covariates. In particular, for $h = T$ (respectively, $h = t$) the covariates are strictly (respectively, weakly) exogenous; otherwise they would be endogenous (see Arellano 2003, Section 8.1). Our methodology is valid irrespective of the value of h mutatis mutandis. Consequently, our framework allows for lagged values of the dependent variable to be part of the covariates.

We also note that Assumption 1 (iii) implies that the idiosyncratic errors are conditionally serially uncorrelated. In practice, it is entirely straightforward to allow for serial correlation of (a) a moving average form by carefully selecting the moment conditions, or (b) an autoregressive form by including further lags of y and x into the model; see e.g. Bond (2002). Finally, Assumption 1 (iii) implies that the idiosyncratic error is conditionally uncorrelated with the factor loadings. This is standard in the panel data literature, and allows for lagged values of y in levels to be used as instruments.

Let $\boldsymbol{S} = diag\,(\boldsymbol{S}_1; \ldots; \boldsymbol{S}_T)$ and $\boldsymbol{Z}_i' \equiv \boldsymbol{S}\,(\boldsymbol{I}_T \otimes \boldsymbol{w}_i)$. Define $\boldsymbol{G} \equiv E\,(\boldsymbol{w}_i \boldsymbol{\lambda}_i')$, a $d \times r$ matrix that contains the unrestricted covariances between the instruments and the factor loadings, and let

$$\boldsymbol{\mu}_{\tau,i}\,(\boldsymbol{\theta}_\tau) \equiv \boldsymbol{Z}_i'\{\boldsymbol{y}_i - \boldsymbol{X}_i^{(1)}\boldsymbol{\beta}_1 - \boldsymbol{X}_i^{(\tau)}\boldsymbol{\beta}_\tau\} - \boldsymbol{S}\,(\boldsymbol{I}_T \otimes \boldsymbol{G})\,\boldsymbol{f}, \tag{3}$$

where $\boldsymbol{\theta}_\tau \equiv (\boldsymbol{g}', \boldsymbol{f}', \boldsymbol{\beta}_1', \boldsymbol{\beta}_\tau')'$ with $\boldsymbol{g} = \mathrm{v}ec\,(\boldsymbol{G})$. Under Assumption 1, taking expectations of (3) yields the following vector-valued moment function:

$$\boldsymbol{\mu}_\tau\,(\boldsymbol{\theta}_\tau) \equiv E\,[\boldsymbol{\mu}_{\tau,i}\,(\boldsymbol{\theta}_\tau)] = \boldsymbol{m} - \boldsymbol{M}^{(1)}\boldsymbol{\beta}_1 - \boldsymbol{M}^{(\tau)}\boldsymbol{\beta}_\tau - \boldsymbol{S}\,(\boldsymbol{I}_T \otimes \boldsymbol{G})\,\boldsymbol{f}, \tag{4}$$

where $\boldsymbol{m} = E[\boldsymbol{Z}_i'\boldsymbol{y}_i]_{\zeta \times 1}$ and $\boldsymbol{M}^{(j)} = E[\boldsymbol{Z}_i'\boldsymbol{X}_i^{(j)}]_{\zeta \times K}$, $(j = 1, \tau)$. It follows that at the true parameter value $\boldsymbol{\theta}_\tau^0$, we have $\boldsymbol{\mu}_\tau\,(\boldsymbol{\theta}_\tau^0) = \mathbf{0}$.

It is worth providing some motivation for the choice of the moment function defined in (3) and (4). In order to develop a method of moments estimator of the parameter in model (1), one approach would be to start with residuals of the form $\boldsymbol{e}_i(\boldsymbol{\theta}_\tau) = \boldsymbol{y}_i - [\boldsymbol{X}_i^{(1)}\boldsymbol{\beta}_1 + \boldsymbol{X}_i^{(\tau)}\boldsymbol{\beta}_\tau + (\boldsymbol{I}_T \otimes \boldsymbol{\lambda}_i')\boldsymbol{f}]$, $(\boldsymbol{\theta}_\tau \in \Theta)$. Then define the moment function as $\boldsymbol{\mu}_\tau\,(\boldsymbol{\theta}_\tau) = E[\boldsymbol{Z}_i'\boldsymbol{e}_i(\boldsymbol{\theta}_\tau)]$, where $\boldsymbol{Z}_i$ are instruments such that $E[\boldsymbol{Z}_i'\boldsymbol{e}_i(\boldsymbol{\theta}_\tau)]$ is zero at the true parameter value $\boldsymbol{\theta}_\tau^0$. Our next step would be to construct a suitable sample counterpart of $E[\boldsymbol{Z}_i'\boldsymbol{e}_i(\boldsymbol{\theta}_\tau)]$, which we denote by $\hat{\boldsymbol{\mu}}_\tau\,(\boldsymbol{\theta}_\tau)$, and finally introduce a GMM-type objective function of the form $\hat{\boldsymbol{\mu}}_\tau'\,(\boldsymbol{\theta}_\tau)\,\hat{\boldsymbol{W}}\,\hat{\boldsymbol{\mu}}_\tau\,(\boldsymbol{\theta}_\tau)$, where $\hat{\boldsymbol{W}}$ is a suitably chosen weight matrix. At first glance, a suitable choice for $\hat{\boldsymbol{\mu}}_\tau\,(\boldsymbol{\theta}_\tau)$ appears to be the sample average $N^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i'\boldsymbol{e}_i(\boldsymbol{\theta}_\tau)$. If we were to do so and estimate $(\boldsymbol{\theta}_\tau, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)$, treating them as unknown parameters, then the number of parameters would increase with the sample size and we would encounter an incidental parameter problem. However, our approach is different. We treat $\boldsymbol{\lambda}_i$ appearing in $\boldsymbol{e}_i(\boldsymbol{\theta}_\tau)$ through the term $(\boldsymbol{I}_T \otimes \boldsymbol{\lambda}_i')\boldsymbol{f}$, as an unobserved random variable. Therefore, we work with the form $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau) = E\left[\boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i^{(1)}\boldsymbol{\beta}_1 - \boldsymbol{X}_i^{(\tau)}\boldsymbol{\beta}_\tau\right)\right] - \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{G})\boldsymbol{f}$, and define $\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1}\sum_{i=1}^N \boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i^{(1)}\boldsymbol{\beta}_1 - \boldsymbol{X}_i^{(\tau)}\boldsymbol{\beta}_\tau\right) - \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{G})\boldsymbol{f}$; note that the sample averaging applies only to the part $E\left[\boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i^{(1)}\boldsymbol{\beta}_1 - \boldsymbol{X}_i^{(\tau)}\boldsymbol{\beta}_\tau\right)\right]$ in $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$. This leads to $\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1}\sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$ with $\boldsymbol{\mu}_{\tau,i}$ as in (3). The individual effects $\{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N\}$ enter the GMM objective function, $\hat{\boldsymbol{\mu}}_\tau'(\boldsymbol{\theta}_\tau)\hat{\boldsymbol{W}}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau)$, only through the expected values $E[w_i\boldsymbol{\lambda}_i]$, which we parameterize using the parameter $\boldsymbol{g}$ of fixed dimension. Consequently, the GMM objective function involves only the parameter $\boldsymbol{\theta}_\tau$ and the observed data, but not the collection $\{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N\}$. Therefore, the dimension of the parameters estimated does not change with N and hence the usual incidental parameter problem does not arise.

The requirement that $\{\boldsymbol{\lambda}_i, i = 1, \ldots, N\}$ be identically distributed can be relaxed to some extent. For example, conditional moments of $\boldsymbol{\lambda}_i$ may depend on i. To illustrate this, consider the case $K = 1$ and $r = 1$. In this case, one may set $E(\lambda_i | x_{i,1}, \ldots, x_{i,T}) = \pi_1 \bar{x}_i$ and $var(\lambda_i | x_{i,1}, \ldots, x_{i,T}) = \pi_2 \bar{x}_i{}^2$. In this case, conditionally upon x, the variance-covariance matrix of the moment conditions involves terms of the form $\pi_2 E[z_{i,t} x_{i,t} \bar{x}_i{}^2]$, which can be estimated consistently using suitable simple sample averages.[1]

In contrast to the setting of this paper, a recent alternative literature treats the factor loadings as fixed (incidental) parameters; for example, see Bai (2013). The advantage of this approach is that it relaxes the i.i.d. assumption on λ_i. On the other hand, the method imposes the nontrivial initial condition $y_{i,0} = 0$, $(i = 1, \ldots, N)$.[2] Moreover, for T fixed the method assumes conditional homoskedasticity in ε_{it}, as well as exogenous covariates.

[1] As our simulations indicate, our estimator appears to be robust even under unconditional heteroskedasticity in $\boldsymbol{\lambda}_i$.

[2] Hsiao (2014), page 88, provides an analysis of the limitations of treating $y_{i,0}$ as fixed when T is small.

Observe that the last term $\boldsymbol{S}\left(\boldsymbol{I}_T \otimes \boldsymbol{G}\right)\boldsymbol{f}$ in (4) can be written as $\boldsymbol{S}\left(\boldsymbol{F} \otimes \boldsymbol{I}_d\right)\boldsymbol{g} = \boldsymbol{S}\text{vec}(\boldsymbol{G}\boldsymbol{F}')$. But since $\boldsymbol{S}\text{vec}(\boldsymbol{G}\boldsymbol{F}') = \boldsymbol{S}\text{vec}(\boldsymbol{G}\boldsymbol{U}\boldsymbol{U}^{-1}\boldsymbol{F}')$ for any $r \times r$ invertible matrix $\boldsymbol{U}$, the parameters $\boldsymbol{G}$ and $\boldsymbol{F}$ are not identified without further restrictions. Therefore, we assume that a set of normalizing restrictions is available to ensure identification. The actual choice is not important; for example, for $r = 1$, one may normalise $f_T = 1$, or $g_0 = 1$. Therefore, in what follows $\boldsymbol{\theta}_\tau^0$ corresponds to the true parameter vector containing the normalized values of $\boldsymbol{f}^0$ and $\boldsymbol{g}^0$. □

Suppose that the null hypothesis H_0 is true. In this case, (1), (2), and (4) reduce to

$$\begin{aligned} y_{it} &= \boldsymbol{x}_{it}'\boldsymbol{\beta}_1^0 + \boldsymbol{\lambda}_i'\boldsymbol{f}_t^0 + \varepsilon_{it}, \quad (t = 1, \ldots, T) && (5) \\ \boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta}_1^0 + (\boldsymbol{I}_T \otimes \boldsymbol{\lambda}_i')\boldsymbol{f}^0 + \boldsymbol{\varepsilon}_i, && (6) \\ \boldsymbol{\mu}_1(\boldsymbol{\theta}_1) &= \boldsymbol{m} - \boldsymbol{M}\boldsymbol{\beta}_1 - \boldsymbol{S}\left(\boldsymbol{I}_T \otimes \boldsymbol{G}\right)\boldsymbol{f}, && (7) \end{aligned}$$

respectively, where $\boldsymbol{X}_i = \boldsymbol{X}_i^{(1)} + \boldsymbol{X}_i^{(\tau)}$, $\boldsymbol{M} = \boldsymbol{M}^{(1)} + \boldsymbol{M}^{(\tau)}$ and $\boldsymbol{\theta}_1 = (\boldsymbol{g}', \boldsymbol{f}', \boldsymbol{\beta}_1')'$. This setting under the null hypothesis of no structural break, is the one studied by Robertson and Sarafidis (2015).

To avoid possible ambiguities in regards to notation, we introduce the following notation for the model under the null hypothesis: $\boldsymbol{\theta}_R = (\boldsymbol{g}', \boldsymbol{f}', \boldsymbol{\beta}_1', \boldsymbol{\beta}_1')'$. Thus, $\boldsymbol{\theta}_R$ is the complete parameter $\boldsymbol{\theta}_\tau = (\boldsymbol{g}', \boldsymbol{f}', \boldsymbol{\beta}_1', \boldsymbol{\beta}_1',)'$ under H_0, which requires $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_\tau$; $\boldsymbol{\theta}_1 = (\boldsymbol{g}', \boldsymbol{f}', \boldsymbol{\beta}_1')'$ is a sub-vector of $\boldsymbol{\theta}_R$ and is also the parameter that defines the model under H_0. Note that $\{\boldsymbol{\theta}_\tau, \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\}$ are well defined for $\tau \geq 1$, but their definitions are different for $\tau = 1$ and for $\tau \geq 2$. It is instructive to illustrate the moment functions for a simple example.

Example 1. Suppose that $T = 3$, $r = 1$, and $\boldsymbol{x}_{it}$ is the one-period lagged scalar value $y_{i,t-1}$, such that $y_{i,t} = \beta_1 y_{i,t-1} + \lambda_i f_t + \varepsilon_{i,t}$. Thus, there are $T = 3$ observations over time and one unobserved factor. In this example, we denote the values of the factor over the three time points by $\{f_1, f_2, f_3\}$ instead of $\{f_{11}, f_{12}, f_{13}\}$. Let $\boldsymbol{z}_{i1} = (y_{i0})$,

$\boldsymbol{z}_{i2} = (y_{i0}, y_{i1})'$, and $\boldsymbol{z}_{i3} = (y_{i0}, y_{i1}, y_{i2})'$. Then, under the null hypothesis,

$$
\begin{aligned}
E(\boldsymbol{\mu}_{1,i}(\boldsymbol{\theta}_1)) &= \begin{pmatrix} E(\boldsymbol{z}_{i1}\varepsilon_{i1}) \\ E(\boldsymbol{z}_{i2}\varepsilon_{i2}) \\ E(\boldsymbol{z}_{i3}\varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} E(y_{i0}\varepsilon_{i1}) \\ E(y_{i0}\varepsilon_{i2}) \\ E(y_{i1}\varepsilon_{i2}) \\ E(y_{i0}\varepsilon_{i3}) \\ E(y_{i1}\varepsilon_{i3}) \\ E(y_{i2}\varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} m_{01} \\ m_{02} \\ m_{12} \\ m_{03} \\ m_{13} \\ m_{23} \end{pmatrix} - \beta_1 \begin{pmatrix} m_{00} \\ m_{01} \\ m_{11} \\ m_{02} \\ m_{12} \\ m_{22} \end{pmatrix} - \begin{pmatrix} g_0 f_1 \\ g_0 f_2 \\ g_1 f_2 \\ g_0 f_3 \\ g_1 f_3 \\ g_2 f_3 \end{pmatrix} \\
&\equiv \boldsymbol{m} - \beta_1 \boldsymbol{m}_{-1} - \boldsymbol{S}\text{vec}(\boldsymbol{G}\boldsymbol{F}') \qquad (8)
\end{aligned}
$$

where $\boldsymbol{\theta}_1 = (g_0, g_1, g_2, f_1, f_2, f_3, \beta_1)'$, $m_{st} = E(y_{is}y_{it})$, $\boldsymbol{m}_{-1} = E[\boldsymbol{Z}_i \boldsymbol{y}_{i,-1}]$ and $\boldsymbol{y}_{i,-1} = (y_{i,t-1})_{T\times 1}$. Observe that the moment conditions are ordered by the time-index t of the equations from which they are derived and then by the time-index s of the instruments.
On the other hand, under $H_{(\tau)}$ with $\tau = 3$, the moment conditions $E(\boldsymbol{\mu}_{3,i}(\boldsymbol{\theta}_3))$ are

$$
\begin{aligned}
&\begin{pmatrix} E(\boldsymbol{z}_{i1}\varepsilon_{i1}) \\ E(\boldsymbol{z}_{i2}\varepsilon_{i2}) \\ E(\boldsymbol{z}_{i3}\varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} E(y_{i0}\varepsilon_{i1}) \\ E(y_{i0}\varepsilon_{i2}) \\ E(y_{i1}\varepsilon_{i2}) \\ E(y_{i0}\varepsilon_{i3}) \\ E(y_{i1}\varepsilon_{i3}) \\ E(y_{i2}\varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} m_{01} \\ m_{02} \\ m_{12} \\ m_{03} \\ m_{13} \\ m_{23} \end{pmatrix} - \beta_1 \begin{pmatrix} m_{00} \\ m_{01} \\ m_{11} \\ 0 \\ 0 \\ 0 \end{pmatrix} - \beta_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ m_{02} \\ m_{12} \\ m_{22} \end{pmatrix} - \begin{pmatrix} g_0 f_1 \\ g_0 f_2 \\ g_1 f_2 \\ g_0 f_3 \\ g_1 f_3 \\ g_2 f_3 \end{pmatrix} \\
&\equiv \boldsymbol{m} - \beta_1 \boldsymbol{m}_{-1}^{(1)} - \beta_3 \boldsymbol{m}_{-1}^{(3)} - \boldsymbol{S}\text{vec}(\boldsymbol{G}\boldsymbol{F}'), \qquad (9)
\end{aligned}
$$

where $\boldsymbol{\theta}_3 = (g_0, g_1, g_2, f_1, f_2, f_3, \beta_1, \beta_3)'$. □

The vector-valued moment function can be simplified when $f_t = 1$ for all t. In this case, the factor component degenerates to a single individual-specific effect, and the last term in $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$ reduces to $\boldsymbol{S}(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_d)\boldsymbol{g}$, where $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones. Therefore, the proposed framework model incorporates the standard *fixed effects panel data model* as a special case. □

2.3 Main Results

Our test statistic builds upon the moment conditions introduced in the previous section. Therefore, our approach retains the traditional attractive feature of Method of Moments estimators in that it exploits only the orthogonality conditions implied by the model and does not require subsidiary assumptions such as normality of the error process.

Recall from the basic model equation (1) that the number of factors, r_0, is assumed known. While this may appear to be restrictive at first glance, in practice the asymptotic results for the test statistic obtained under the assumption that the number of factors r_0 is known, remain unchanged when r_0 is replaced by an estimate $\hat{r}$, so long as $\hat{r}$ is consistent. As an example, let $\hat{r}$ denote the value of r that minimises the following Bayesian Information Criterion [BIC]: $BIC(r) = N \times \widehat{Q}_\tau(r) - ln(N)/T^\xi \times \phi \times df(r)$, where $\widehat{Q}_\tau(r)$ denotes the minimum value of the objective function with r factors, $\xi \in (0,1)$, ϕ is a finite positive constant, and $df(r)$ denotes the number of degrees of freedom of the model, i.e. the number of moment conditions minus the number of estimable parameters. Then it is straightforward to show that $P(\hat{r} = r_0) \to 1$ (see Robertson and Sarafidis 2015). Therefore, it suffices to present asymptotic results for the case when r_0 is known. The small-sample performance of the test statistic when r_0 is estimated by $\hat{r}$ is evaluated in the simulation study of the paper.

The remainder of this section lists the other assumptions and establishes the main asymptotic results. To begin with, let Θ denote the parameter space that is obtained by a particular set of normalizing restrictions on $(\boldsymbol{G}, \boldsymbol{F})$. Let $\boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau) = E_{\boldsymbol{\theta}_\tau^0}[\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\boldsymbol{\mu}'_{\tau,i}(\boldsymbol{\theta}_\tau)]$ and $\boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau) = E_{\boldsymbol{\theta}_\tau^0}[(\partial/\partial\boldsymbol{\theta}'_\tau)\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)]$, $(\tau \geq 1)$. Let $\boldsymbol{\Phi}_\tau$ and $\boldsymbol{\Gamma}_\tau$ denote $\boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0)$ and $\boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0)$ respectively; note that $\boldsymbol{\Gamma}_1$ has fewer columns than $\boldsymbol{\Gamma}_\tau$, $(\tau \geq 2)$.

Assumption 2. *The parameter space Θ is compact, contains the true value $\boldsymbol{\theta}_\tau^0$ in its interior, and the population moment function $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$ is equal to $\boldsymbol{0}$ if and only if $\boldsymbol{\theta}_\tau = \boldsymbol{\theta}_\tau^0$ $(\tau \geq 2)$.*

Assumption 3. *The variance-covariance matrix* $\boldsymbol{\Phi}_\tau \equiv \boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0)$ *of the moment functions evaluated at* $\boldsymbol{\theta}_\tau^0$*, and* $\boldsymbol{\Gamma}_\tau \equiv \boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0)$*, the matrix of derivatives of the moment functions, both exist and have full rank* ($\tau \geq 1$)*.*

The aforementioned assumptions provide the main conditions to ensure consistency and asymptotic normality of the estimator proposed in this paper. Let $\boldsymbol{W}$ be a given positive definite weighting matrix and $Q_\tau(\boldsymbol{\theta}_\tau) = \boldsymbol{\mu}'_\tau(\boldsymbol{\theta}_\tau)\boldsymbol{W}\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$. Let $\widehat{\boldsymbol{W}}$ be a given consistent estimator of $\boldsymbol{W}$ under the null hypothesis, $\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1}\sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$, $\hat{\boldsymbol{\Phi}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1}\sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\boldsymbol{\mu}'_{\tau,i}(\boldsymbol{\theta}_\tau)$, and $\hat{Q}_\tau(\boldsymbol{\theta}_\tau) = \hat{\boldsymbol{\mu}}'_\tau(\boldsymbol{\theta}_\tau)\widehat{\boldsymbol{W}}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau)$. Define the GMM estimator $\hat{\boldsymbol{\theta}}_\tau$ of $\boldsymbol{\theta}_\tau^0$ by ($\tau \geq 1$)

$$\hat{\boldsymbol{\theta}}_\tau = \underset{\boldsymbol{\theta}_\tau \in \Theta}{\arg\min}\ \hat{Q}_\tau(\boldsymbol{\theta}_\tau). \tag{10}$$

The consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_\tau$ are established later in the paper. The optimal choice of the weighting matrix is obtained by setting $\widehat{\boldsymbol{W}} = \hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau)$ for the $\hat{Q}_\tau$ in (10) (see Hansen 1982). Since this requires an initial consistent estimate of $\boldsymbol{\theta}_\tau^0$, efficient estimation can be implemented in two stages: in the first stage one sets $\widehat{\boldsymbol{W}} = \boldsymbol{I}_\zeta$, which provides an initial consistent estimate $\hat{\boldsymbol{\theta}}_\tau^{(1)}$ of $\boldsymbol{\theta}_\tau^0$. This can be used in the second stage to obtain a consistent estimate of the inverse of the variance-covariance matrix of the moment conditions. Subsequently, one may rely on (10) with $\widehat{\boldsymbol{W}} = \hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)})$ if $\hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)})$ is non-singular, otherwise $\widehat{\boldsymbol{W}} = [\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_\tau^{(1)}) + N^{-1}\boldsymbol{I}]^{-1}$. Next, we state a lemma, the proof of which is given in the Supplementary Material.

Lemma 1. *Suppose that Assumptions 1-3 are satisfied. Then, as* $N \to \infty$*, we have (i)* $\hat{\boldsymbol{\theta}}_\tau \xrightarrow{p} \boldsymbol{\theta}_\tau^0$*, (ii)* $\sqrt{N}(\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}_\tau^0) = -(\boldsymbol{\Gamma}'_\tau\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{\Gamma}_\tau)^{-1}\boldsymbol{\Gamma}'_\tau\boldsymbol{\Phi}_\tau^{-1}\sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) + o_p(1)$*, (iii)* $\sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_\tau)$*, and (iv)* $\sqrt{N}(\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}_\tau^0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\boldsymbol{\Gamma}'_\tau\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{\Gamma}_\tau)^{-1})$.

In what follows we develop a distance-based test that resembles a 'likelihood ratio' type test, and an LM based test for a single structural break.

(A) Distance based test

In studies involving structural breaks, the case where the break point τ is known, and the case where it is unknown are both of interest. Hence, we will study both.

Recall that $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_\tau$ are estimators of the true value of the parameter under the null H_0 and under the alternative $H_{(\tau)}$ respectively, obtained by minimizing the same quadratic distance function. Therefore, a suitable *distance based* statistic for testing H_0 vs $H_{(\tau)}$ is the following reduction in the quadratic distance function:

$$D_\tau = N[\hat{Q}_1(\hat{\boldsymbol{\theta}}_1) - \hat{Q}_\tau(\hat{\boldsymbol{\theta}}_\tau)]. \tag{11}$$

This type of test statistics are also sometimes referred to as *likelihood ratio type* [D-type] statistics, because of their resemblance to likelihood ratio statistic.

To provide an asymptotic representation of the test statistic D_τ and deduce its asymptotic distribution, under the null hypothesis, let us introduce the following notation: For a full column rank matrix, $\boldsymbol{B}$, let $\boldsymbol{M}_B$ and $\boldsymbol{P}_B$ be two projection matrices defined by $\boldsymbol{P}_B \equiv \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'$ and $\boldsymbol{M}_B \equiv \boldsymbol{I} - \boldsymbol{P}_B$ respectively. For a positive definite symmetric matrix $\boldsymbol{A}$, let $\boldsymbol{A}^{-1/2}$ denote the symmetric square root of $\boldsymbol{A}$, so that $\boldsymbol{A} = \boldsymbol{A}^{-1/2}\boldsymbol{A}^{-1/2}$. Let $\boldsymbol{V}_\tau(\boldsymbol{\theta}_\tau^0) = \boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1} - \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau}$, where all the quantities are evaluated at the true value of $\boldsymbol{\theta}_\tau$ under the null hypothesis. Let $\boldsymbol{z}_N = \sqrt{N}\boldsymbol{\Phi}_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0)$. Then, it follows from Lemma 1 that, under the null hypothesis, $\boldsymbol{z}_n \xrightarrow{d} \boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_\zeta)$. The distance-type and LM-type test statistics proposed in this paper turn out to be continuous functions of $\boldsymbol{z}_N$, except for an additive $o_p(1)$ term. Consequently, the asymptotic null distributions of these test statistics can be deduced by the continuous function theorem. The proof of the following theorem is given in the Appendix.

Theorem 1. *Suppose that Assumptions 1-3 hold, $\tau \geq 2$, and that the null hypothesis H_0 is satisfied. Then $D_\tau = \boldsymbol{z}_N' V_\tau \boldsymbol{z}_N + o_p(1)$, where V_τ is a projection matrix of rank dim($\boldsymbol{\beta}_\tau$), and hence D_τ is asymptotically distributed as chi-squared with degrees of freedom equal to dim($\boldsymbol{\beta}_\tau$).*

Next, we use this to develop a test of H_0 against the more general alternative H_1 wherein the breakpoint τ_0 is unknown, but is known to lie in $\{\tau_1, \ldots, \tau_L\} \subseteq \{2, \ldots, T\}$. To this end, we consider the foregoing test for each possible value of τ in $\{\tau_1, \ldots, \tau_L\}$, and then combine them. Let $\boldsymbol{D} = [D_{\tau_1}, \ldots, D_{\tau_L}]'$, where $D_{\tau_\ell} =$

$N[\hat{Q}_1(\hat{\boldsymbol{\theta}}_1) - \hat{Q}_{\tau_\ell}(\hat{\boldsymbol{\theta}}_{\tau_\ell})]$ is the statistic proposed in (11) for testing H_0 against $H_{(\tau_\ell)}$. Clearly, if H_0 is true then $\hat{Q}_1(\hat{\boldsymbol{\theta}}_1)$ and $\{\hat{Q}_{\tau_\ell}(\hat{\boldsymbol{\theta}}_{\tau_\ell}), \ell = 1, \dots, L\}$ are all estimators of the same quantity and hence $\max\{D_{\tau_1}, \dots, D_{\tau_L}\}$ is expected to be small. On the other hand, if H_1 is true, then $\hat{Q}_{\tau_0}(\hat{\boldsymbol{\theta}}_{\tau_0})$ is expected to be smaller than $\hat{Q}_1(\hat{\boldsymbol{\theta}}_1)$ and hence $\max\{D_{\tau_1}, \dots, D_{\tau_L}\}$ is expected to be large, assuming that $\tau_0 \in \{\tau_1, \dots, \tau_L\}$. Therefore, in order to test H_0 vs H_1 we propose to use the statistic

$$D_{\max} := \max_{\ell=1,\dots,L} \{D_{\tau_\ell}\}, \tag{12}$$

and reject the null for large values of $D_{\max}$; this is essentially the test based on the *Union-Intersection Principle* (for example, see Section 5.2 in Silvapulle and Sen 2011). Further, we propose to estimate the unknown break point τ_0 by $\hat{\tau}$, where $D_{\hat{\tau}} = D_{\max}$; The next theorem provides the essential result for applying D_{max} for testing H_0 against H_1.

Theorem 2. *Suppose that Assumptions 1-3 and the null hypothesis H_0 are satisfied. Let $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_\zeta)$ where ζ is the number of moment conditions, and let $\boldsymbol{V}_{\tau_1}, \dots, \boldsymbol{V}_{\tau_L}$ be evaluated at the true value of the parameter specified by the null hypothesis. Then, $(D_{\tau_1}, \dots, D_{\tau_L}) = \big(\boldsymbol{z}_N' \boldsymbol{V}_{\tau_1} \boldsymbol{z}_N, \dots, \boldsymbol{z}_N' \boldsymbol{V}_{\tau_L} \boldsymbol{z}_N\big) + o_p(1) \xrightarrow{d} \big(\boldsymbol{z}' \boldsymbol{V}_{\tau_1} \boldsymbol{z}, \dots, \boldsymbol{z}' \boldsymbol{V}_{\tau_L} \boldsymbol{z}\big)$. Consequently, $D_{\max} = \max\{\boldsymbol{z}_N' \boldsymbol{V}_{\tau_1} \boldsymbol{z}_N, \dots, \boldsymbol{z}_N' \boldsymbol{V}_{\tau_L} \boldsymbol{z}_N\} + o_p(1)$, and hence $D_{\max}$ is asymptotically distributed as $\max\{\boldsymbol{z}' \boldsymbol{V}_{\tau_1} \boldsymbol{z}, \dots, \boldsymbol{z}' \boldsymbol{V}_{\tau_L} \boldsymbol{z}\}$.*

The asymptotic null distribution of D_{max} depends on the nuisance parameter $\boldsymbol{\theta}_1^0$ through $\boldsymbol{V}_{\tau_\ell}(\boldsymbol{\theta}_1^0)$. We propose to estimate the distribution of the test statistic D_{max} by that of $\max\{\boldsymbol{z}' \boldsymbol{V}_{\tau_1}(\boldsymbol{\theta}_1) \boldsymbol{z}, \dots, \boldsymbol{z}' \boldsymbol{V}_{\tau_L}(\boldsymbol{\theta}_1) \boldsymbol{z}\}$ at $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_1$. Therefore, critical values can be obtained by simulation, using the following steps:

(a) Generate one observation of $\boldsymbol{z}$ from $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_\zeta)$, where $\boldsymbol{z}$ is $\zeta \times 1$;

(b) Compute $\hat{c} = \max\{\boldsymbol{z}' \hat{\boldsymbol{V}}_{\tau_1}(\hat{\boldsymbol{\theta}}_1) \boldsymbol{z}, \dots, \boldsymbol{z}' \hat{\boldsymbol{V}}_{\tau_L}(\hat{\boldsymbol{\theta}}_1) \boldsymbol{z}\}$, where $\hat{\boldsymbol{\theta}}_1$ is the estimator of $\boldsymbol{\theta}_\tau^0$, under the null hypothesis;

(c) Repeat steps (a)-(b) n times, say $n = 10,000$, and generate n values of $\hat{c}$, which we denote as $\hat{c}_1, \dots, \hat{c}_n$.

(d) Let $\hat{c}_{(0.95)}$ be the 95^{th} percentile of $\{\hat{c}_1, \dots, \hat{c}_n\}$.

Note that $\boldsymbol{V}_{\tau_\ell}(\boldsymbol{\theta}_\tau)$ is continuously differentiable and $\hat{\boldsymbol{\theta}}_1$ is consistent for $\boldsymbol{\theta}_\tau^0$ under H_0 for $\tau = 1$. Therefore, the distribution of $\boldsymbol{z}'\boldsymbol{V}_{\tau_\ell}(\boldsymbol{\theta}_\tau^0)\boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_\zeta)$, can be approximated by that of $\boldsymbol{z}'\hat{\boldsymbol{V}}_{\tau_\ell}(\hat{\boldsymbol{\theta}}_\tau)\boldsymbol{z}$, where $\hat{\boldsymbol{\theta}}_\tau$ is a consistent estimator of $\boldsymbol{\theta}_\tau^0$ under H_0. This in turn provides the justification for the method employed in our numerical study to estimate the critical value of the test statistic at the estimated null value.

(B) Lagrange Multiplier (LM) Test

In this subsection, we use the building blocks of the distance based test in the previous subsection to develop a Lagrange Multiplier (LM) test. The LM test has advantages and disadvantages compared to the foregoing distance based tests. A feature of LM and score type tests is that their implementation requires estimation only of the restricted model under the null hypothesis, but not the unrestricted full model. If it is desired to avoid estimation of the full model, then this feature is important. On the other hand, as it is well-known, LM test statistics are not invariant under change of parametrization.

Let $\tau \geq 2$ be given, $\hat{\boldsymbol{\Gamma}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1}\sum_{i=1}^{N}(\partial/\partial\boldsymbol{\theta}_\tau')\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$,

$$\hat{\boldsymbol{A}}_\tau = [\hat{\boldsymbol{\Gamma}}_\tau(\hat{\boldsymbol{\theta}}_R)]'[\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_R)]^{-1}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_R); \quad \hat{\boldsymbol{U}}_\tau = [\hat{\boldsymbol{\Gamma}}_\tau(\hat{\boldsymbol{\theta}}_R)]'[\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_R)]^{-1}[\hat{\boldsymbol{\Gamma}}_\tau(\hat{\boldsymbol{\theta}}_R)], \tag{13}$$

where $\hat{\boldsymbol{\theta}}_R$ is the estimator of $\boldsymbol{\theta}_\tau^0$ under H_0. Then the LM-statistic, LM_τ, for testing H_0 against $H_{(\tau)}$ is[3]

$$LM_\tau = N\hat{\boldsymbol{A}}_\tau'\hat{\boldsymbol{U}}_\tau^{-1}\hat{\boldsymbol{A}}_\tau. \tag{14}$$

Theorem 3 stated below shows that the asymptotic null distribution of LM_τ is the same as that of D_τ.

Theorem 3. *Let the setting be as in Theorem 1 and $\tau \geq 2$. Then $LM_\tau = \boldsymbol{z}_N' V_\tau \boldsymbol{z}_N + o_p(1) = D_\tau + o_p(1)$, and hence LM_τ is asymptotically distributed as chi-squared with degrees of freedom equal to $dim(\boldsymbol{\beta}_\tau)$.*

Next, as in the setting of Theorem 2, let us consider testing H_0 against H_1 wherein the breakpoint τ_0 is unknown, but is known to lie in $\{\tau_1, \ldots, \tau_L\} \subseteq \{2, \ldots, T\}$.

[3]For example, see section 4.5.3 in Silvapulle and Sen (2011); (2.9) in Newey and West (1987).

By motivations similar to those for D_{max}, we propose the test statistic $LM_{max} = \max_{\ell=1,\dots,L}\{LM_{\tau_\ell}\}$, and reject the null for large enough values of LM_{max}.

Theorem 4. *Let the setting be as in Theorem 2. Then,* $LM_{\max} = D_{max} + o_p(1) \xrightarrow{d} \max\{\boldsymbol{z}'\boldsymbol{V}_{\tau_1}\boldsymbol{z}, \dots, \boldsymbol{z}'\boldsymbol{V}_{\tau_L}\boldsymbol{z}\}$ *as* $N \to \infty$.

The test based on LM_{max} can be implemented by the simulation method for D_{max} outlined earlier.

3 Monte Carlo Simulations

3.1 Simulation Design

This section investigates the finite sample properties of the tests introduced previously. Our focus is on the impact of sample size, and the location and magnitude of the break on the performance of our tests. We study the pure panel AR(1) with one-factor; the choice of this model was motivated mainly by the application presented in Section 4. The DGP is

$$y_{it} = \begin{cases} \beta_1^0 y_{i,t-1} + \lambda_i f_t^0 + \varepsilon_{it}; & t = 1, \dots, \tau - 1; \\ \beta_\tau^0 y_{i,t-1} + \lambda_i f_t^0 + \varepsilon_{it}; & t = \tau, \dots, T, \end{cases} \tag{15}$$

for $i = 1, \dots, N$, where $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2)$, $\sigma_{it}^2 \sim U[0, 2]$, $\lambda_i \sim \mathcal{N}(0, \sigma_i^2)$, $\sigma_i^2 = \sigma_\lambda^2 s_i$, $s_i \sim U[0, 2]$, $f_t^0 \sim \mathcal{N}(0, \sigma_f^2)$, and $\sigma_\lambda^2 = \sigma_f^2 = 1$. We consider the case when the true number of factors, r_0, is known and also the case when it is unknown.

The initial observation is generated as $y_{i0} = \lambda_i/(1 - \beta_1^0) + \mathcal{N}(0, 1)$. The vector of possible instruments is $\boldsymbol{w}_i = (y_{i0}, \dots, y_{i,T-1})'$. For each t, we choose $\boldsymbol{z}_{it} = (y_{i0}, \dots, y_{i,t-1})'$. We set $N \in \{100, 300, 600\}$, $T \in \{6, 9\}$, $\tau_0 \in \{4, 6\}$, and $\omega^0 \equiv \beta_\tau^0 - \beta_1^0 \in \{0, 0.10, 0.15\}$. We fix $\beta_1^0 = 0.50$. All simulations are conducted using $5,000$ replications.

3.2 Simulation Results

Number of factors is known

Table 1 provides the percentage of times that the null hypothesis was rejected when the breakpoint location τ_0 and the number of factors r_0 are known. Recall that $\omega^0 = \beta_\tau^0 - \beta_1^0$. Therefore, the entries under $\omega^0 = 0$ are the empirical sizes of the tests; the other entries are estimated powers. Nominal level is fixed at 5% throughout.

Table 1 shows that the performance of the $D_\tau-$test is quite satisfactory. For example, the Type I error rate is close to the nominal level of 5% for every case; further, the power of test increases close to 100% as the number of observations and/or the size of the break ω^0 increases. These observations are consistent with the general result that the $D_\tau-$test is valid and consistent.

While the large sample behavior of the $LM_\tau-$test is similar to that of the $D_\tau-$test, there is one departure that is worthy of note. For $T = 6$ and $N = 100$ (i.e. small N), the $LM_\tau-$test is substantially oversized, with the Type I error rate being approximately 16% while the nominal level is only 5%. However, as N increases to 300 and 600 with T fixed at 6, the Type I error rate moves very close to the nominal level 5%. We note that similar size distortions associated with large values of T are well-known in the dynamic panel data literature (c.f. Bun and Sarafidis 2015); the main reason is that as T increases, the total number of moment conditions available becomes larger, which may distort inference unless N is large. An obvious way to alleviate such size distortions may appear to be to reduce the number of moment conditions used when N is relatively "small"; exactly, what is meant by "small" is difficult to quantify and we did not explore this approach. Based on the simulation results in Table 1, the indications are that D_τ-test is better than the LM_τ-test.

For $T = 6$, the power of both D_τ- and LM_τ-tests are substantially smaller when τ_0 is also equal to 6, compared to the case when $\tau_0 = 4$. This is to be expected since in the former case the break takes place at the very end of the sample period, and therefore the only moment conditions available to detect the break are those in the last sample period; there are only six of them in this case. By contrast, when $T = 9$, the break point $\tau_0 = 6$ no longer corresponds to the final period of the sample, and hence the difference between $\tau_0 = 6$ and $\tau_0 = 4$ in terms of power is smaller, as expected.

Table 1: **Frequencies (%) that the null hypothesis was rejected when the breakpoint τ_0 and the number of factors r_0 are known.**

		$\tau_0 = 4$						$\tau_0 = 6$					
		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$	
T	N	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ
6	100	6.3	6.7	10.8	11.4	19.2	19.5	6.2	6.9	6.7	6.0	7.5	8.0
	300	5.0	4.8	29.7	29.0	55.8	55.1	5.9	5.7	9.4	8.9	14.9	14.2
	600	5.0	4.9	53.0	52.7	81.8	80.4	5.2	5.3	14.8	13.7	26.3	24.8
9	100	7.2	16.4	12.8	12.7	23.1	23.7	6.7	15.9	15.6	14.2	29.8	28.4
	300	5.3	5.2	43.0	43.7	75.0	75.5	5.5	5.8	47.9	47.2	80.0	77.9
	600	6.1	6.3	70.8	71.2	94.5	94.5	5.90	6.0	79.0	78.2	96.0	95.0

Notes: The value of β_1^0 was fixed at 0.5, and the value of $\omega^0 := \beta_{\tau_0} - \beta_1$ was varied. Therefore, the estimates under $\omega^0 = 0.00$ correspond to size, and the other ones correspond to power.

Next, we consider the case when the breakpoint τ_0 is unknown. For this case, we present the results only for $T = 6$ and $\tau_0 = 4$. The results for the other values of T and τ_0 do not add much; they are available upon request [4]. The performance of the tests in terms Type I error rate, power, and accuracy of $\hat{\tau}$ as an estimator of τ_0 is summarised in Table 2. The Type I error rates are similar to the corresponding estimates in Table 1 and hence are quite satisfactory, and the power is slightly lower, which is expected since the breakpoint is unknown. The performance of the Distance

[4] Vasilis: It would be better to present them in the Supplementary Materials.

test relative to the LM test, in terms of size and power, appears to be similar that for the case when the break point is known. The frequency of detecting the correct break point τ_0 increases with N and ω^0, as expected. [5]

Table 2: **Rejection Frequencies, unknown breakpoint, r_0 known, $T = 6$, $\tau_0 = 4$.**

		D					*LM*				
			When null rejected, breakpoint detected at (%)					When null rejected, breakpoint detected at (%)			
ω^0	N		$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$		$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$
0.00	100	7.10	-	-	-	-	6.74	-	-	-	-
	300	5.72	-	-	-	-	5.22	-	-	-	-
	600	5.00	-	-	-	-	5.04	-	-	-	-
0.10	100	7.94	15	33	16	36	8.30	21	35	21	23
	300	21.52	12	59	14	16	22.78	13	54	15	17
	600	42.5	11	67	12	11	42.28	9	66	13	12
0.15	100	14.76	14	42	14	30	15.10	17	46	16	21
	300	44.46	9	68	12	11	44.52	9	68	12	11
	600	75.82	7	78	9	7	73.78	6	78	8	8

For power ($\omega^0 \neq 0.00$) the entries under different values of τ in each row sum up to 100%. As an example, consider the 4^{th} row of results, corresponding to $N = 100$ and $\omega^0 = 0.10$, in which empirical power is 5.78 %. The interpretation of this row of results is that among the 5.78% of the 5000 samples that rejected the null hypothesis of no structural break, 29%, 19%, 20%, and 31% of the samples estimated the breakpoint to be at $\tau = 4, 3, 5, 6$ respectively.

Number of factors is unknown

In practice, the number of factors is likely to be unknown. This section reports simulation results for both D and LM tests when the number of factors is estimated

[5] Vasilis: Please see the suggested format, Table 8 for Table 2, at the end. Note that, tables are required to be as self-contained as possible.

$\hat{r}$, the value of r that minimizes BIC over $r = 0, 1, 2$, with $\xi = 0.3$ and $\phi = 0.75$ (see Section 3.2). [6] Thus, the differences between Tables 1 and 3 (or between Tables 2 and 4), highlights the impact of estimating r_0.

A major difference between the D and LM tests is that the former uses the $\hat{r}$ obtained by minimizing the BIC value corresponding to the model under H_1, while in the latter case $\hat{r}$ is obtained by minimizing the BIC value corresponding to the model under H_0. This ensures that the LM test does not use any sample estimates computed for the model H_1.

Table 3 presents results on rejection (i.e. rejection of the null hypothesis) frequencies (%) when the breakpoint is known. The performance of the D-test is very similar to that reported in Table 1; therefore, the effect of estimating the number factors is small. The same applies to the performance of LM with respect to *size*. However, power of the LMtest appears to be substantially lower than that of the D-test. Moreover, for $T = 9$ and $\omega^0 = 0.15$, power fails to increase as N increases from $N = 300$ to $N = 600$.

We investigated the underlying reason for this undesirable bahaviour of the LM-test. We observed that when $\hat{r}$ is obtained by minimizing the BIC value corresponding to the model under H_0, there is an identification problem in that when H_0 is violated, the structural break tends to be absorbed by an additional factor. To illustrate, consider the example provided in Section 2.2. Suppose that there is a structural break and hence H_0 is not true. In this case, imposing the null under the assumption that the number of factors is unknown, leads to the following alternative representation of the model:

$$E(\boldsymbol{\mu}_{3,i}(\boldsymbol{\theta}_3)) = \boldsymbol{m} - \beta_1 \boldsymbol{m}_{-1} - \boldsymbol{S}\text{vec}(\boldsymbol{G}\boldsymbol{F}') - (\beta_3 - \beta_1)\, \boldsymbol{m}_{-1}^{(3)}, \tag{16}$$

where $\boldsymbol{G}\boldsymbol{F}'$ takes exactly the same form as in (9), whereas the last term on the right-hand side can be captured by an additional factor component $\tilde{f}_t$; [7] this is achieved

[6] The finite sample properties of the BIC with these parameter values were investigated by Robertson and Sarafidis (2015).

[7] Vasilis: This is a really nice insight of yours! Well-done. Please check my changes, because the

by $\tilde{f}_1 = \tilde{f}_2 = 0$, $\tilde{f}_3 = (\beta_3 - \beta_1)$, $\tilde{g}_0 = m_{02}$, $\tilde{g}_1 = m_{12}$ and $\tilde{g}_2 = m_{22}$. Therefore, in this example, when the true model violates H_0 and r_0 is estimated under H_0, the structural break can be absorbed by an additional factor. In consequence, BIC may select $\hat{r} = 2$, in which case H_0 is not likely to be rejected. As a result, the LM-test is likely to under-reject the null hypothesis. In our simulations, we observed that the problem was more pronounced when the deviation of ω^0 from zero becomes larger. Intuitively, this is because when $\beta_3 - \beta_1 = 0$, [8] the last term in (16) is eliminated.

An implication of the above result is that in order to ensure consistency of the LM-test in the case where r_0 is unknown, $\hat{r}$ needs to be obtained based on the BIC value corresponding to the model under H_1. However, if we were to do so, the appeal of the LM-test relative to D-test diminishes.[9]

Table 3: **Rejection Frequencies (%), known breakpoint, r_0 unknown.**

		$\tau_0 = 4$, $\beta_1^0 = 0.5$						$\tau_0 = 6$, $\beta_1^0 = 0.5$					
		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$	
T	N	D	LM	D	LM	D	LM	D	LM	D	LM	D	LM
6	100	6.2	5.7	13.2	11.2	21.4	16.8	8.9	6.3	10.3	6.9	11.3	7.6
	300	4.9	4.5	28.5	24.2	54.5	39.5	5.8	5.4	11.2	9.4	17.3	13.5
	600	4.9	4.8	52.8	44.0	80.6	48.2	5.1	5.1	15.5	14.3	28.0	23.6
9	100	6.9	14.3	14.9	21.9	25.3	30.5	7.0	14.6	16.7	22.2	29.2	31.7
	300	5.3	4.7	43.2	38.3	73.7	58.1	5.6	5.4	49.4	42.3	81.3	56.3
	600	6.2	6.1	72.2	63.3	95.1	58.8	5.9	5.8	80.1	65.5	96.4	46.1

Note: $\omega^0 = 0.00$ corresponds to size, otherwise power.

details need to be bit more precise.

[8] Vasilis: Please check. This is unclear. Do you mean $\beta_3 - \beta_1 \neq 0$?

[9] We have conducted further simulations confirming that the power properties of LM are similar to those of D when they use the same $\hat{r}$ based on the model under H_1. The results are available upon request.

Table 4 reports results on rejection frequencies when the breakpoint and the number of factors are both unknown. The conclusions are qualitatively similar in that the performance of the D test remains similar to that when r_0 is known (see Table 2 for comparison). On the other hand, the power of the LM test is much lower than that of the D-test; the reason for this is likely to be the aforementioned identification problem.

Table 4: **Rejection Frequencies, unknown breakpoint, r_0 unknown, $T = 6$, $\tau_0 = 4$.**

		D					LM				
			When null rejected, breakpoint detected at (%)					When null rejected, breakpoint detected at (%)			
ω^0	N		$\hat{\tau} = 3$	$\hat{\tau} = 4$	$\hat{\tau} = 5$	$\hat{\tau} = 6$		$\hat{\tau} = 3$	$\hat{\tau} = 4$	$\hat{\tau} = 5$	$\hat{\tau} = 6$
0.00	100	7.46	-	-	-	-	5.16	-	-	-	-
	300	5.68	-	-	-	-	4.38	-	-	-	-
	600	5.12	-	-	-	-	4.64	-	-	-	-
0.10	100	9.98	17	35	14	34	5.80	21	30	24	24
	300	20.98	14	55	16	15	18.80	12	55	15	17
	600	42.46	10	68	13	9	34.52	10	63	12	15
0.15	100	15.24	15	43	17	26	9.06	13	41	21	25
	300	44.22	9	67	11	13	31.38	9	61	14	16
	600	74.28	6	78	9	7	45.40	7	67	13	13

For power ($\omega^0 \neq 0.00$) the entries under different values of $\hat{\tau}$ in each row sum up to 100%. As an example, consider the 4$^{\text{th}}$ row of results, corresponding to $N = 100$ and $\omega^0 = 0.10$, in which empirical power is 9.98 %. The interpretation of this row of results is that among the 5.78% of the 5000 samples that rejected the null hypothesis of no structural break, 17%, 35%, 14%, and 34% of the samples estimated the breakpoint to be at $\tau = 3, 4, 5, 6$ respectively.

4 Empirical Illustration

In this section we apply our method to investigate the empirical validity of the well-known "Law of Proportionate Effect", also known as Gibrat's 'Law', for the US finance industry. Gibrat's law postulates that the size of a firm and its growth rate are independent.

Consider the model $y_{it} = \beta^0 y_{i,t-1} + u_{it}$, where y_{it} denotes some measure of size (expressed in natural logarithms) for firm i at time t ($i = 1, \ldots, N; \quad t = 1, \ldots, T$). Subtracting $y_{i,t-1}$ from both sides yields $\Delta y_{it} = \delta^0 y_{i,t-1} + u_{it}$, where $\delta^0 = \beta^0 - 1$. The growth rate of firm i at time t is Δy_{it}. Gibrat's 'Law' implies the restriction $\beta^0 = 1$, or equivalently $\delta^0 = 0$. To see this more closely, notice that for $\beta^0 = 1$ firm size is a random walk and so it can be expressed as $y_{it} = y_{i0} + \sum_{s=1}^{t} u_{is}$. Since the first term therein, y_{i0}, does not depend on t, it is clear that the growth rate of firm i, Δy_{it}, is white noise.

Gibrat's 'Law' has attracted considerable attention in economics (see e.g. Santarelli et al. 2006), mainly because it is consistent with an empirical regularity observed across several industries, namely that the distribution of firm size is often highly skewed to the right. This is due to the fact that many industries consist of a small (respectively, large) number of big (respectively, small or medium-sized) firms. In addition, as pointed out by Simon and Bonini (1958), there is also a connection between Gibrat's 'Law' and the returns to scale in a given industry. In particular, under constant returns to scale, the probability of a given firm increasing in size relative to its existing size is expected to be constant across all firms in the industry that lie above a critical minimum size.

Our data set spans 13 years $(2002 - 2014)$ and contains observations on $4,128$ banking institutions. The error term in our model is assumed to obey the multi-factor structure,

$$u_{it} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t^0 + \varepsilon_{it}. \tag{17}$$

Allowing for a common factor component is important in the present case for several reasons. For instance, a subset of the factors may be adequate to capture the effect of the age of the banking institution since 'age' is not observed in our sample. For example, let x_{it} denote the variable age and β_i denote the bank-specific age effect. Since age increases at the same rate for all companies every year, one may set

$$\beta_i x_{it} = \beta_i(x_{i0} + t) = \gamma_i + \beta_i t, \tag{18}$$

where x_{i0} denotes the age of the bank at the beginning of the sample period, and $\gamma_i \equiv \beta_i x_{i0}$. Thus, the effect of the unobserved variable age can be captured using two factors, one of which is constant over time and resembles a fixed effect, while the other one is a deterministic time trend. Since we do not know a priori whether age has a linear effect or not, it is prudent not to impose specific restrictions on how the factors vary over time. [10]

The common factor approach may capture the presence of common shocks, such as the GFC, that have hit all individual banks, albeit with different intensities. Indeed, due to the pervasiveness of the recent GFC, it is important to be able to investigate whether the parameter β^0 appears to be constant over time. Therefore, the application of our methodology to this research question is particularly relevant.

We consider the monetary value of assets expressed in constant prices as a measure of bank size. The results are reported in Table 5.[11] The null hypothesis for the Wald-test is $\beta_1^0 = 1$. Let r denote the number of factors fitted in the model, while β_1^0 and β_τ^0 denote the value of the autoregressive parameters prior to and after the break, respectively. Hence the magnitude of the break is $\beta_\tau^0 - \beta_1^0$. Let P_W, P_J, and $P_{D_{\max}}$ denote the p-values for the Wald-test, overidentifying restrictions test (Hansen's test), and the structural break test proposed in this paper, respectively. The number of

[10] In practice, it is impossible to distinguish between the effect of age and the presence of a separate deterministic linear trend. Our aim is not to identify these factors per se, but rather to identify whether the parameter δ^0 is subject to a break or not, controlling for a flexible form of unobserved heterogeneity. Failing to do so may result in invalid inferences.

[11] To avoid possible confusion, we only report results for D. The results for LM are similar, subject to the caveat discussed in the simulations. These results are available upon request.

factors was selected using the BIC model information criterion. [12] In this case, there are 78 moment conditions in total.

Table 5: **Evaluation of the Gibrat's 'Law' for Banks in the USA**

		Est.	S.E.	W	P_t	[95% C. I.]	P_J	$P_{D_{\max}}$	Break Date	BIC
$r = 1$	β_1^0	0.60	0.02	25	0.00	$[0.55, 0.65]$	0.00	0.00	2010	-56
	β_τ^0	0.72	0.03	24	0.00	$[0.66, 0.78]$				
$r = 2$	β_1^0	0.48	0.08	5.9	0.00	$[0.32, 0.63]$	0.68	0.00	2012	-81
	β_τ^0	0.99	0.08	12	0.00	$[0.82, 1.15]$				
$r = 3$	β_1^0	0.41	0.13	3.1	0.00	$[0.15, 0.67]$	0.99	0.00	2011	-50
	β_τ^0	0.94	0.28	4.4	0.00	$[0.39, 1.49]$				

[a.] The sample contains data for 4,128 US bank institutions, spanning the period 2002-2014.

[b.] Est: Estimate of the relevant β parameter; S.E.: The Standard Error of the estimate; P_t: p-value for the t-statistic; C.I.: Confidence Interval; P_J: the p-value for the Hansen's test'; $P_{D_{\max}}$: the p-value of the structural break test; 'Break Date': the estimated year of the structural break; BIC: Bayesian Information Criterion for choosing the number of factors.

[c.] W denotes the Wald statistic.

Based on the results of Table 5, the optimal number of factors is two (i.e. $\widehat{r} = 2$), which is the value corresponding to the smallest BIC value based on the unrestricted model. Notice that in the two-factor model, the null hypothesis for the validity of the instruments is not rejected by the overidentifying restrictions test and therefore the model appears to be correctly specified. The null hypothesis of no structural break is rejected [13] at the 1% level of significance and the break is estimated to have occurred in 2012, which may be regarded as the end of the GFC. In particular, the

[12] Vasislis: Better to mention that number of factors was selected using the BIC under H_1, assuming that is true; otherwise we may have an inconsistent $\hat{r}$.

[13] rejected by which test ?

null hypothesis $\beta_1^0 = 1$ is rejected prior to the GFC (against the alternative $\beta_1^0 < 1$) but it is not rejected afterwards [14] This suggests that during the period 2002-2011, the growth rate of financial institutions was negatively correlated with their size, and small banks grew at a higher rate than large banks. On the other hand, the break in 2012 is such that the new value of the autoregressive parameter provides empirical support towards Gibrat's 'Law'.

One reason for such a development might be the establishment of the so-called 'Basel III' (or the Third Basel Accord) capital regulatory framework in 2010-2011. This is a global, voluntary regulatory framework on bank capital adequacy, stress testing and market liquidity risk, which was agreed upon by the members of the Basel Committee on Banking Supervision in 2010-–11. In particular, in order to prevent a further collapse of the financial sector during a potential future GFC, governments around the world decided to introduce more stringent capital requirements. It can be expected that higher capital requirements may make banks better able to absorb losses on their own resources. In response, banks did appear to change their behaviour by raising equity, cutting down lending, and reducing asset risk; as it has been argued in the relevant literature, well-capitalized banks managed to perform better during and after the GFC (e.g. Demirguc-Kunt et al. 2013). Overall, our results provide support for the claim that following the establishment of the 2011 capital regulatory framework, the growth of financial institutions depends more on capitalized structure than on size.

5 Discussion

Our proposed estimation and testing procedures remain valid in a pure AR(1) model, even when the autoregressive coefficient is equal to one, provided that the matrix of

[14] Vasilis: Is the claim 'not rejected' based on a proper test, if so, we need to make that clear. The table shows that the p-value P_t is zero for every case; something is not clear to me. When $r = 2$ the estimate of β_τ^0 is 0.99 (se=0.08). How was $W = 12$ obtained? Earlier, it was mentioned that the null for the Wald test is $\beta_1^0 = 1$. Is this related to the column of W in the table?

derivatives of the moment functions has full rank (Assumption 3 in this paper). The full-rank assumption on the Jacobian matrix essentially implies that identification of a unit root process requires that T is fixed and, in addition, some factors are 'genuine', in the sense that not all factors reduce to $f_t = 1$ for all t, which is the one-way error components model.[15]

As an anonymous referee pointed out, at first glance, the foregoing framework might appear to be a special case of the general framework of Andrews (1993), although this is not the case. The setting in Andrews (1993) is for time series data with one observation at each time point (i.e. $N = 1$) and $T \to \infty$, whilst in our setting T is fixed and $N \to \infty$. Consequently, the two settings are based on different model assumptions and different structures. For example, in the former case, moment functions are estimated by averaging over time using terms of the form $T^{-1} \sum_{t=1}^{T} (\cdot)$ and then letting $T \to \infty$. By contrast, in our setting the moment functions are estimated by averaging over i using terms of the form $N^{-1} \sum_{i=1}^{N} (\cdot)$ and letting $N \to \infty$. As a result, some differences are worth noting: (i) in the present paper no restrictions are placed on the inter-temporal variation of the data, other than those in Assumption 1. For example, the data may exhibit deterministic or stochastic trends, for example, $f_t = t$ is allowed. By contrast, in Andrews (1993) it is assumed that the data do not exhibit deterministic or stochastic trends (see top of page 822 therein); (ii) in Andrews (1993) the structural break date is assumed to be bounded away from the end of the sample in order to ensure that there is an adequate number of observations to estimate the parameters appearing after the structural break. By contrast, in our theoretical setting, since T is fixed and $N \to \infty$, estimation of the parameters of interest is feasible even if the structural break takes place at the end of the sample period.

The results obtained in Section 2 can also be extended in other directions. An anonymous referee suggested that it would be of interest to compare and contrast the present setting with the more flexible model $y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta}_t + \boldsymbol{\lambda}'_i \boldsymbol{f}_t + \varepsilon_{it}$, which

[15]Robertson et al. (2018) extend Robertson and Sarafidis (2015) and develop a unit root test for T fixed.

has time-specific $\boldsymbol{\beta}_t$ $(t = 1, \dots, T)$, and test $H_0 : \{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T\}$ *are all equal* vs $H_1^* : \{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T\}$ *are not all equal*. Such a test would be suitable as a general diagnostic test of model adequacy (see Andrews 1993, pg. 825 for such a global test when $T \to \infty$).

The methodology proposed in the previous sections under a known breakpoint generalizes in a straightforward way to the case of testing against H_1^*. To this end, let us define $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i' \left(\boldsymbol{y}_i - \sum_{t=1}^{T} \boldsymbol{X}_i^{(t)} \boldsymbol{\beta}_t \right) - \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{G})\boldsymbol{f}$ and $\hat{\boldsymbol{\mu}}_i(\boldsymbol{\theta}) = N^{-1} \sum_{i=1}^{N} \boldsymbol{\mu}_i(\boldsymbol{\theta})$ and consider (say) a 'global' LM-type test, which we denote by LM_G, of H_0 against H_1^*. Note that the H_1 in the previous sections is a special case of H_1^*. Therefore, one would conjecture that a test against H_1 is also likely to have power when the true model is not in H_1 but in H_1^*.

We extended the simulation study reported earlier to include the foregoing test against H_1^* as well. In particular, we considered the model (15) modified to $y_{it} = \beta_t y_{i,t-1} + \lambda_i f_t^0 + \varepsilon_{it}$, $(t = 1, \dots, T)$, where $\beta_t = 0.5 + (t/2)\omega^0$, for $t = 2, \dots, 6$. [16] As expected, when the true model has only one structural break, as is the case in model (1), LM_G-test against H_1^* exhibits lower power than the D_τ- and LM_τ-tests introduced in this paper. For example, for $N = 300$, $\omega^0 = 0.15$, $\tau_0 = 4$, the powers of D_τ, LM_τ and LM_G are 56%, 55% and 35%, respectively. On the other hand, for the case where β_t depends on t, we observed that LM_G tends to have moderately higher power compared to D_τ and LM_τ. For example, for $N = 300$, $\omega^0 = 0.15$, $\tau_0 = 4$, the powers for D_τ, LM_τ and LM_G tests are 77%, 76% and 84%, respectively.

[17] [18]

The results in Theorems 1 and 3 can be extended to obtain the efficiency of the test under a sequence of local hypotheses. To illustrate this, let us consider the sequence of local hypotheses $H_{1N} : \boldsymbol{\beta}_\tau = \boldsymbol{\beta}_1 + N^{-1/2}\boldsymbol{\delta}$, for a given break point τ,

[16] We assumed that $\beta_1 = \beta_2$, to ensure that all parameters are identified.

[17] **Mervyn: please see discussion above if you are happy. Should we mentioned we also developed theory for this case? Should we mention about supplement?**

[18] Vasilis: I think, we do not need to mention about these proofs, because it should be clear that we have these proofs from the fact that we chose $\beta_1 = \beta_2$ to overcome identifiability. If the referee or editor needs them, they will ask.

where $\boldsymbol{\delta} \neq 0$. Then, under H_{1N}, the limiting distribution of the test statistic in Theorem 1, for the case of known break point, is a noncentral chi-square. A similar result holds for the LM-test in Theorem 3. Hence, the proposed tests have nontrivial power against local hypotheses converging at the rate $N^{-1/2}$.

6 Conclusion

This paper developed a structural break detection test for panel data models with multi-factor error structure. The stochastic framework considered in the paper is very general because it allows for (i) multiple sources of unobserved heterogeneity, which are represented by common factors, and (ii) endogenous regressors. This is important because often the covariates receive some form of 'feedback' from the dependent variable. The proposed structural break tests are based on a distance type statistic and an LM-type statistic, both derived within the GMM framework. The asymptotic properties of the statistic are established for both known and unknown breakpoints, and when the number of factors is known and when it is unknown. The simulation study demonstrates that the proposed method performs well in terms of size and power, as well as in terms of locating the breakpoint correctly.

References

Ahn, S. C., Lee, Y. H. and Schmidt, P. (2013), 'Panel data models with multiple time-varying individual effects', *Journal of Econometrics* **174**(1), 1–14.

Andrews, D. W. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica: Journal of the Econometric Society* **61**(4), 821–856.

Andrews, D. W. (2003), 'End-of-sample instability tests', *Econometrica* **71**, 1661–1694.

Arellano, M. (2003), *Panel data econometrics*, Oxford University Press.

Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009), 'Break detection in the covariance structure of multivariate time series models', *The Annals of Statistics* **37**(6B), 4046–4087.

Bai, J. (2013), 'Fixed-effects dynamic panel models, a factor analytical method', *Econometrica* **81**(1), 285–314.

Bai, J. and Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**(1), 47–78.

Baltagi, B. H., Feng, Q. and Kao, C. (2016), 'Estimation of heterogeneous panels with structural breaks', *Journal of Econometrics* **191**(1), 176–195.

Bond, S. R. (2002), 'Dynamic panel data models: a guide to micro data methods and practice', *Portuguese Economic Journal* **1**, 141–162.

Bun, M. J. and Sarafidis, V. (2015), Dynamic panel data models, *in* B. H. Baltagi, ed., 'The Oxford Handbook of Panel Data', Oxford University Press.

Chan, F., Manchini-Grifolli, T. and Pauwels, L. (2008), Testing structural stability in heterogeneous panel data, Working paper, Curtin University of Technology, Paris-Jourdan Sciences Economiques and Hong Kong Monetary Authority and Graduate Institute of International Studies.
URL: *http://www.mssanz.org.au/MODSIM07/papers/16_s9/TestingStructura_s9_Chan_.pdf*

Chen, J. and Gupta, A. (1997), 'Testing and locating variance changepoints with application to stock prices', *Journal of the American Statistical Association* **92**(438), 739–747.

Demirguc-Kunt, A., Detragiache, E. and Merrouche, O. (2013), 'Bank capital: lessons from the financial crisis', *Journal of Money, Credit and Banking* **45**(6), 1147–1164.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica: Journal of the Econometric Society* **50**(4), 1029–1054.

Harchaoui, Z. and Lévy-Leduc, C. (2010), 'Multiple change-point estimation with a total variation penalty', *Journal of the American Statistical Association* **105**(492), 1480–1493.

Hsiao, C. (2014), *Analysis of Panel Data*, Cambridge University Press.

Kuersteiner, G. and Prucha, I. R. (2013), 'Limit theory for panel data models with cross sectional dependence and sequential exogeneity', *Journal of Econometrics* **174**(2), 107–126.

Li, D., Qian, J. and Su, L. (2016), 'Panel data models with interactive fixed effects and multiple structural breaks', *Journal of the American Statistical Association* **111**(516), 1804–1819.

Newey, W. K. and West, K. D. (1987), 'Hypothesis testing with efficient method of moments estimation', *International Economic Review* **28**(3), 777–787.

Perron, P. and Qu, Z. (2006), 'Estimating restricted structural change models', *Journal of Econometrics* **134**(2), 373–399.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Qian, J. and Su, L. (2016), 'Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso', *Journal of Econometrics* **191**(1), 86–109.

Qu, Z. and Perron, P. (2007), 'Estimating and testing structural changes in multivariate regressions', *Econometrica* **75**(2), 459–502.

Robertson, D. and Sarafidis, V. (2015), 'IV estimation of panels with factor residuals', *Journal of Econometrics* **185**, 526–541.

Robertson, D., Sarafidis, V. and Westerlund, J. (2018), 'Unit root inference in generally trending and cross-correlated fixed-t panels', *Journal of Business and Economic Statistics* **36**(31), 493–504.

Santarelli, E., Klomp, L. and Thurik, A. R. (2006), Gibrat′s law: an overview of the empirical literature, *in* 'Entrepreneurship, growth, and innovation', Springer US, pp. 41–73.

Silvapulle, M. J. and Sen, P. K. (2011), *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*, J Wiley, New Jerzy.

Simon, H. A. and Bonini, C. P. (1958), 'The size distribution of business firms', *The American Economic Review* **48**(4), 607–617.

Appendix

For an $m \times n$ real matrix $\boldsymbol{A}$, let $\boldsymbol{A}'$ denote its transpose, $\|\boldsymbol{A}\|$ denote its Frobenius norm $[tr(\boldsymbol{A}'\boldsymbol{A})]^{1/2}$, $vec(\boldsymbol{A})$ denote the vectorization of $\boldsymbol{A}$, $\boldsymbol{P}_A = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, and $\boldsymbol{M}_A = \boldsymbol{I}_m - \boldsymbol{P}_A$ where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix. Finally, let $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and in distribution respectively.

We assume that Assumptions 1–3 are satisfied throughout this Appendix. To simplify the notation and details, the proof Theorems 1–2 are provided here for the special case $K = 1$ and $r = 1$. This does not sacrifice the intricacies of the ideas, but simplifies the details. For this special case, the regressor x_{it}, the factor f_t^0, and the factor loading λ_i are scalars. Thus, our model under the alternative hypothesis H_1, reduces to the following form for some unknown τ in $\{2, 3, \ldots, T\}$:

$$y_{it} = \begin{cases} x_{it}\beta_1^0 + \lambda_i f_t^0 + \varepsilon_{it}; & t = 1, \ldots, \tau - 1; \\ x_{it}\beta_\tau^0 + \lambda_i f_t^0 + \varepsilon_{it}; & t = \tau, \ldots, T. \end{cases} \tag{19}$$

Lemma A.1. *Suppose that H_0 holds and $\tau \geq 2$. Let $\boldsymbol{A}_{\zeta\times(s+K)} = \boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1$, $\boldsymbol{B}_{\zeta\times(s+2K)} = \boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau$, $\mathcal{L}_1 = \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^{s+K}\}$, and $\mathcal{L}_2 = \{\boldsymbol{B}\boldsymbol{y} : \boldsymbol{y} \in \mathbb{R}^{s+2K}\}$. Then*
(a) $\boldsymbol{P_A} := \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ is the projection matrix onto $\mathcal{L}_1$, $\boldsymbol{P_B} := \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'$ is the projection matrix onto $\mathcal{L}_2$, $\mathcal{L}_1$ is a linear subspace of $\mathcal{L}_2$, and $\boldsymbol{P_B}\boldsymbol{P_A} = \boldsymbol{P_A}\boldsymbol{P_B} = \boldsymbol{P_A}$.
(b) $\boldsymbol{V}_\tau = \boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1} - \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau} = \boldsymbol{P_B} - \boldsymbol{P_A}$ is the projection matrix onto $\mathcal{L}_1^\perp \cap \mathcal{L}_2$ where $\mathcal{L}_1^\perp$ is the orthogonal complement of $\mathcal{L}_1$ in $\mathcal{L}_2$, and the rank of $\boldsymbol{V}_\tau$ is $\dim(\beta_\tau)$.

Proof. It follows from the theory of least squares estimation, that $\boldsymbol{P_A} := \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ is the projection matrix onto $\mathcal{L}_1$, and $\boldsymbol{P_B} := \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'$ is the projection matrix onto $\mathcal{L}_2$. Let s denote the dimension of $(\boldsymbol{g}', \boldsymbol{f}')$. Recall that $\boldsymbol{\theta}_1' = (\boldsymbol{g}', \boldsymbol{f}', \boldsymbol{\beta}_1')_{1\times(s+K)}$ and $\boldsymbol{\theta}_\tau' = (\boldsymbol{g}', \boldsymbol{f}', \boldsymbol{\beta}_1', \boldsymbol{\beta}_\tau')_{1\times(s+2K)}$. Recall the definition of $\{\boldsymbol{M}, \boldsymbol{M}^{(\tau)}, \boldsymbol{M}^{(1)}\}$ in (4). Differentiating $\boldsymbol{\mu}_{1,i}(\boldsymbol{\theta}_1)$ with respect to $\boldsymbol{\theta}_1$ and taking expectation, we may write $\boldsymbol{\Gamma}_1 = -[\boldsymbol{H}_{\zeta\times s} \mid \boldsymbol{M}_{\zeta\times K}]$, for some $\boldsymbol{H}(\boldsymbol{\theta}_1)$. Similarly $\boldsymbol{\Gamma}_\tau = \boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau) = -[\boldsymbol{H}_{\zeta\times s} \mid \boldsymbol{M}^{(1)}_{\zeta\times K} \mid \boldsymbol{M}^{(\tau)}_{\zeta\times K}]$. Since $\boldsymbol{M} = \boldsymbol{M}^{(\tau)} + \boldsymbol{M}^{(1)}$, we have $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_\tau\boldsymbol{R}$, where $\boldsymbol{R}$ is the 3×2 partitioned matrix $[I_s,\ \boldsymbol{0}; \boldsymbol{0},\ I_K; \boldsymbol{0},\ I_K]$. Since $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{R}$, it follows that $\mathcal{L}_1$ is a linear subspace of the linear space $\mathcal{L}_2$. Therefore, projecting a vector $\boldsymbol{z}$ onto $\mathcal{L}_2$ and then onto $\mathcal{L}_1$ provides the same result as projecting $\boldsymbol{z}$ onto $\mathcal{L}_1$. Hence, we have $\boldsymbol{P_A}\boldsymbol{P_B}\boldsymbol{z} = \boldsymbol{P_B}\boldsymbol{P_A}\boldsymbol{z} = \boldsymbol{P_A}\boldsymbol{z}$ for any $\boldsymbol{z}$; this may also be verified directly as $\boldsymbol{P_A}\boldsymbol{P_B} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}' = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{R}'\boldsymbol{B}'\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}' = \boldsymbol{P_A}$. Similarly, $\boldsymbol{P_B}\boldsymbol{P_A} = \boldsymbol{P_A}$. Now, it is easily seen that $\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1} - \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau} = \boldsymbol{P_B} - \boldsymbol{P_A}$, is a projection matrix; it projects onto $\mathcal{L}_1^\perp \cap \mathcal{L}_2$, where $\mathcal{L}_1^\perp$ is the orthogonal complement of $\mathcal{L}_1$ in $\mathcal{L}_2$. The rank of $\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1} - \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau}$ is $trace[\boldsymbol{P_B}] - trace[\boldsymbol{P_A}] = K$, the number of equality constraints in the null hypothesis. □

Proof of Theorem 1. Let $\boldsymbol{C}_\tau = \boldsymbol{\Gamma}_\tau'(\boldsymbol{\theta}_\tau^0)\boldsymbol{\Phi}_\tau^{-1}(\boldsymbol{\theta}_\tau^0)\boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0)$ and $\boldsymbol{D}_\tau = \boldsymbol{\Gamma}_\tau'(\boldsymbol{\theta}_\tau^0)\boldsymbol{\Phi}_\tau^{-1}(\boldsymbol{\theta}_\tau^0)$. Since the null hypothesis is assumed to hold, we also have, $\beta_\tau = \beta_1$, $\boldsymbol{\mu}_1(\boldsymbol{\theta}_1^0) = \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau^0)$, $\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0) = \hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0)$, and $\boldsymbol{\Phi}_1(\boldsymbol{\theta}_1^0) = \boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0)$.

By part (ii) of Lemma 1 and the Mean Value Theorem applied to $\sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau)$, we have, for $\tau \geq 2$

$$\begin{aligned}\sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau) &= \sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) + \hat{\boldsymbol{\Gamma}}_\tau(\bar{\boldsymbol{\theta}}_\tau)\sqrt{N}(\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}_\tau^0)\\ &= \sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) - \boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0)\boldsymbol{C}_\tau^{-1}\boldsymbol{D}_\tau\sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) + o_p(1)\\ &= \left[\boldsymbol{I}_\zeta - \boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0)\boldsymbol{C}_\tau^{-1}\boldsymbol{D}_\tau\right]\sqrt{N}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) + o_p(1).\end{aligned}$$

Hence, with $\boldsymbol{\Gamma}_\tau = \boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0)$ and $\boldsymbol{\Phi}_\tau = \boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0)$ for simplicity,

$$\begin{aligned}\sqrt{N}\boldsymbol{\Phi}_\tau^{-1/2}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau) &= [\boldsymbol{I}_\zeta - \boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau(\boldsymbol{\Gamma}_\tau'\boldsymbol{\Phi}_\tau^{-1}\boldsymbol{\Gamma}_\tau)^{-1}\boldsymbol{\Gamma}_\tau'\boldsymbol{\Phi}_\tau^{-1/2}]\sqrt{N}\boldsymbol{\Phi}_\tau^{-1/2}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) + o_p(1)\\ &= \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau}\left(\sqrt{N}\boldsymbol{\Phi}_\tau^{-1/2}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0)\right) + o_p(1)\\ &= \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau}\left(\sqrt{N}\boldsymbol{\Phi}_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0)\right) + o_p(1)\\ &= \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau}\boldsymbol{z}_N + o_p(1). \qquad (20)\end{aligned}$$

Then

$$N\{\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau)\}'\boldsymbol{\Phi}_\tau^{-1}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau) = \boldsymbol{z}_N'\boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau}\boldsymbol{z}_N + o_p(1). \qquad (21)$$

By similar arguments, we have the following for $\tau = 1$:

$$\begin{aligned}\sqrt{N}\boldsymbol{\Phi}_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) &= [\boldsymbol{I}_\zeta - \boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1(\boldsymbol{\Gamma}_1'\boldsymbol{\Phi}_1^{-1}\boldsymbol{\Gamma}_1)^{-1}\boldsymbol{\Gamma}_1'\boldsymbol{\Phi}_1^{-1/2}]\sqrt{N}\boldsymbol{\Phi}_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0) + o_p(1)\\ &= \boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\left(\sqrt{N}\boldsymbol{\Phi}_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0)\right) + o_p(1)\\ &= \boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{z}_N + o_p(1), \qquad (22)\end{aligned}$$

where $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_1(\boldsymbol{\theta}_1^0)$, $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_1(\boldsymbol{\theta}_1^0)$. Then

$$N\{\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1)\}'\boldsymbol{\Phi}_1^{-1}\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) = \boldsymbol{z}_N'\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{z}_N + o_p(1). \qquad (23)$$

Since $\hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)}) \xrightarrow{p} \boldsymbol{\Phi}_\tau^{-1}$ $(\tau \geq 1)$, the statistic for testing H_0 against $H_{(\tau)}$ is

$$\begin{aligned}D_\tau = N\left[\hat{Q}_1(\hat{\boldsymbol{\theta}}_1) - \hat{Q}_\tau(\hat{\boldsymbol{\theta}}_\tau)\right] &= N\left[\hat{\boldsymbol{\mu}}_1'(\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\Phi}}_1^{-1}(\hat{\boldsymbol{\theta}}_1^{(1)})\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) - \hat{\boldsymbol{\mu}}_\tau'(\hat{\boldsymbol{\theta}}_\tau)\hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)})\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau)\right]\\ &= N\left[\hat{\boldsymbol{\mu}}_1'(\hat{\boldsymbol{\theta}}_1)\boldsymbol{\Phi}_1^{-1}\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) - \hat{\boldsymbol{\mu}}_\tau'(\hat{\boldsymbol{\theta}}_\tau)\boldsymbol{\Phi}_\tau^{-1}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau)\right] + o_p(1),\\ &= \boldsymbol{z}_N'[\boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau} - \boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}]\boldsymbol{z}_N + o_p(1), \qquad \text{by (21) and (23)}\\ &= \boldsymbol{z}_N'\boldsymbol{V}_\tau\boldsymbol{z}_N + o_p(1). \qquad (24)\end{aligned}$$

Since $\boldsymbol{z}_N$ converges in distribution to $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_\zeta)$, we have $D_\tau \xrightarrow{d} \boldsymbol{z}'\boldsymbol{V}_\tau\boldsymbol{z}$. Since $\boldsymbol{V}_\tau$ is a projection matrix by Lemma A.1, it follows that $D_\tau \xrightarrow{d} \chi^2_K$ where K is the rank of $\boldsymbol{V}_\tau$, which in turn is $dim(\boldsymbol{\beta}_\tau)$. □

Proof of Theorem 2. Let $\tau_\ell \in \{\tau_1, \ldots, \tau_L\}$. By Theorem 1, we have

$$D_{\tau_\ell} = \sqrt{N}\hat{\boldsymbol{\mu}}_1'(\boldsymbol{\theta}_1^0)\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{V}_{\tau_\ell}\sqrt{N}\boldsymbol{\Phi}_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0) + o_p(1).$$

Let $\boldsymbol{D} = (D_{\tau_1}, \ldots, D_{\tau_L})'$.

Then, since $\boldsymbol{z}_N = \sqrt{N}[\boldsymbol{\Phi}_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0)]$, we have

$$\boldsymbol{D} = (\boldsymbol{z}_N'\boldsymbol{V}_{\tau_1}\boldsymbol{z}_N, \ldots, \boldsymbol{z}_N'\boldsymbol{V}_{\tau_L}\boldsymbol{z}_N) + o_p(1) \xrightarrow{d} (\boldsymbol{z}'\boldsymbol{V}_{\tau_1}\boldsymbol{z}, \ldots, \boldsymbol{z}'\boldsymbol{V}_{\tau_L}\boldsymbol{z}). \tag{25}$$

The Distance type test statistic, $\max\limits_{l=1,\ldots,L}\{D_{\tau_\ell}\}$, is a continuous function of $[D_{\tau_1}, \ldots, D_{\tau_L}]'$. Therefore, by the continuous mapping theorem, the asymptotic distribution of $\max\limits_{l=1,\ldots,L}\{D_{\tau_\ell}\}$ is the distribution of the maximum of the components in $\{\boldsymbol{z}'\boldsymbol{V}_{\tau_1}\boldsymbol{z}, \ldots, \boldsymbol{z}'\boldsymbol{V}_{\tau_L}\boldsymbol{z}\}$. □

Proof of Theorem 3.

It follows form (3), (4), and (7) that $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}) = \boldsymbol{\mu}_1(\boldsymbol{\theta})$ and $\boldsymbol{\Phi}_\tau(\boldsymbol{\theta}) = \boldsymbol{\Phi}_1(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ satisfies H_0. Also, since $\hat{\boldsymbol{\theta}}_R$ is in H_0, we have $\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_R) = \hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1)$ and $\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_R) = \hat{\boldsymbol{\Phi}}_1(\hat{\boldsymbol{\theta}}_1)$. Let $\hat{\boldsymbol{E}}_\tau = [\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_R)]^{-1/2}[\hat{\boldsymbol{\Gamma}}_\tau(\hat{\boldsymbol{\theta}}_R)]$ and $\boldsymbol{E}_\tau = [\boldsymbol{\Phi}_\tau(\boldsymbol{\theta}^0)]^{-1/2}[\boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}^0)]$. Then $\hat{\boldsymbol{E}}_\tau = \boldsymbol{E}_\tau + o_p(1), \hat{\boldsymbol{U}}_\tau = \hat{\boldsymbol{E}}_\tau'\hat{\boldsymbol{E}}_\tau$ and $\hat{\boldsymbol{U}}_\tau^{-1} = [\boldsymbol{E}_\tau'\boldsymbol{E}_\tau]^{-1} + o_p(1)$. By (22), we have (see (13) for definitions of $\hat{\boldsymbol{A}}_\tau$ and $\hat{\boldsymbol{U}}_\tau$)

$$\begin{aligned}
\sqrt{N}\hat{\boldsymbol{A}}_\tau &= \sqrt{N}[\hat{\boldsymbol{\Gamma}}_\tau(\hat{\boldsymbol{\theta}}_R)]'[\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_R)]^{-1}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_R) = \sqrt{N}[\hat{\boldsymbol{\Gamma}}_\tau(\hat{\boldsymbol{\theta}}_R)]'[\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_R)]^{-1/2}[\hat{\boldsymbol{\Phi}}_1(\hat{\boldsymbol{\theta}}_1)]^{-1/2}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_R) \\
&= \hat{\boldsymbol{E}}_\tau'[\boldsymbol{\Phi}_1(\boldsymbol{\theta}^0)]^{-1/2}\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) + o_p(1) = \boldsymbol{E}_\tau'\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{z}_N + o_p(1).
\end{aligned}$$

Using the foregoing representation of $\sqrt{N}\hat{\boldsymbol{A}}_\tau$, we obtain

$$\begin{aligned}
LM_\tau &= N\hat{\boldsymbol{A}}_\tau'\hat{\boldsymbol{U}}_\tau^{-1}\hat{\boldsymbol{A}}_\tau \\
&= [\boldsymbol{E}_\tau'\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{z}_N + o_p(1)]'\{[\boldsymbol{E}_\tau'\boldsymbol{E}_\tau]^{-1} + o_p(1)\}\{\boldsymbol{E}_\tau'\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{z}_N + o_p(1)\} \\
&= \boldsymbol{z}_N'\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{E}_\tau[\boldsymbol{E}_\tau'\boldsymbol{E}_\tau]^{-1}\boldsymbol{E}_\tau'\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{z}_N + o_p(1) \\
&= \boldsymbol{z}_N'\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}(\boldsymbol{I} - \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau})\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{z}_N + o_p(1) \\
&= \boldsymbol{z}_N'(\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1} - \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau})\boldsymbol{z}_N + o_p(1) = \boldsymbol{z}_N'\boldsymbol{V}_\tau\boldsymbol{z}_N + o_p(1);
\end{aligned}$$

the second last step follows since, $\boldsymbol{M}^2_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1} = \boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1} = \boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}\boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau} = \boldsymbol{M}_{\boldsymbol{\Phi}_\tau^{-1/2}\boldsymbol{\Gamma}_\tau}\boldsymbol{M}_{\boldsymbol{\Phi}_1^{-1/2}\boldsymbol{\Gamma}_1}$, by Lemma 1. □

The proof of Theorem 4 is similar to that of Theorem 2, hence omitted.

Table 6: **Percent of times that the null hypothesis was rejected when the breakpoint τ_0 and the number of factors r_0 are known.**

		$\tau_0 = 4^{(a)}$						$\tau_0 = 6$					
		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$	
T	N	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ
6	100	6.3	6.7	10.8	11.4	19.2	19.5	6.2	6.9	6.7	6.0	7.5	8.0
	300	5.0	4.8	29.7	29.0	55.8	55.1	5.9	5.7	9.4	8.9	14.9	14.2
	600	5.0	4.9	53.0	52.7	81.8	80.4	5.2	5.3	14.8	13.7	26.3	24.8
9	100	7.2	16.4	12.8	12.7	23.1	23.7	6.7	15.9	15.6	14.2	29.8	28.4
	300	5.3	5.2	43.0	43.7	75.0	75.5	5.5	5.8	47.9	47.2	80.0	77.9
	600	6.1	6.3	70.8	71.2	94.5	94.5	5.9	6.0	79.0	78.2	96.0	95.0

[a] The parameter β_1^0 is fixed at 0.5. By definition, $\omega^0 = \beta_{\tau_0} - \beta_1$; therefore, the estimates under $\omega^0 = 0.00$ are the type I error rates in %; each entry under $\omega^0 = 0.10$ and $\omega^0 = 0.15$ is the estimated power.

The next page provides the same table with two significant digits; I prefer table 7, but OK with Table 6.

Table 7: **Percent of times that the null hypothesis was rejected when the breakpoint τ_0 and the number of factors r_0 are known.**

		$\tau_0 = 4^{(a)}$						$\tau_0 = 6$					
		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$		$\omega^0 = 0.00$		$\omega^0 = 0.10$		$\omega^0 = 0.15$	
T	N	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ	D_τ	LM_τ
6	100	6.3	6.7	11	11	19	20	6.2	6.9	6.7	6.0	7.5	8.0
	300	5.0	4.8	30	29	56	55	5.9	5.7	9.4	8.9	15	14
	600	5.0	4.9	53	53	82	80	5.2	5.3	15	14	26	25
9	100	7.2	16	13	13	23	24	6.7	16	16	14	30	28
	300	5.3	5.2	43	44	75	76	5.5	5.8	48	47	80	78
	600	6.1	6.3	71	71	95	95	5.9	6.0	79	78	96	95

(a) The parameter β_1^0 is fixed at 0.5. By definition, $\omega^0 = \beta_{\tau_0} - \beta_1$; therefore, the estimates under $\omega^0 = 0.00$ are the type I error rates in %; each entry under $\omega^0 = 0.10$ and $\omega^0 = 0.15$ is the estimated power.

Table 8: **Performance of the tests when the break point τ_0 is unknown and the number of factors r_0 is known; $T = 6$ and $\tau_0 = 4$.**

		D_{max}					LM_{max}				
			Distribution of $\hat{\tau}^{(b)}$					Distribution of $\hat{\tau}^{(b)}$			
ω^0	N	$P^{(a)}$	$\hat{\tau} = 3$	$\hat{\tau} = 4^{(c)}$	$\hat{\tau} = 5$	$\hat{\tau} = 6$	$P^{(a)}$	$\hat{\tau} = 3$	$\hat{\tau} = 4^{(c)}$	$\hat{\tau} = 5$	$\hat{\tau} = 6$
0.00	100	7.1	-	-	-	-	6.7	-	-	-	-
	300	5.7	-	-	-	-	5.2	-	-	-	-
	600	5.0	-	-	-	-	5.0	-	-	-	-
0.10	100	7.9	15	33	16	36	8.3	21	35	21	23
	300	21.5	12	59	14	16	22.8	13	54	15	17
	600	42.5	11	67	12	11	42.3	9	66	13	12
0.15	100	14.8	14	42	14	30	15.1	17	46	16	21
	300	44.5	9	68	12	11	44.5	9	68	12	11
	600	75.2	7	78	9	7	73.8	6	78	8	8

(a) The variable P denotes the percent of times that the null hypothesis was rejected by the relevant test. For example, when $\omega^0 = 0$ and $N = 100$, the null hypothesis was rejected by the D_{max}-test 7.1% times in the 5000 iid samples. The values of P in the three rows corresponding to $\omega^0 = 0.00$ are Type I error rates; the other values of P are estimated powers.

(b) The 'Distribution of $\hat{\tau}$' provides the observed distribution (in %) of $\hat{\tau}$ among the samples for which the null hypothesis was rejected by the test. As an example, for $\omega^0 = 0.10$ and $N = 100$, the empirical power of the D_{max}-test is 7.9%. Further, among the 7.9% of the 5000 samples that rejected the null hypothesis of no structural break, 15%, 33%, 16%, and 36% of the samples estimated the breakpoint to be $\hat{\tau} = 3, 4, 5, 6$ respectively; therefore, the sum of these four percentages is 100%.

(c) Since the true breakpoint is $\tau_0 = 4$, the column under $\hat{\tau} = 4$ provides the percent of times the break point was corrected estimated, conditional on the null hypothesis being rejected.

The results obtained in Section 2 can also be extended in other directions. An anonymous referee suggested that it would be of interest to compare and contrast the present setting with the more flexible model $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta}_t + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{it}$, which has time-specific $\boldsymbol{\beta}_t$ ($t = 1, \ldots, T$), and test $H_0 : \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T\}$ *are all equal* vs $H_1^* : \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T\}$ *are not all equal*. Such a test would be suitable as a general diagnostic test of model adequacy (see Andrews 1993, pg. 825 for such a global test when $T \to \infty$).

The methodology proposed in the previous sections under a known breakpoint generalizes to the case of testing against H_1^*. Let LM_G and D_G denote the LM and distance statistics for testing H_0 against H_1^*. Then the asymptotic null distribution of these two test statistics is χ_k^2 where k is the number of parameter equality constraints. Since H_1 in the previous sections is a special case of H_1^*, one would conjecture that a test against H_1 is also likely to have power when the true model is not in H_1 but in H_1^*. We would expect that if the true model has only break, then the tests introduced in this paper are likely to be more powerful than the aforementioned more global tests, D_G and LM_G. Similarly, if the true model has several breaks, then the tests introduced in the previous sections of this paper are likely to be less powerful than D_G and LM_G. Our simulation results, reported below, corroborate these conjectures.

We extended the simulation study in Table 1 with the corresponding model (15) modified to $y_{it} = \beta_t y_{i,t-1} + \lambda_i f_t^0 + \varepsilon_{it}$, ($t = 1, \ldots, T$), where $\beta_t = 0.5 + (t/2)\omega^0$, for $t = 2, \ldots, 6$; we assumed that $\beta_1 = \beta_2$, to ensure that the parameters are identified. As expected, when the true DGP is as in Table 1 with exactly one break and $(N, \omega^0, \tau_0) = (300, 0.15, 4)$, the powers of D_{τ_0}, LM_{τ_0} and LM_G are 56%, 55% and 35%, respectively. On the other hand, when the true DGP is modified to have the aforementioned time-specific β_t, the powers the powers of D_{τ_0}, LM_{τ_0} and LM_G are 77%, 76% and 84%, respectively.