



**MONASH** University

**Australia**

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Testing for a Structural Break in Dynamic Panel  
Data Models with Common Factors**

**Huanjun Zhu, Vasilis Sarafidis,  
Mervyn Silvapulle, Jiti Gao**

**September 2015**

**Working Paper 20/15**

# Testing for a Structural Break in Dynamic Panel Data Models with Common Factors

HUANJUN ZHU, VASILIS SARAFIDIS, MERVYN SILVAPULLE AND JITI GAO\*

Department of Econometrics and Business Statistics, Monash University, Australia

## Abstract

This paper develops a method for testing for the presence of a single structural break in panel data models with unobserved heterogeneity represented by a factor error structure. The common factor approach is an appealing way to capture the effect of unobserved variables, such as skills and innate ability in studies of returns to education, common shocks and cross-sectional dependence in models of economic growth, law enforcement acts and public attitudes towards crime in statistical modelling of criminal behaviour. Ignoring these variables may result in inconsistent parameter estimates and invalid inferences. We focus on the case where the time frequency of the data may be yearly and thereby the number of time series observations is small, even if the sample covers a rather long period of time. We develop a Distance type statistic based on a Method of Moments estimator that allows for unobserved common factors. Existing structural break tests proposed in the literature are not valid under these circumstances. The asymptotic properties of the test statistic are established for both known and unknown breakpoints. In our simulation study, the method performed well, both in terms of size and power, as well as in terms of successfully locating the time at which the break occurred. The method is illustrated using data from a large sample of banking institutions, providing empirical evidence on the well-known Gibrat's 'Law'.

**Key words:** Method of Moments, unobserved heterogeneity, break-point detection, fixed T asymptotics.

*JEL:* C12, C23, C26.

---

\**Corresponding Author:* Email: jiti.gao@monash.edu. The fourth author acknowledges financial support by the Australian Research Council Discovery Grants Program under Grant Numbers: DP1314229 and DP150101012.

# 1 Introduction

There is a vast literature in time series analysis on testing for structural breaks in various moments of the distribution of random variables, such as the unconditional mean (e.g. Harchaoui and Lévy-Leduc 2010); the unconditional variance (e.g. Chen and Gupta 1997); and the covariance structure of multivariate models (e.g. Aue et al. 2009). Structural break detection methods on conditional models have also been prevalent in econometrics, such as those developed by Andrews (1993), Bai and Perron (1998), Perron and Qu (2006), and Qu and Perron (2007), among others.

With the increasing availability of longitudinal data sets, methods for testing for structural breaks in panel data models are in demand. Panel data regression analysis is a popular tool in many areas of economics, finance and statistics. The main motivation behind using such analysis is that it provides a framework for dealing with unobserved heterogeneity (i.e. what is left over after conditioning on a set of regressors) in a more effective way compared to univariate time series, or cross-sectional models.

In particular, statistical inferences may be erroneous if, in addition to the observed variables under study, there exist other relevant variables that are unobserved but correlated with the observed variables. An increasingly popular method for dealing with such problems is the common factor approach; see Sarafidis and Wansbeek (2012) for a recent overview of common factor models in panel data analysis.

To illustrate, consider the study area of analysing returns to education, or earnings and inequality; the individual wage rate is often modeled as a function of observed characteristics, such as education, experience, tenure, gender and race; however, wages also depend on individual-specific characteristics that are unobserved and typically difficult to measure, like innate ability, skills, and so forth. These characteristics are likely to be correlated with the regressors - for example, all other things being equal, it would be the most able individuals who embark on higher education. This implies a positive correlation between education and ability. Furthermore, the returns to unobserved skills are likely to vary over time, e.g. with the business cycle of the economy. Failure to take into account such complex features of the model may result in inconsistent parameter estimates and invalid inferences. The common factor approach is appealing because it presents a natural framework for capturing the effect of these variables. In particular, letting the regression error term be given by  $u_{it} = \boldsymbol{\lambda}'_i \mathbf{f}_t + \varepsilon_{it}$ , where both  $\boldsymbol{\lambda}_i$  and  $\mathbf{f}_t$  are  $r \times 1$  vectors and  $\varepsilon_{it}$  is a purely idiosyncratic error component, individual-specific unmeasured skills may be

represented by the factor loadings,  $\lambda_i$ , while the prices of these skills that affect wages are captured by the factors. In this way, skill prices vary over time in an intertemporally arbitrary way. A classic application is given by Cawley et al. (1997).

There are several other areas where the application of the factor structure has been useful. For instance, in empirical growth models the factors may represent different sources of time-varying technology that is potentially available to all countries, while the factor loadings may reflect the rate at which countries absorb such technological advances; see, e.g. Bun and Sarafidis (2015). In the context of estimating private returns to R&D, the factor component often reflects knowledge spill overs and cross-sectional dependence (e.g. Eberhardt et al. 2013). Similarly, when considering yearly stock price movements in financial analysis, the presence of ‘common shocks’, such as economy-wide recessions, technical innovation and exchange rates movements can be represented by common factors (e.g. Geweke and Zhou 1996).

In the aforementioned applications the number of time series observations,  $T$ , is relatively small. This is because typically the frequency of the data is yearly, and therefore (say)  $T = 10$  spans a rather long period of time. Nevertheless, the above models often rely on the assumption of parameter stability over time. In practice, however, economic agents may inhabit an ever changing economic environment; the impact of globalization and of the recent global financial crisis, as well as events of major technological advances are just few of the factors that may contribute to possible structural changes in the economic mechanisms that generate the data we observe. Ignoring breaks of such nature is likely to result in inconsistent parameter estimates and invalid inferences.

Clearly, from a modeling point of view it is important to be able to employ statistical methods for testing against structural breaks in the aforementioned empirical studies. Currently the only way of doing this is based on the seminal approach put forward by De Wachter and Tzavalis (2012). However, this method imposes that unobserved heterogeneity is *additive* and therefore common factors in the error term are ruled out.

The present paper advances the current state of econometric literature and fills in an important gap. In particular, we build upon the method proposed by De Wachter and Tzavalis (2012) and develop a Distance type test statistic that is valid under a common factor structure. Our approach essentially involves modifying appropriately the Method of Moments estimator developed by Robertson and Sarafidis (2015) in order to allow for changes in the structural parameters. In contrast to De Wachter and Tzavalis (2012), our method does not eliminate unobserved heterogeneity using some form of differencing of the

data, but instead it introduces extra parameters that capture the covariances between the instruments and the unobserved common factor component. Our estimator is more efficient than existing estimators that rely on differencing of the data and possesses the traditional attraction of Method of Moments without imposing further distributional assumptions (see Robertson and Sarafidis 2015). Thus, our method is semiparametric.

It is worth mentioning that there do exist alternative test statistics against structural breaks in the context of panel data models, although the literature remains quite sparse. In particular, Chan et al. (2008) extend the test statistic developed by Andrews (2003) for time series data to heterogeneous panel data models; Baltagi et al. (2013) study estimation of static heterogeneous panels with a common break using the common correlated effects estimator of Pesaran (2006); and Qian and Su (2014) consider estimation and inference of possibly multiple common breaks in panel data models via adaptive group fused lasso, allowing for common factors and cross-sectional dependence. However, these methods are valid only for large  $T$ . In actual fact, there is a large number of applications where this condition is not satisfied, even if the sampling period spans over a relatively long time interval. For instance, the empirical section tests for Gibrat's 'Law' (or the Law of Proportionate Effect) using data from a large sample of 4,128 banking institutions, each one being observed over 13 years. Since the sampling time period overlaps with the GFC, which implies potentially major 'shocks' in the model, and there are omitted variables that are common across individual banks, it is crucial to be able to test for the presence of a structural break using a method that allows for common factors and cross-sectional dependence.

The rest of the paper is organized as follows. Section 2 describes the model and develops the test statistic, deriving its asymptotic properties. Section 3 and 4 examine the performance of the test statistic in finite samples using simulated data and real data respectively. Section 5 concludes. All proofs are contained in the Appendix.

## 2 A New Structural Break Test

We start with a description of the model and its assumptions, while the next section describes the moment conditions that we employ and provides an illustrative example.

## 2.1 Stochastic Framework

We study a linear dynamic 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 autoregressive and/or slope coefficients. Consider the following model:

$$y_{it} = \begin{cases} \rho^0 y_{i,t-1} + \mathbf{x}'_{it} \boldsymbol{\beta}^0 + \boldsymbol{\lambda}'_i \mathbf{f}_t^0 + \varepsilon_{it} & t < \tau; \\ \eta_\tau^0 y_{i,t-1} + \mathbf{x}'_{it} \boldsymbol{\delta}_\tau^0 + \boldsymbol{\lambda}'_i \mathbf{f}_t^0 + \varepsilon_{it} & t \geq \tau, \end{cases} \quad (1)$$

where both  $\boldsymbol{\lambda}_i$  and  $\mathbf{f}_t^0$  are  $r \times 1$  vectors, while  $(\eta_\tau^0, \boldsymbol{\delta}_\tau^0)$  replaces  $(\rho^0, \boldsymbol{\beta}^0)$  from the break at time  $\tau$ ,  $\tau \geq 2$ ;  $t = 1, \dots, T$ . Notice that since  $T$  is held fixed in our study,  $\mathbf{f}_t^0$  is treated as a parameter vector to be estimated together with the structural parameters of the model.

Our testing problem can be studied more formally by defining the following hypotheses:

- $H_0$  : There are no structural breaks
- $H_{(\tau)}$  : There is a structural break at time  $\tau$ , where  $\tau$  is known
- $H_1$  : There is a structural break at time  $\tau$ , where  $\tau$  is unknown.

If  $H_1$  is true then we shall denote the true point in time where the break occurs, by  $\tau_0$ .

The model above can be expressed in vector form as

$$\mathbf{y}_i = \rho^0 \mathbf{y}_{i,-1}^{(b)} + \eta_\tau^0 \mathbf{y}_{i,-1}^{(a)} + \mathbf{X}_i^{(b)} \boldsymbol{\beta}^0 + \mathbf{X}_i^{(a)} \boldsymbol{\delta}_\tau^0 + (\mathbf{I}_T \otimes \boldsymbol{\lambda}'_i) \mathbf{f}^0 + \boldsymbol{\varepsilon}_i, \quad (2)$$

where  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'_{T \times 1}$ ,  $\mathbf{y}_{i,-1}^{(b)} = (y_{i0}, y_{i1}, \dots, y_{i,\tau-2}, 0, \dots, 0)'_{T \times 1}$ ,  $\mathbf{y}_{i,-1}^{(a)} = (0, \dots, 0, y_{i,\tau-1}, \dots, y_{i,T-1})'_{T \times 1}$ ,  $\mathbf{X}_i^{(b)} = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i,\tau-1}, \mathbf{0}_{k \times 1}, \dots, \mathbf{0}_{k \times 1})'_{T \times k}$ ,  $\mathbf{X}_i^{(a)} = (\mathbf{0}_{k \times 1}, \dots, \mathbf{0}_{k \times 1}, \mathbf{x}_{i\tau}, \dots, \mathbf{x}_{iT})'_{T \times k}$ ,  $\mathbf{f}^0 = \text{vec}[(\mathbf{F}^0)']$ ,  $\mathbf{F}^0 = (\mathbf{f}_1^0, \mathbf{f}_2^0, \dots, \mathbf{f}_T^0)'_{T \times r}$  and  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'_{T \times 1}$ ; the value ‘ $-1$ ’ in the suffix indicates that the corresponding values have been lagged by one time period; the superscripts ‘ $(b)$ ’ and ‘ $(a)$ ’ indicate that the vectors correspond to the periods before  $\tau$  and from  $\tau$  onwards respectively, regardless of whether or not a break has occurred.

The following assumption is employed throughout the paper.

**Assumption 1:** (i)  $(\mathbf{x}_{it}, \boldsymbol{\lambda}_i, \varepsilon_{it}, y_{i0})$  are independently and identically distributed for  $i = 1, \dots, N$ , with each component having finite fourth moment. (ii) There exists  $a \in (0, \infty)$  such that  $|\rho^0| \leq a$ ,  $|\eta_\tau^0| \leq a$ ,  $\|\boldsymbol{\beta}^0\| \leq a$ , and  $\|\boldsymbol{\delta}_\tau^0\| \leq a$ . (iii)  $\mathbf{f}^0$  is non-stochastic and there exists  $b \in (0, \infty)$  such that  $\|\mathbf{f}^0\| \leq b$ . (iv)  $E(\varepsilon_{it} | y_{i0}, \dots, y_{i,t-1}, \boldsymbol{\lambda}_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{ih}) = 0$ ,

for  $t = 1, \dots, T$ ,  $i = 1, \dots, N$ , and some positive integer  $h$ . The value of  $h$  depends on whether regressor  $\mathbf{x}_{it}$  is strictly (weakly) exogenous, or endogenous.

Assumption 1 is fairly standard in this literature, for example, it is in line with Assumption 2 in Robertson and Sarafidis (2015), and Assumptions BA.1, BA.3, and BA.4 in Ahn et al. (2013). The independence assumption over  $i = 1, \dots, N$ , in the first part of Assumption 1, can be relaxed so long as the moment functions that will be defined below converge to their respective expectation in probability. The requirement of identical distribution in the same assumption, can also be relaxed. For example,  $\varepsilon_{it}$  could be heterogeneously distributed across both  $i$  and  $t$ , while conditional moments of  $\boldsymbol{\lambda}_i$  could also depend on  $i$  (see e.g. Juodis and Sarafidis 2015). We do not consider such generalizations in this paper in order to avoid unnecessary notational complexity.

Assumption 1 (iv) implies that the idiosyncratic errors are conditionally serially uncorrelated. This can be relaxed in a straightforward way and 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. In addition, this assumption implies that the idiosyncratic error is conditionally uncorrelated with the factor loadings. This is standard in the dynamic panel data literature, and allows for lagged values of  $y$  in levels to be used as instruments. Moreover, the value of  $h$  in Assumption 1 (iv) characterises the exogeneity properties of the covariates. For example, for  $h = T$  (respectively,  $h = t$ ) the covariates are strictly (respectively, weakly) exogenous, or otherwise they would be endogenous (see Arellano 2003, pg. Section 8.1). Our methodology is valid regardless of the value of  $h$  mutatis mutandis.

## 2.2 Moment Conditions

Our approach is based on a class of Method of Moments estimators known as *Factor Instrumental Variables* [FIV] estimators, for consistent inference in model (1); see Robertson and Sarafidis (2015). This approach involves building moment functions that contain parameters to capture the unobserved, unrestricted covariances between the instruments employed and the unobserved factor component. In particular, notice that under Assumption 1(iv) there exists a  $d \times 1$  vector  $\mathbf{w}_i$  that contains potential “instruments”, i.e. variables within the model that are potentially orthogonal to the purely idiosyncratic error component; the precise set of variables that satisfies this orthogonality property may depend on period  $t$ . Thus, let  $\mathbf{S}_t$  be a  $\zeta_t \times d$  selector matrix of 0’s and 1’s that picks up from  $\mathbf{w}_i$  variables at period  $t$  that are uncorrelated with  $\varepsilon_{it}$ . Thus, in each period  $t$ ,

$\zeta_t > 0$  instruments are available, expressed in vector form as  $\mathbf{z}_{it} = \mathbf{S}_t \mathbf{w}_i$ , for which the *orthogonality condition*  $E(\mathbf{z}_{it} \varepsilon_{it}) = \mathbf{0}$  holds true. The total number of moment conditions is given by  $\zeta = \sum_{t=1}^T \zeta_t$ . Let  $\mathbf{S} = \text{diag}(\mathbf{S}_1; \dots; \mathbf{S}_T)$ , and  $\mathbf{Z}'_i \equiv \mathbf{S}(\mathbf{I}_T \otimes \mathbf{w}_i)$ , while

$$\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau) \equiv \mathbf{Z}'_i \{ \mathbf{y}_i - \rho \mathbf{y}_{i,-1}^{(b)} - \eta_\tau \mathbf{y}_{i,-1}^{(a)} - \mathbf{X}_i^{(b)} \boldsymbol{\beta} - \mathbf{X}_i^{(a)} \boldsymbol{\delta}_\tau \} - \mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f}, \quad (3)$$

where  $\mathbf{G} = E(\mathbf{w}_i \boldsymbol{\lambda}'_i)$  and  $\boldsymbol{\theta}_\tau = (\mathbf{g}', \mathbf{f}', \rho, \boldsymbol{\beta}', \eta_\tau, \boldsymbol{\delta}'_\tau)'$  with  $\mathbf{g} = \text{vec}(\mathbf{G})$ .  $\mathbf{G}$  denotes the matrix that contains the unrestricted covariances between the matrix of instruments and the factor loadings.

**Remark 1.** The above definition for  $\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$  requires some explanation. To develop a Method of Moments estimator for inference based on the foregoing moment conditions, one typically considers terms of the form

$$\mathbf{e}_i(\boldsymbol{\theta}_\tau) = \mathbf{y}_i - [\rho \mathbf{y}_{i,-1}^{(b)} + \eta_\tau \mathbf{y}_{i,-1}^{(a)} + \mathbf{X}_i^{(b)} \boldsymbol{\beta} + \mathbf{X}_i^{(a)} \boldsymbol{\delta}_\tau + (\mathbf{I}_T \otimes \boldsymbol{\lambda}'_i) \mathbf{f}]$$

with  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau) = E[\mathbf{Z}'_i \mathbf{e}_i(\boldsymbol{\theta}_\tau)]$ . Thus,  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau^0) = \mathbf{0}$  is treated as a moment condition and  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$  as the corresponding moment function. The method adopted in the literature for simple settings suggests to estimate  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau^0)$  by  $N^{-1} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{e}_i(\boldsymbol{\theta}_\tau)$ . However, in the present case  $\boldsymbol{\lambda}_i$  is an unobserved random variable and so the foregoing suggestive method is not suitable. Therefore, we adopt a method which essentially integrates out the unobserved random variable.

Taking expectations in the expression above yields the following vector-valued moment function:

$$\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau) \equiv E[\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)] = \mathbf{m} - \rho \mathbf{m}_{-1}^{(b)} - \eta_\tau \mathbf{m}_{-1}^{(a)} - \mathbf{M}^{(b)} \boldsymbol{\beta} - \mathbf{M}^{(a)} \boldsymbol{\delta}_\tau - \mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f}, \quad (4)$$

with  $\mathbf{m} = E[\mathbf{Z}'_i \mathbf{y}_i]_{\zeta \times 1}$ ,  $\mathbf{m}_{-1}^{(j)} = E[\mathbf{Z}'_i \mathbf{y}_{i,-1}^{(j)}]_{\zeta \times 1}$ , and  $\mathbf{M}^{(j)} = E[\mathbf{Z}'_i \mathbf{X}_i^{(j)}]_{\zeta \times k}$ , for  $j = \{a, b\}$ . It follows that for the true parameter vector  $\boldsymbol{\theta}_\tau^0$ , we have  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau^0) = \mathbf{0}$ .

**Remark 2.** Observe that the last term of expression (4) can be written as  $\mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f} = \mathbf{S}(\mathbf{F} \otimes \mathbf{I}_d) \mathbf{g} = \mathbf{S} \text{vec}(\mathbf{G} \mathbf{F}')$ . But since  $\mathbf{S} \text{vec}(\mathbf{G} \mathbf{F}') = \mathbf{S} \text{vec}(\mathbf{G} \mathbf{U} \mathbf{U}^{-1} \mathbf{F}')$  for any  $r \times r$  invertible matrix  $\mathbf{U}$ , the parameters  $\mathbf{G}$  and  $\mathbf{F}$  are not identified per se without normalizing restrictions. This issue is well-known in factor models. Following standard practice, in what follows we assume that a set of normalizing restrictions is available to ensure identification. The actual choice is not important; see, for example, Robertson and Sarafidis (2015). Therefore,  $\boldsymbol{\theta}_\tau^0$  corresponds hereafter to the true parameter vector containing the normalized values of  $\mathbf{f}^0$  and  $\mathbf{g}^0$ .

Now, suppose that the null hypothesis is true. Hence model (1) reduces to

$$y_{it} = \rho^0 y_{i,t-1} + \mathbf{x}'_{it} \boldsymbol{\beta}^0 + \boldsymbol{\lambda}'_i \mathbf{f}_t^0 + \varepsilon_{it}, \quad (t = 1, \dots, T). \quad (5)$$

By arguments very similar to those leading to (4), we obtain the moment function

$$\boldsymbol{\mu}_1(\boldsymbol{\theta}_\tau) = \mathbf{m} - \rho \mathbf{m}_{-1} - \mathbf{M} \boldsymbol{\beta} - \mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f}, \quad (6)$$

where  $\mathbf{m}_{-1} = \mathbf{m}_{-1}^{(b)} + \mathbf{m}_{-1}^{(a)}$  and  $\mathbf{M} = \mathbf{M}^{(b)} + \mathbf{M}^{(a)}$ . Note that the foregoing moment function involves only the parameter  $\boldsymbol{\theta}_1 := (\mathbf{g}', \mathbf{f}', \rho, \boldsymbol{\beta}')'$ , which is a subvector of  $\boldsymbol{\theta}_\tau$ . In what follows, we will slightly abuse notation and define  $\boldsymbol{\mu}_1(\boldsymbol{\theta}_1)$  and  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$  as having the same value when the null hypothesis is satisfied. Consequently,  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$  is well defined for  $\tau \geq 1$ . More specifically, it is the moment function in (4) for the full model (1) if  $\tau \geq 2$ , and the moment function in (6) for the reduced model (5) if  $\tau = 1$ .

Before we formalize the testing methodology proposed in our paper, it will be instructive to use a simple example in order to illustrate the moment functions described above, which are employed by our estimator. To this end, we consider for simplicity the case where  $T = 3$ ,  $r = 1$  and  $\boldsymbol{\beta} = \mathbf{0}$ .

**Example 1.** Under the null hypothesis we have

$$\begin{aligned} E(\boldsymbol{\mu}_{1,i}(\boldsymbol{\theta}_1)) &= \begin{pmatrix} E(\mathbf{z}_{i1} \varepsilon_{i1}) \\ E(\mathbf{z}_{i2} \varepsilon_{i2}) \\ E(\mathbf{z}_{i3} \varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} E(y_{i0} \varepsilon_{i1}) \\ E(y_{i0} \varepsilon_{i2}) \\ E(y_{i1} \varepsilon_{i2}) \\ E(y_{i0} \varepsilon_{i3}) \\ E(y_{i1} \varepsilon_{i3}) \\ E(y_{i2} \varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} m_{01} \\ m_{02} \\ m_{12} \\ m_{03} \\ m_{13} \\ m_{23} \end{pmatrix} - \rho \begin{pmatrix} m_{00} \\ m_{01} \\ m_{11} \\ m_{02} \\ m_{12} \\ m_{22} \end{pmatrix} - \begin{pmatrix} g_0 f_1 \\ g_0 f_2 \\ g_1 f_2 \\ g_0 f_3 \\ g_1 f_3 \\ g_2 f_3 \end{pmatrix} \\ &= \mathbf{m} - \rho \mathbf{m}_{-1} - \mathbf{S} \mathbf{vec}(\mathbf{G} \mathbf{F}'), \end{aligned} \quad (7)$$

where  $m_{s,t} = E(y_{is} y_{it})$  and  $\boldsymbol{\theta}_1 = (g_0, g_1, g_2, f_1, f_2, f_3, \rho)'$ . 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

$\tau = 3$  we have

$$\begin{aligned}
E(\boldsymbol{\mu}_{3,i}(\boldsymbol{\theta}_3)) &= \begin{pmatrix} E(\mathbf{z}_{i1}\varepsilon_{i1}) \\ E(\mathbf{z}_{i2}\varepsilon_{i2}) \\ E(\mathbf{z}_{i3}\varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} E(y_{i0}\varepsilon_{i1}) \\ E(y_{i0}\varepsilon_{i2}) \\ E(y_{i1}\varepsilon_{i2}) \\ E(y_{i0}\varepsilon_{i3}) \\ E(y_{i1}\varepsilon_{i3}) \\ E(y_{i2}\varepsilon_{i3}) \end{pmatrix} = \begin{pmatrix} m_{01} \\ m_{02} \\ m_{12} \\ m_{03} \\ m_{13} \\ m_{23} \end{pmatrix} - \rho \begin{pmatrix} m_{00} \\ m_{01} \\ m_{11} \\ 0 \\ 0 \\ 0 \end{pmatrix} - \eta_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ m_{02} \\ m_{12} \\ m_{22} \end{pmatrix} - \begin{pmatrix} g_0 f_1 \\ g_0 f_2 \\ g_1 f_2 \\ g_0 f_3 \\ g_1 f_3 \\ g_2 f_3 \end{pmatrix} \\
&= \mathbf{m} - \rho \mathbf{m}_{-1}^{(b)} - \eta_3 \mathbf{m}_{-1}^{(a)} - \mathbf{S}vec(\mathbf{G}\mathbf{F}'), \tag{8}
\end{aligned}$$

with  $\boldsymbol{\theta}_3 = (g_0, g_1, g_2, f_1, f_2, f_3, \rho, \eta_3)'$ .  $\square$

**Remark 3.** As it is typically the case with breakpoint detection in general, identification of the structural break requires certain restrictions on the date when the break occurs. For example, in a time series autoregressive model it is required that the break occurs at period  $\tau > p + 1$ , where  $p$  denotes the order of the AR process. On the other hand, in a standard dynamic panel data model of first order, identification requires that the break takes place at period  $\tau \geq 2$  because first-differencing of the model removes the first time period. In the present paper identification also depends on the number of factors, as well as on the properties of the regressors. For  $r = 1$ , identification requires  $\tau \geq 3$ ; to see this, notice that if the break occurs at period  $t = 2$ , then  $\rho$  is not identified because it is only “observable” at period  $t = 1$ , which contains a single estimating equation, given by  $m_{01} - \rho m_{00} - g_0 f_1 = 0$ , and 2 parameters that do not appear elsewhere, namely  $\rho$  and  $f_1$ . On the other hand, it is worth mentioning that if instruments with respect to exogenous regressors or other exogenous variables are used, identification can be accomplished even with  $\tau = 2$ .

**Remark 4.** The vector-valued moment function above 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 (say)  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$  reduces to  $\mathbf{S}(\boldsymbol{\nu}_T \otimes \mathbf{I}_d)\mathbf{g}$ , where  $\boldsymbol{\nu}_T$  is a  $T \times 1$  vector of ones. Therefore, our model incorporates the fixed effects panel data model as a special case.

As it will be shown in the next section, our estimator of  $\boldsymbol{\theta}_\tau^0$  is the minimiser of the following quadratic form:  $\hat{Q}_\tau(\boldsymbol{\theta}_\tau) = \hat{\boldsymbol{\mu}}_\tau'(\boldsymbol{\theta}_\tau) \widehat{\mathbf{W}} \hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau)$ , where  $\widehat{\mathbf{W}}$  is some positive definite weighting matrix, and  $\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1} \sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$ .

## 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 homoskedasticity or other distributional properties of the error process. In what follows, we list the remaining assumptions required to establish the main asymptotic properties of our method. Let  $\Theta$  denote the parameter space that is obtained by a particular set of normalizing restrictions on  $(\mathbf{G}, \mathbf{F})$ .

**Assumption 2.** *The parameter space  $\Theta$  is compact and contains the true value  $\boldsymbol{\theta}_\tau^0$  in its interior. For  $\tau \geq 1$ , the population moment function vector  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$  is equal to  $\mathbf{0}$  if and only if  $\boldsymbol{\theta}_\tau = \boldsymbol{\theta}_\tau^0$ .*

**Assumption 3.** *The variance-covariance matrix of the moment functions evaluated at  $\boldsymbol{\theta}_\tau^0$ , which is defined as  $\boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0) \equiv E[\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau^0)\boldsymbol{\mu}'_{\tau,i}(\boldsymbol{\theta}_\tau^0)]$ , and the derivative matrix of moment functions  $\boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0) \equiv E[(\partial/\partial\boldsymbol{\theta}'_\tau)\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau^0)]$ , 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. In particular, let  $\hat{Q}_\tau(\boldsymbol{\theta}_\tau) = \hat{\boldsymbol{\mu}}'_\tau(\boldsymbol{\theta}_\tau)\widehat{\mathbf{W}}\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau)$ , and define the *Factor Instrumental Variable* [FIV] estimator  $\hat{\boldsymbol{\theta}}_\tau$  of  $\boldsymbol{\theta}_\tau^0$  as follows:

$$\hat{\boldsymbol{\theta}}_\tau = \arg \min_{\boldsymbol{\theta}_\tau \in \Theta} \hat{Q}_\tau(\boldsymbol{\theta}_\tau). \quad (9)$$

We will show that  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} |\hat{Q}_\tau(\boldsymbol{\theta}_\tau) - Q_\tau(\boldsymbol{\theta}_\tau)|$  converges to zero in probability, where  $Q_\tau(\boldsymbol{\theta}_\tau) = \boldsymbol{\mu}'_\tau(\boldsymbol{\theta}_\tau)\mathbf{W}\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$  and  $\mathbf{W}$  is a positive definite weighting matrix. The consistency of  $\hat{\boldsymbol{\theta}}_\tau$  follows from this result.

The optimal choice of the weighting matrix is obtained by setting  $\widehat{\mathbf{W}} = \hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau)$  in (9) (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  $\widehat{\mathbf{W}} = \mathbf{I}_\zeta$ , which provides a first-step 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 and then rely on (9) with  $\widehat{\mathbf{W}} = \hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)})$  if  $\hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)})$  is non-singular, otherwise  $\widehat{\mathbf{W}} = [\hat{\boldsymbol{\Phi}}_\tau(\hat{\boldsymbol{\theta}}_\tau^{(1)}) + N^{-1}\mathbf{I}]^{-1}$ .

Below we establish an important lemma for this paper; the proof is given in the Appendix.

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

In studies involving structural breaks, the case when the break point  $\tau$  is given, and the case when it is unknown are both of interest. Hence, we will study both cases. Since  $\hat{\boldsymbol{\theta}}_1$  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, a suitable statistic for testing  $H_0$  vs  $H_{(\tau)}$  is the following difference in the squared lengths of the sample moment functions:

$$\psi_\tau = N[\hat{Q}_1(\hat{\boldsymbol{\theta}}_1) - \hat{Q}_\tau(\hat{\boldsymbol{\theta}}_\tau)]. \quad (10)$$

We now state the main result about the asymptotic null distribution of  $\psi_\tau$ . The proof is given in the Appendix.

**Theorem 1.** *Suppose that Assumptions 1-3 hold, and  $\tau \geq 2$ . Then, under the null hypothesis  $H_0$ , the test statistic  $\psi_\tau = N[\hat{Q}_1(\hat{\boldsymbol{\theta}}_1) - \hat{Q}_\tau(\hat{\boldsymbol{\theta}}_\tau)]$  is asymptotically distributed as chi-squared with degrees of freedom equal to  $\dim(\eta_\tau) + \dim(\boldsymbol{\delta}_\tau) = 1 + k$ .*

This result is sufficient for us to test  $H_0$  vs  $H_{(\tau)}$ . Now, we use this to develop a test of  $H_0$  against the more general alternative  $H_1$  wherein the breakpoint  $\tau_0$  is unknown. To this end, we consider the foregoing test for each possible value of  $\tau_0$  and then combine them. Recall that, we are interested to test  $H_0$  against the alternative that a structural break occurs at some time  $\tau_0$  in  $\{\tau_1, \dots, \tau_L\}$ . Let

$$\boldsymbol{\psi} = [\psi_{\tau_1}, \dots, \psi_{\tau_L}]', \quad (11)$$

where  $\psi_{\tau_\ell} = N[\hat{Q}_1(\hat{\boldsymbol{\theta}}_1) - \hat{Q}_{\tau_\ell}(\hat{\boldsymbol{\theta}}_{\tau_\ell})]$  has been proposed for testing  $H_0$  against  $H_{(\tau_\ell)}$ . Clearly, if  $H_0$  is true then  $\hat{Q}_1(\hat{\boldsymbol{\theta}}_1)$  and  $\{\hat{Q}_{\tau_1}(\hat{\boldsymbol{\theta}}_{\tau_1}), \dots, \hat{Q}_{\tau_L}(\hat{\boldsymbol{\theta}}_{\tau_L})\}$  are all estimators of the same quantity and hence  $\max\{\psi_{\tau_1}, \dots, \psi_{\tau_L}\}$  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  $\max\{\psi_{\tau_1}, \dots, \psi_{\tau_L}\}$  is expected to be large. Therefore, in order to test  $H_0$  vs  $H_1$  we propose using the statistic

$$\psi_{\max} := \max_{\ell=1, \dots, L} \{\psi_{\tau_\ell}\}, \quad (12)$$

and reject the null for large enough values of  $\psi_{\max}$ . The unknown break point  $\tau_0$  can be estimated by  $\tau_s$ , where  $\psi_{\tau_s} = \psi_{\max}$ .

Now, to state the main result about the distribution of the test statistic,  $\psi_{max}$ , let us introduce the following notation: For a full column rank matrix,  $\mathbf{B}$ , let  $\mathbf{M}_B$  and  $\mathbf{P}_B$  be two projection matrices defined by  $\mathbf{P}_B \equiv \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$  and  $\mathbf{M}_B \equiv \mathbf{I} - \mathbf{P}_B$  respectively.

Suppose that the null hypothesis holds true. Let  $\Phi_{\tau_\ell}^{-1/2}$  denote the symmetric square root of  $\Phi_{\tau_\ell}^{-1}$  so that  $\Phi_{\tau_\ell}^{-1} = \Phi_{\tau_\ell}^{-1/2}\Phi_{\tau_\ell}^{-1/2}$ . Let  $\mathbf{V}_{\tau_\ell} = \mathbf{M}_{\Phi_{\tau_\ell}^{-1/2}\mathbf{\Gamma}_{\tau_\ell}} - \mathbf{M}_{\Phi_{\tau_\ell}^{-1/2}\mathbf{\Gamma}_{\tau_\ell}}$ , where all the quantities are evaluated at the true value of the unknown parameter.

The next theorem provides the essential result for applying  $\psi_{max}$  for testing  $H_0$  against  $H_1$ .

**Theorem 2.** *Suppose that Assumptions 1-3 and the null hypothesis  $H_0$  are satisfied. Let  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_\zeta)$  where  $\zeta$  is the number of moment conditions, and let  $\mathbf{V}_{\tau_1}, \dots, \mathbf{V}_{\tau_L}$  be evaluated at the true value of the parameter specified by the null hypothesis. Then,  $\psi \xrightarrow{d} \{\mathbf{z}'[\mathbf{V}_{\tau_1}, \dots, \mathbf{V}_{\tau_L}](\mathbf{I}_L \otimes \mathbf{z})\}'$ . Consequently, the test statistic  $\psi_{max}$  is asymptotically distributed as  $\Xi \equiv \max\{\mathbf{z}'\mathbf{V}_{\tau_1}\mathbf{z}, \dots, \mathbf{z}'\mathbf{V}_{\tau_L}\mathbf{z}\}$ .*

Since the asymptotic null distribution of  $\psi_{max}$  depends on the nuisance parameter  $\theta_1^0$  through  $\mathbf{V}_{\tau_\ell}(\theta_1^0)$ , we propose to approximate the distribution of  $\psi_{max}(\theta_1^0)$  by that of  $\psi_{max}(\hat{\theta}_1)$  conditional on  $\hat{\theta}_1$ . Therefore, critical values can be estimated by simulation. Details are provided in Section 3 below.

## 3 Monte Carlo Simulations

### 3.1 Simulation Design

In this section we investigate the finite-sample properties of the test introduced previously. Our focus is on the impact of sample size, as well as on the location and magnitude of the break on the performance of our statistic. We study the pure AR(1) model, the choice of which is mainly motivated by our application that follows. The DGP is given by

$$y_{it} = \begin{cases} \rho^0 y_{i,t-1} + \lambda_i f_t^0 + \varepsilon_{it}; & t < \tau \\ \eta_\tau^0 y_{i,t-1} + \lambda_i f_t^0 + \varepsilon_{it}; & t \geq \tau, \end{cases} \quad (13)$$

for  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ , where  $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ ,  $\lambda_i \sim \mathcal{N}(0, \sigma_\lambda^2)$ ,  $f_t^0 \sim \mathcal{N}(0, \sigma_f^2)$ , and  $\sigma_\varepsilon^2 = \sigma_f^2 = 1$ . We set the factor number as  $r = 1$ , noting that we have also examined a two-factor model; the results are very similar and therefore we are not including them here to save space. Define  $\pi \equiv \sigma_\lambda^2 / (1 + \sigma_\lambda^2)$ , which is the variance of the factor component  $\lambda_i f_t^0$  as a proportion of the variance of the total error term  $\lambda_i f_t^0 + \varepsilon_{it}$ .

The initial observation is generated as  $y_{i0} = \lambda_i/(1 - \rho^0) + \mathcal{N}(0, 1)$ . The vector of possible instruments is given by  $\mathbf{w}_i = (y_{i0}, \dots, y_{i,T-1})'$ . For each  $t$ , we choose  $\mathbf{z}_{it} = (y_{i0}, \dots, y_{i,t-1})'$ . We set  $N \in \{60, 120, 300, 600, 1200\}$ ,  $T \in \{6, 8\}$ ,  $\tau_0 \in \{4, 6\}$ , and  $\omega^0 = \eta_\tau^0 - \rho^0 \in \{-0.05, 0, 0.05, 0.10, 0.15\}$ . We fix  $\rho^0 = 0.5$  and  $\pi = 0.5$ . All the simulations are conducted using 10,000 replications.

The estimation algorithm that we employ involves an iterative procedure. In particular, notice that if either  $\mathbf{f}$  or  $\mathbf{g}$  is held fixed then equations (4) and (6) become linear in the remaining parameters. This feature can simplify the computations; specifically, to compute the global minimum of the objective function  $\hat{Q}_\tau(\boldsymbol{\theta}_\tau)$  one can proceed as follows: (a) for a given fixed value of  $\mathbf{f}$ , minimize  $\hat{Q}_\tau(\boldsymbol{\theta}_\tau)$  with respect to the remaining parameters, (b) hold the value of  $\mathbf{g}$  obtained in the previous step fixed and minimize  $\hat{Q}_\tau(\boldsymbol{\theta}_\tau)$  with respect to the remaining parameters, (c) hold the value of  $\mathbf{f}$  obtained in the previous step fixed and minimize  $\hat{Q}_\tau(\boldsymbol{\theta}_\tau)$  with respect to the remaining parameters, (d) repeat the previous two steps until convergence.

In terms of the implementation of our test statistic, in practice this requires to compute simulated critical values. To this end, we adopted the following steps:

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

$$\hat{c} = \max\{\mathbf{z}'\hat{\mathbf{V}}_{\tau_1}(\hat{\boldsymbol{\theta}}_1)\mathbf{z}, \dots, \mathbf{z}'\hat{\mathbf{V}}_{\tau_L}(\hat{\boldsymbol{\theta}}_1)\mathbf{z}\},$$

where  $\hat{\boldsymbol{\theta}}_1$  is the estimator of  $\boldsymbol{\theta}_\tau^0$ , under the null hypothesis.

- (c) Repeat steps (a)-(b)  $K$  times, say  $K = 10,000$ , and generate  $K$  values of  $\hat{c}$ , which we denote as  $\hat{c}_1, \dots, \hat{c}_K$ .
- (d) Let  $\hat{c}_{(0.95)}$  be the 95<sup>th</sup> percentile of  $\{\hat{c}_1, \dots, \hat{c}_K\}$ .

Note that  $\mathbf{V}_{\tau_\ell}(\boldsymbol{\theta}_\tau)$  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  $\mathbf{z}'\mathbf{V}_{\tau_\ell}(\boldsymbol{\theta}_\tau^0)\mathbf{z}$ , where  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_\zeta)$ , can be approximated by that of  $\mathbf{z}'\hat{\mathbf{V}}_{\tau_\ell}(\hat{\boldsymbol{\theta}}_1)\mathbf{z}$ . 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.

## 3.2 Simulation Results

Table 1 reports results on size and power of our test statistic for the case where the breakpoint is known. Since  $\omega^0 = \eta_\tau^0 - \rho^0$ , results with respect to  $\omega^0 = 0$  correspond to the empirical size and otherwise to (size-adjusted) power (nominal size is set equal 5% throughout). The panel on the left (correspondingly, on the right) corresponds to the case where the break point occurs at  $t = 4$  (correspondingly,  $t = 6$ ). Clearly, the performance of our test statistic is more than satisfactory in both cases. In particular, while the test appears to exhibit small size distortions for  $N$  small, empirical size approaches quickly its nominal value as  $N$  gets larger, while power increases at a high rate as well, as  $N$  grows. As expected, whether the break has a positive or negative sign appears to make little difference to the results. On the other hand, a larger value of  $T$  is associated with a further slight size distortion when  $\tau = 4$  but the performance remains satisfactory. We note that the dependence of the performance on the value of  $T$  is well established in the literature - the main reason being that the number of moment conditions becomes larger with higher values of  $T$ , while the size of the cross-sectional dimension remains constant (c.f. Bun and Sarafidis 2015).

When  $\tau = 6$ , such that for  $T = 6$  the break takes place at the very end of the sample, the test appears to be slightly more size-distorted and has less power compared to the case where  $\tau = 4$ . This is not surprising because for  $T = 6$ ,  $\tau = 6$  means that the only moment conditions available to detect the break are those in the last sample period (6 moment conditions). However, again as  $N$  increases the performance of the statistic improves considerably in this case. A substantial improvement is also visible for  $T = 8$ , for which value the break point  $\tau = 6$  no longer corresponds to the final period of the sample.

Table 1: **Size and Power (%) for Given Breakpoint Detection**

T	N	$\tau_0 = 4, \rho^0 = 0.5$					$\tau_0 = 6, \rho^0 = 0.5$				
		Size $\omega^0 = 0$	Power for various values of $\omega^0$				Size $\omega^0 = 0$	Power for various values of $\omega^0$			
		$\omega^0 = 0$	-0.05	0.05	0.10	0.15	$\omega^0 = 0$	-0.05	0.05	0.10	0.15
6	60	6.77	5.67	5.54	7.98	11.94	11.17	4.91	4.99	5.81	6.24
	120	6.19	6.82	6.73	12.41	24.40	8.32	5.38	5.59	6.46	8.37
	300	5.23	10.95	11.02	28.90	55.59	6.02	5.89	5.74	8.64	14.46
	600	5.43	16.90	17.34	51.35	81.46	5.46	7.07	7.00	14.37	27.92
	1200	5.11	29.70	30.45	78.97	93.91	4.82	9.76	10.16	25.47	45.70
8	60	7.39	5.84	6.07	8.87	13.47	8.04	5.65	5.78	8.41	12.71
	120	6.41	7.46	7.39	15.56	29.89	6.97	7.23	6.85	14.57	27.71
	300	5.80	12.79	13.46	39.62	70.46	5.44	13.81	14.49	39.49	67.98
	600	5.35	23.27	23.71	69.62	94.18	5.14	23.84	23.47	67.16	90.25
	1200	5.36	40.27	42.14	91.75	99.51	5.33	39.39	40.73	88.91	97.12

Table 2 reports results on empirical size and power for our test statistic when the breakpoint is unknown. We focus on the case  $\tau_0 = 4$ , while the results for  $\tau_0 = 6$  are analogous to those in Table 1. As before,  $\omega^0 = 0$  corresponds to empirical size and otherwise to (size-adjusted) power. Compared to the case where the breakpoint is known, size performance remains satisfactory and power is slightly smaller, which is expected. For power, we also report the proportion of times the breakpoint is located successfully. Thus, the sum of the values along a particular row equals 1. As an example, consider the row that corresponds to  $N = 60$ ,  $\omega^0 = -0.05$ , where empirical power is 5.17. This means that conditionally upon rejecting the null hypothesis 5.17% of the times, the breakpoint is estimated to be at  $\tau = 4$  for 23.60% of the times, and so on. For higher values of  $N$ , the frequency of detecting the correct date quickly increases uniformly.

Table 2: **Size and Power (%) for Unknown Breakpoint Detection:**  $T = 6$

$\omega^0$	N	Rejection Rate	Breakpoint Detected; $\tau_0 = 4$			
			$\tau = 4$	$\tau = 3$	$\tau = 5$	$\tau = 6$
0.00	60	8.21	N/A	N/A	N/A	N/A
	120	7.26	N/A	N/A	N/A	N/A
	300	5.71	N/A	N/A	N/A	N/A
	600	4.84	N/A	N/A	N/A	N/A
	1200	5.76	N/A	N/A	N/A	N/A
-0.05	60	5.17	23.60	15.86	15.28	45.26
	120	6.00	27.33	22.67	17.67	32.33
	300	8.17	40.64	21.42	18.12	19.83
	600	12.72	49.53	19.18	17.93	13.37
	1200	21.06	60.87	14.20	13.68	11.25
0.05	60	5.38	16.36	16.17	18.59	48.89
	120	5.82	21.13	19.93	18.73	40.21
	300	8.88	36.04	23.54	17.79	22.64
	600	13.73	47.56	17.92	18.72	15.81
	1200	21.81	60.34	13.30	15.27	11.10
0.10	60	6.60	23.33	14.70	18.94	43.03
	120	9.31	36.95	16.65	16.43	29.97
	300	22.12	57.19	14.01	14.01	14.78
	600	43.55	66.41	11.39	12.10	10.10
	1200	69.76	79.09	6.45	8.19	6.28
0.15	60	8.63	29.43	11.94	17.38	41.25
	120	16.38	46.95	13.25	14.47	25.34
	300	44.67	68.79	8.98	11.04	11.19
	600	75.23	78.79	5.93	8.04	7.24
	1200	91.20	87.29	2.87	5.81	4.02

For power ( $\omega^0 \neq 0$ ) the entries in the last 4 columns in each row sum up to 1. As an example, consider the row that corresponds to  $N = 60$ ,  $\omega^0 = -0.05$ , where empirical power is 5.17. This means that conditionally upon rejecting the null hypothesis 5.17% of the times, the breakpoint is estimated to be at  $\tau = 4$  for 23.60% of the times, and so on.

### 3.3 Comparison Simulation

For the purposes of comparison, we also examine the performance of the testing procedure developed by De Wachter and Tzavalis (2012), denoted as DWT hereafter. This is based on the first-differenced GMM estimator and is developed for the dynamic panel data model with individual-specific effects. We consider two models; the first one is the one-way error components structure, i.e.  $f_t^0 \equiv 1$  for all  $t$ , while the second contains a single genuine factor component. In the factor case, the first-differenced GMM estimator is not consistent. Notwithstanding, it is of practical interest to see how far DWT can go in addressing the problem. In the fixed effects case, both statistics are valid. However, our estimator over-parameterizes the model, and therefore it is interesting to see if this has a negative effect on our test statistic. Essentially in the first model DWT plays at home while in the second model it plays away.

The reported results in Table 3 correspond to the known breakpoint case when  $\tau_0 = 4$  is the alternative hypothesis. As we can see, in the fixed effects case the size performance for our statistic,  $\psi_4$ , and DWT is similar and close to the nominal 5% level. On the other hand, in the single factor case DWT appears to be largely distorted in terms of size, and in fact the probability of rejecting the null hypothesis increases as  $N$  grows. In terms of (size-adjusted) power, DWT has larger power than  $\psi_4$  in the fixed effects case, which is plausible given that our model specification is more general. For both statistics power increases with  $N$ , as well as  $\omega^0$ . However, for the factor model our test statistic remains consistent and its power increases with the sample size as well as with the size of  $\omega^0$ , i.e. the magnitude of the break. By contrast, the power of DWT appears to be largely invariant to changes in  $\omega^0$  and  $N$ , which suggests that DWT is not consistent in this case.

Table 3: **Size and Power (%) for  $\psi_4$  and DWT for Panel Data Models with Fixed Effects and a Factor Structure:  $T = 6$**

$\omega^0$	N	Fixed Effects Model		Factor Model	
		$\psi_4$	DWT	$\psi_4$	DWT
0.00	60	10.17	9.65	6.77	13.84
	120	8.08	8.27	6.19	23.18
	300	5.52	6.69	5.23	33.93
	600	5.29	5.82	5.43	41.32
	1200	5.41	5.39	5.11	46.02
-0.05	60	5.17	7.38	5.67	5.23
	120	5.61	10.05	6.82	5.16
	300	8.10	21.89	10.95	5.13
	600	11.17	38.62	16.90	4.99
	1200	16.33	51.90	29.70	5.06
0.05	60	4.89	6.23	5.54	5.06
	120	5.31	10.16	6.73	4.89
	300	6.90	24.79	11.02	5.44
	600	9.62	42.23	17.34	4.61
	1200	14.45	53.71	30.45	4.87
0.10	60	5.01	11.36	7.98	5.20
	120	6.65	27.44	12.41	5.22
	300	13.15	51.50	28.90	5.62
	600	21.89	56.74	51.35	4.89
	1200	39.51	59.21	78.97	4.70
0.15	60	6.21	22.19	11.94	5.04
	120	8.10	45.35	24.40	5.17
	300	21.61	56.15	55.59	5.55
	600	39.82	59.85	81.46	4.83
	1200	69.34	65.87	93.91	4.97

Note:  $\omega^0$  represents different levels of the magnitude of the break.

## 4 Empirical Application

In this section we will apply our methodology to investigate the empirical validity of the well known “Law of Proportionate Effect”, or otherwise Gibrat’s ‘Law’ with regards to the US financial industry. Gibrat’s law postulates that the size of a firm and its growth rate are independent. More specifically, consider the following equation:

$$y_{it} = \rho^0 y_{i,t-1} + u_{it}; \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (14)$$

where  $y_{it}$  denotes some measure of size (expressed in natural logarithms) for firm  $i$  at time  $t$ . Subtracting  $y_{i,t-1}$  from both sides yields

$$\Delta y_{it} = \delta^0 y_{i,t-1} + u_{it}, \quad (15)$$

where  $\delta^0 = \rho^0 - 1$ .  $\Delta y_{it}$  is the firm’s growth rate at period  $t$ . Gibrat’s ‘Law’ implies the restriction  $\rho^0 = 1$ , or equivalently  $\delta^0 = 0$ . To see this more closely, notice that for  $\rho^0 = 1$  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,  $y_{i0}$ , remains constant regardless of the value of  $t$ , it is clear that the growth rate of firm  $i$ ,  $\Delta y_{it}$ , 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 what appears to be an empirical regularity with respect to the distribution of firm size across several industries. In particular, this is often highly skewed, given that many industries consist of a small (respectively, large) number of large (respectively, small or medium-sized) firms. According to Simon and Bonini (1958), there is a connection between Gibrat’s ‘Law’ and the returns to scale in a given industry. That is, under (approximate) 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 value.

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 a multi-factor structure:

$$u_{it} = \boldsymbol{\lambda}'_i \mathbf{f}_t + \varepsilon_{it}. \quad (16)$$

Allowing for a common factor component is important in the present case for several reasons. First of all, a subset of the factors might capture the effect of the age of the banking institution, if this effect is statistically significant, given that ‘age’ is not observed

in our sample. Let  $x_{it}$  denote the variable age with the bank-specific coefficient given by  $\beta_i$ . Since age increases at the same rate for all companies every year, we may set

$$\beta_i x_{it} = \beta_i(x_{i0} + t) = \gamma_i + \beta_i t, \quad (17)$$

where  $x_{i0}$  denotes the age of the bank at the beginning of the sample period, and  $\gamma_i = \beta_i + x_{i0}$ . 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 an effect or not on the growth of firms in the banking industry, it is easier to allow this variable to be absorbed by the factors.

**Remark 5.** Notice of course that in practice it is impossible to distinguish between the effect of age and the presence of a separate, genuine deterministic linear trend. However, our aim is not to identify these factors, but rather to identify correctly whether there is a break in the autoregressive coefficient or not, controlling for such factors. Failing to do so may result in invalid inferences.

Secondly, factors may capture common shocks that have hit all individual banks, albeit with different intensities, due to (say) the recent global financial crisis [GFC]. In fact, the presence of the GFC in our sample may cast doubt about the common assumption employed by dynamic panel data estimators – namely, that the coefficients of the model remain constant over time. Therefore, the application of our methodology to this research question is particularly important.

We consider the monetary value of assets expressed in constant prices as a measure of bank size. The results are reported in the Table 4.  $r$  denotes the number of factors fitted in the model, while  $\rho^0$  and  $\eta^0$  denote the value of the autoregressive parameters prior to and after the break, respectively. Hence the magnitude of the break is given by  $\eta^0 - \rho^0$ .  $P_t$ ,  $P_J$  and  $P_{\psi_{\max}}$  denote the  $p$ -value of the  $t$ -test, the  $p$ -value of the overidentifying restrictions test (Hansen's test) statistic and the  $p$ -value of the structural break test statistic proposed in this paper, respectively. The number of factors can be selected using a model information criterion, such as BIC. In the present context this is given by

$$BIC(r) = N \times \widehat{Q}_\tau(r) - \ln(N)/T^{0.3} \times 0.75 \times df(r), \quad (18)$$

where  $\widehat{Q}_\tau(r)$  denotes the minimum value of the objective function using  $r$  factors, while  $df(r)$  denotes the number of degrees of freedom of the model, i.e. the number of moment conditions minus the number of estimable parameters. In this case, there are 78 moment

conditions in total. The finite sample properties of this criterion for FIV estimators has been studied by Robertson and Sarafidis (2015). As a result, we choose

$$\hat{r} = \arg \min_r BIC(r). \quad (19)$$

Table 4: **Evaluation of the Gibrat’s ‘Law’ for Banks in the USA**

Assets		Coef.	Std. Err.	$t$	$P_t$	[95% Conf. Interval]	$P_J$	$P_{\psi_{\max}}$	Break Date	BIC
$r = 1$	$\rho^0$	0.60	0.02	25.00	0.00	[0.55, 0.65]	0.00	0.00	2010	-56.13
	$\eta^0$	0.72	0.03	24.00	0.00	[0.66, 0.78]				
$r = 2$	$\rho^0$	0.48	0.08	5.88	0.00	[0.32, 0.63]	0.98	0.00	2012	-80.89
	$\eta^0$	0.99	0.08	11.76	0.00	[0.82, 1.15]				
$r = 3$	$\rho^0$	0.41	0.13	3.06	0.00	[0.15, 0.67]	0.99	0.00	2011	-49.72
	$\eta^0$	0.94	0.28	4.43	0.00	[0.39, 1.49]				

<sup>a</sup>. The data are for 4,128 bank institutions in the USA for the period 2002-2014.

<sup>b</sup>. Coef., Std. Err.,  $t$ ,  $P_t$ , and [95% Conf. Interval] show the estimation results, including the estimates, standard errors,  $t$ -Statistics,  $p$ -values of the  $t$ -tests and 95% Confidence Intervals.

<sup>c</sup>.  $P_J$  and  $P_{\psi_{\max}}$  present the  $p$ -value of the Hansen’s test and the  $p$ -value of the structural break test statistic; ‘Break Date’ is the estimated year when the break occurs; BIC is used as a criterion for determining the number of factors.

Based on the results of Table 4, the optimal number of factors is two (i.e.  $\hat{r} = 2$ ). 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 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 null hypothesis  $\rho^0 = 1$  is rejected prior to the GFC (over the alternative  $\rho^0 < 1$ ) but it’s not rejected afterwards. This implies that before the GFC the growth rate of financial institutions was negatively correlated with their size, while the break in 2012 is associated with a structural change that provides empirical support towards Gibrat’s ‘Law’. One reason for this 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). Thus, our results provide support for the claim that the growth of financial institutions depends more on capitalized structure than on size nowadays.

## 5 Discussion

In this article we have proposed a structural break detection test statistic for dynamic panel data models with a 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, i.e. covariates that are correlated with the purely idiosyncratic error term.

This is very important because often the covariates receive some form of ‘feedback’ from the dependent variable. As an example, modelling the rate of crime as a function of police forces requires taking into account the fact that while the size of a police department in a particular area may exert an effect on crime, the opposite is also true.

We focus on the case where  $N$  is large and  $T$  fixed; this is an empirically relevant scenario in several applications involving socioeconomic data, such as in the field of micro-econometrics. The proposed structural break testing procedure is based on a Distance type statistic and the Method of Moments. The latter is appealing because it allows for substantial heteroskedasticity in the error term and does not impose strong distributional assumptions, such as normality. The asymptotic properties of the statistic are established for both known and unknown breakpoints since both can be of interest, depending on the application considered.

The simulation study demonstrates that our method performs well in terms of size and power, as well as in terms of locating the breakpoint correctly. Even in the special case where the common factor structure reduces to the much simpler fixed effects framework, i.e. when  $f_t = 1$  for all  $t$ , our method evidently continues to work well.

We have applied our methodology on a long-standing issue in banking, which is the

empirical validity of Gibrat's 'Law'. This postulates that the size of the firm and its growth rate are independent. Note that Gibrat's 'Law' has also been applied to examine the relationship between size and growth rate of cities. The results show that before the GFC the growth rate of financial institutions was negatively correlated with their size, i.e. small banks grew faster than larger ones. However, this pattern changed following the GFC, providing support towards Gibrat's 'Law'. This is an interesting result that may be attributed to the recently introduced regulatory framework on bank capital adequacy and market liquidity risk.

We have focused on detecting the presence of a single break. The case of multiple breaks is possible to explore, based on a model information criterion. We leave this topic for future research.

## 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. 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. (2013), Estimation of heterogeneous panels with structural breaks, Working paper, Syracuse University, Nanyang Technological University.  
**URL:** [http://takeacon.sixcore.jp/20thIPDC\\_web/All/2\\_4.1.2\\_Feng/Abstract.pdf](http://takeacon.sixcore.jp/20thIPDC_web/All/2_4.1.2_Feng/Abstract.pdf)
- 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.
- Cawley, J., Conneely, K., Heckman, J. and Vytlacil, E. (1997), *Cognitive ability, wages, and meritocracy*, Springer.

- Chan, F., Manchin-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](http://www.mssanz.org.au/MODSIM07/papers/16_s9/TestingStructura_s9_Chan_.pdf)
- Chen, J. and Gupta, A. (1997), ‘Testing and locating variance change-points with application to stock prices’, *Journal of the American Statistical Association* **92**(438), 739–747.
- De Wachter, S. and Tzavalis, E. (2012), ‘Detection of structural breaks in linear dynamic panel data models’, *Computational Statistics & Data Analysis* **56**(11), 3020–3034.
- 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.
- Eberhardt, M., Helmers, C. and Strauss, H. (2013), ‘Do spillovers matter when estimating private returns to r&d?’, *Review of Economics and Statistics* **95**(2), 436–448.
- Geweke, J. and Zhou, G. (1996), ‘Measuring the pricing error of the arbitrage pricing theory’, *Review of Financial Studies* **9**(2), 557–587.
- Greene, W. H. (2012), *Econometric Analysis*, Pearson Education.
- 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.
- Juodis, A. and Sarafidis, V. (2015), Fixed T dynamic panel data estimators with multi-factor errors, Technical report, University of Amsterdam, Monash University.
- Newey, W. K. and McFadden, D. (1994), ‘Large sample estimation and hypothesis testing’, *Handbook of Econometrics* **4**, 2111–2245.
- 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. (2014), Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso, Working paper, Shanghai Jiao Tong University and Singapore Management University.  
**URL:** [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2417560](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2417560)
- 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.

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.

Sarafidis, V. and Wansbeek, T. (2012), ‘Cross-sectional dependence in panel data analysis’, *Econometric Reviews* **31**(5), 483–531.

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

Except for Lemmas A.2 and A.3, we shall assume that Assumptions 1–3 are satisfied throughout this Appendix. To simplify the notation and details, the proofs of Lemma 1 and 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  $\lambda_i$  are scalars. Thus, our model under the alternative hypothesis  $H_1$ , reduces to the following form for some unknown  $\tau$  in  $\{2, 3, \dots, T\}$ :

$$y_{it} = \begin{cases} \rho^0 y_{i,t-1} + x_{it}\beta^0 + \lambda_i f_t^0 + \varepsilon_{it}, & t < \tau; \\ \eta_\tau^0 y_{i,t-1} + x_{it}\delta_\tau^0 + \lambda_i f_t^0 + \varepsilon_{it}, & t \geq \tau. \end{cases} \quad (20)$$

In this case, the lagged value  $y_{is}$ , where  $s = 0, 1, \dots, t-1$ , can be used as a valid instrument for individual  $i$ . We start with some necessary preliminary lemmas.

**Lemma A.1.**  $E|y_{it}|^4 < \infty$  for  $t = 1, 2, \dots, T$ .

*Proof.* We shall consider the cases  $t < \tau$  and  $t \geq \tau$  separately.

**Case 1:**  $t < \tau$ . By repeated substitution of the first part of (20) for  $t < \tau$ , we have that

$$y_{it} = (\rho^0)^t y_{i0} + \beta^0 \sum_{l=0}^{t-1} (\rho^0)^l x_{i,t-l} + \lambda_i \sum_{l=0}^{t-1} (\rho^0)^l f_{t-l}^0 + \sum_{l=0}^{t-1} (\rho^0)^l \varepsilon_{i,t-l}. \quad (21)$$

When  $T$  is fixed, each term can be bounded, following the Cauchy-Schwartz inequality and Assumption 1, which assumes  $E|x_{it}|^4$ ,  $E|\lambda_i|^4$ ,  $E|y_{i0}|^4$  and  $E|\varepsilon_{it}|^4$  are finite for  $t < \tau$ . Thus,  $E|y_{it}|^4 < \infty$  follows from (21) and the Minkowski inequality.

**Case 2:**  $t \geq \tau$ .

By repeated substitution of the second part of equation (20) for  $t \geq \tau$ , we have

$$y_{it} = (\eta_\tau^0)^{t-\tau+1} y_{i,\tau-1} + \delta_\tau^0 \sum_{l=0}^{t-\tau} (\eta_\tau^0)^l x_{i,t-l} + \lambda_i \sum_{l=0}^{t-\tau} (\eta_\tau^0)^l f_{t-l}^0 + \sum_{l=0}^{t-\tau} (\eta_\tau^0)^l \varepsilon_{i,t-l}. \quad (22)$$

The fourth moment of the first term in (22) is bounded as shown in Case 1. The last three terms in (22) have finite moments of the fourth order by the Cauchy-Schwartz inequality and Assumption 1. Finally, by the Minkowski inequality,  $E|y_{it}|^4$  is bounded. This completes the proof.  $\square$

The following two lemmas from Newey and McFadden (1994) are stated here for clarity.

**Lemma A.2.** (*Lemma 2.4 in Newey and McFadden 1994*)

Let  $\{\mathbf{z}_i\}_{i \in \mathbb{N}}$  be a sequence of finite dimensional i.i.d random vectors,  $\Theta$  be a compact subset of  $\mathbb{R}^k$ , and let  $\mathbf{a}(\mathbf{z}_i, \boldsymbol{\theta})$  denote a vector-valued function of  $\mathbf{z}_i$  and  $\boldsymbol{\theta}$ , where  $\boldsymbol{\theta} \in \Theta$ . Suppose that (i)  $\mathbf{a}(\mathbf{z}_i, \boldsymbol{\theta})$  is continuous at each  $\boldsymbol{\theta}$  with probability one, (ii) there exists a function  $d(\mathbf{z})$  such that  $\|\mathbf{a}(\mathbf{z}, \boldsymbol{\theta})\| \leq d(\mathbf{z})$  for all  $\boldsymbol{\theta} \in \Theta$ , and (iii)  $E[d(\mathbf{z})] < \infty$ . Then  $\sup_{\boldsymbol{\theta} \in \Theta} \|n^{-1} \sum_{i=1}^n \mathbf{a}(\mathbf{z}_i, \boldsymbol{\theta}) - E[\mathbf{a}(\mathbf{z}, \boldsymbol{\theta})]\| \xrightarrow{P} 0$ .

**Lemma A.3.** (*Theorem 2.1 in Newey and McFadden 1994*)

Let  $\Theta$  be a compact set in  $\mathbb{R}^k$  and  $\boldsymbol{\theta} \in \Theta$ . Suppose that there exists a function  $Q(\boldsymbol{\theta})$  such that (i)  $Q(\boldsymbol{\theta})$  is uniquely minimized at  $\boldsymbol{\theta}^0$ , (ii)  $Q(\boldsymbol{\theta})$  is continuous, and (iii)  $\hat{Q}(\boldsymbol{\theta})$  converges uniformly in probability to  $Q(\boldsymbol{\theta})$  in  $\Theta$ . Then  $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}^0$ , where  $\hat{\boldsymbol{\theta}} = \arg \min \hat{Q}(\boldsymbol{\theta})$ .

Now, we return to the proof of our main results. Lemma A.4 provides the asymptotic distribution of the sample mean of the moment functions. This will be used to derive the asymptotic normality of the Factor Instrumental Variables [FIV] estimator  $\hat{\boldsymbol{\theta}}_\tau$  in Lemma 1. Recall that  $\boldsymbol{\theta}_\tau^0$  denotes the true parameter,  $\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1} \sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$ ,  $\boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0) = E[\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau^0) \boldsymbol{\mu}'_{\tau,i}(\boldsymbol{\theta}_\tau^0)]$ ,  $\boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau^0) = E[(\partial/\partial \boldsymbol{\theta}'_\tau) \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau^0)]$ ,  $\hat{\boldsymbol{\Gamma}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1} \sum_{i=1}^N (\partial/\partial \boldsymbol{\theta}'_\tau) \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$ .

The following lemma is part (iii) of Lemma 1. Since its proof does not require the other parts of Lemma 1, it is presented here separately.

**Lemma A.4.** For  $\tau \geq 1$ , we have  $\sqrt{N} \hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0))$  as  $N \rightarrow \infty$ .

*Proof.* By Assumption 1,  $\{(y_{i0}, x_{it}, \lambda_i, \varepsilon_{it}), i = 1, \dots, N\}$  are i.i.d. Therefore,  $\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau^0)$  is also i.i.d for  $i = 1, \dots, N$ . It follows from the orthogonality condition that  $E[\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau^0)] = \mathbf{0}$ . Further, by Assumption 3 the variance-covariance matrix of  $\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau^0)$ ,  $\boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau^0)$ , is nonsingular. The statement of Lemma A.4 follows from the Multivariate Lindeberg-Levy CLT (see, e.g. page 1122 in Greene 2012).  $\square$

The next lemma provides some intermediate results that will be used in the proofs of the main results.

**Lemma A.5.**

1. There exists  $d(\mathbf{y}_i, \mathbf{X}_i)$  such that  $\|\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\|^2 \leq d(\mathbf{y}_i, \mathbf{X}_i)$  and  $E[d(\mathbf{y}_i, \mathbf{X}_i)] < \infty$ ; hence  $\|\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\| \leq \sqrt{d(\mathbf{y}_i, \mathbf{X}_i)}$  and  $E[\sqrt{d(\mathbf{y}_i, \mathbf{X}_i)}] < \infty$ .
2.  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\| \xrightarrow{P} 0$  and  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\hat{\boldsymbol{\Gamma}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\Gamma}_\tau(\boldsymbol{\theta}_\tau)\| \xrightarrow{P} 0$ .
3.  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\hat{\boldsymbol{\Phi}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau)\| \xrightarrow{P} 0$ , where  $\hat{\boldsymbol{\Phi}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1} \sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau) \boldsymbol{\mu}'_{\tau,i}(\boldsymbol{\theta}_\tau)$ .
4.  $\widehat{\mathbf{W}}$  is a consistent estimator of  $\boldsymbol{\Phi}_\tau^{-1}(\boldsymbol{\theta}_\tau)$ , where  $\widehat{\mathbf{W}} = \hat{\boldsymbol{\Phi}}_\tau^{-1}(\boldsymbol{\theta}_\tau)$  or  $\widehat{\mathbf{W}} = [\hat{\boldsymbol{\Phi}}_\tau(\boldsymbol{\theta}_\tau) + N^{-1} \mathbf{I}]^{-1}$ .
5. Let  $\hat{\mathbf{A}}$  and  $\mathbf{A}$  be positive definite matrices such that  $\hat{\mathbf{A}} \xrightarrow{P} \mathbf{A}$ . Let  $\hat{R}(\boldsymbol{\theta}_\tau) = \hat{\boldsymbol{\mu}}'_\tau(\boldsymbol{\theta}_\tau) \hat{\mathbf{A}} \hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau)$ ,  $R(\boldsymbol{\theta}_\tau) = \boldsymbol{\mu}'_\tau(\boldsymbol{\theta}_\tau) \mathbf{A} \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$ , and  $\tilde{\boldsymbol{\theta}}_\tau = \arg \min_{\boldsymbol{\theta}_\tau \in \Theta} \hat{R}(\boldsymbol{\theta}_\tau)$ . Then,  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\hat{R}(\boldsymbol{\theta}_\tau) - R(\boldsymbol{\theta}_\tau)\| \xrightarrow{P} 0$ , and  $\tilde{\boldsymbol{\theta}}_\tau \xrightarrow{P} \boldsymbol{\theta}_\tau^0$ .

*Proof.* 1. Since the proof invokes Lemmas A.2 and A.3, we shall first show that the conditions of these two lemmas are satisfied. For a positive finite constant  $c$ , by Assumptions 1 and 2, we have

$$\begin{aligned} \|\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\|^2 &= \|\mathbf{m}_i - \rho \mathbf{m}_{i,-1}^{(b)} - \eta_\tau \mathbf{m}_{i,-1}^{(a)} - \mathbf{M}_i^{(b)} \beta - \mathbf{M}_i^{(a)} \delta_\tau - \mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f}\|^2 \\ &\leq c \|\mathbf{m}_i\|^2 + c \|\mathbf{m}_{i,-1}^{(b)}\|^2 + c \|\mathbf{m}_{i,-1}^{(a)}\|^2 + c \|\mathbf{M}_i^{(b)}\|^2 + c \|\mathbf{M}_i^{(a)}\|^2 + c \|\mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f}\|^2 \\ &\equiv d(\mathbf{y}_i, \mathbf{X}_i). \end{aligned} \tag{23}$$

Let us verify that each term on the right hand side has finite expectation. First note that,

$$\|\mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f}\|^2 = \sum_{t=1}^T \sum_{s=0}^{t-1} g_s^2 f_t^2 \leq \sum_{t=1}^T f_t^2 \sum_{s=0}^T g_s^2.$$

By Assumption 1, there exists a finite  $b \in (0, \infty)$ , such that  $\|\mathbf{f}^0\| \leq b$  and by Assumption 2, the parameter space  $\Theta$  is compact. Further,  $g_s = E(y_{is} \lambda_i) \leq (E|y_{is}|^2 E|\lambda_i|^2)^{1/2} < b$  by the Cauchy-Schwartz inequality, Lemma A.1, and Assumption 1. Therefore, there exists a positive finite number, say  $d$ , such that  $E\|\mathbf{S}(\mathbf{I}_T \otimes \mathbf{G}) \mathbf{f}\|^2 < d$ . Moreover,

$$E\|\mathbf{m}_i\|^2 = E \left( \sum_{t=1}^T \sum_{s=0}^{t-1} y_{is}^2 y_{it}^2 \right) \leq \sum_{t=1}^T \sum_{s=0}^{t-1} E(y_{it}^2 y_{is}^2).$$

The last term is finite since  $E|y_{it}|^4 < \infty$  by Lemma A.1 and  $T$  is fixed. Similarly,  $E\|\mathbf{m}_{i,-1}^{(b)}\|^2 = E \left( \sum_{t=1}^{\tau-1} \sum_{s=0}^{t-1} y_{is}^2 y_{it}^2 \right)$  and  $E\|\mathbf{m}_{i,-1}^{(a)}\|^2 = E \left( \sum_{t=\tau}^T \sum_{s=0}^{t-1} y_{is}^2 y_{it}^2 \right)$  are also bounded by Assumption 1 and Lemma A.1. Furthermore,

$$E\|\mathbf{M}_i^{(b)}\|^2 = E \left( \sum_{t=1}^{\tau-1} \sum_{s=0}^{t-1} y_{is}^2 x_{it}^2 \right) = \sum_{t=1}^{\tau-1} \sum_{s=0}^{t-1} E(y_{is}^2 x_{it}^2) \leq \sum_{t=1}^{\tau-1} \sum_{s=0}^{t-1} [E(y_{is}^4) E(x_{it}^4)]^{1/2}.$$

The upper bound is finite since  $y_{is}$  and  $x_{it}$  have bounded fourth moments. Similarly,

$$E\|\mathbf{M}_i^{(a)}\|^2 = E\left(\sum_{t=\tau}^T \sum_{s=0}^{t-1} y_{is}^2 x_{it}^2\right) < \infty.$$

Thus, each term on the right hand side of (23) has finite expectation. Therefore, we have  $\|\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\|^2 \leq d(\mathbf{y}_i, \mathbf{X}_i)$ ,  $E[d(\mathbf{y}_i, \mathbf{X}_i)] < \infty$ ,  $\|\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\| \leq \sqrt{d(\mathbf{y}_i, \mathbf{X}_i)}$ , and  $E[\sqrt{d(\mathbf{y}_i, \mathbf{X}_i)}] < \infty$ . This completes the proof of part 1.

2. It follows from the previous part and Lemma A.2, that as  $N \rightarrow \infty$ , we have that

$$\sup_{\boldsymbol{\theta}_\tau \in \Theta} \left\| \frac{1}{N} \sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau) \right\| = \sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\| \xrightarrow{P} 0.$$

Note that  $\mathbf{\Gamma}_\tau(\boldsymbol{\theta}_\tau)$  is the Jacobian matrix of  $\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)$  and  $\hat{\mathbf{\Gamma}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1} \sum_{i=1}^N (\partial/\partial \boldsymbol{\theta}_\tau') \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$ . Since  $\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$  is quadratic in  $\boldsymbol{\theta}_\tau$ , the conditions of Lemma A.2 are also satisfied by  $\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)$ . Therefore, the required uniform convergence of  $\hat{\mathbf{\Gamma}}_\tau(\boldsymbol{\theta}_\tau)$  follows. This completes part 2.

3. Recall that  $\hat{\boldsymbol{\Phi}}_\tau(\boldsymbol{\theta}_\tau) = N^{-1} \sum_{i=1}^N \boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau) \boldsymbol{\mu}'_{\tau,i}(\boldsymbol{\theta}_\tau)$ . Since  $\|\boldsymbol{\mu}_{\tau,i}(\boldsymbol{\theta}_\tau)\| \leq \sqrt{d(\mathbf{y}_i, \mathbf{X}_i)}$  and  $E[\sqrt{d(\mathbf{y}_i, \mathbf{X}_i)}] < \infty$ , we have that  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\hat{\boldsymbol{\Phi}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\Phi}_\tau(\boldsymbol{\theta}_\tau)\|$  converges to zero in probability.

4. This follows from the continuous function theorem.

5. By the triangle and the Cauchy-Schwartz inequalities, we have that

$$\begin{aligned} \sup_{\boldsymbol{\theta}_\tau \in \Theta} \{|\hat{R}_\tau(\boldsymbol{\theta}_\tau) - R_\tau(\boldsymbol{\theta}_\tau)|\} &= \sup_{\boldsymbol{\theta}_\tau \in \Theta} \{|\hat{\boldsymbol{\mu}}'_\tau(\boldsymbol{\theta}_\tau) \hat{\mathbf{A}} \hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}'_\tau(\boldsymbol{\theta}_\tau) \mathbf{A} \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)|\} \\ &\leq \sup_{\boldsymbol{\theta}_\tau \in \Theta} \{ |(\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau))' \hat{\mathbf{A}} (\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau))| \\ &\quad + 2|\boldsymbol{\mu}'_\tau(\boldsymbol{\theta}_\tau) \hat{\mathbf{A}} (\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau))| + |\boldsymbol{\mu}'_\tau(\boldsymbol{\theta}_\tau) (\hat{\mathbf{A}} - \mathbf{A}) \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)| \} \\ &\leq \sup_{\boldsymbol{\theta}_\tau \in \Theta} \{ \|\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\|^2 \|\hat{\mathbf{A}}\| + 2 \sup_{\boldsymbol{\theta}_\tau \in \Theta} \{ \|\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\| \|\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\| \} \|\hat{\mathbf{A}}\| \\ &\quad + \sup_{\boldsymbol{\theta}_\tau \in \Theta} \{ \|\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\|^2 \} \|\hat{\mathbf{A}} - \mathbf{A}\| \} \xrightarrow{P} 0. \end{aligned}$$

where  $\xrightarrow{P} 0$  follows since  $\|\hat{\mathbf{A}}\| = O_p(1)$ ,  $\|\hat{\mathbf{A}} - \mathbf{A}\| = o_p(1)$ ,  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\|$  is finite, and  $\sup_{\boldsymbol{\theta}_\tau \in \Theta} \|\hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau) - \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau)\| = o_p(1)$ . Therefore,  $\hat{R}_\tau(\boldsymbol{\theta}_\tau)$  converges uniformly in probability to  $R_\tau(\boldsymbol{\theta}_\tau)$ . By Assumption 2,  $R_\tau(\boldsymbol{\theta}_\tau)$  is uniquely minimized at true value  $\boldsymbol{\theta}_\tau^0$ . Hence, for  $\tilde{\boldsymbol{\theta}}_\tau = \arg \min_{\boldsymbol{\theta}_\tau \in \Theta} \hat{R}_\tau(\boldsymbol{\theta}_\tau)$ ,  $\tilde{\boldsymbol{\theta}}_\tau \xrightarrow{P} \boldsymbol{\theta}_\tau^0$ , by Lemma A.3.  $\square$

**Proof of Lemma 1 .** Recall  $\widehat{\mathbf{W}} = \hat{\boldsymbol{\Phi}}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)})$  as the weight matrix in the second and final step. (i) By Lemma A.5 we have the first-step estimator  $\hat{\boldsymbol{\theta}}_\tau^{(1)} \xrightarrow{P} \boldsymbol{\theta}_\tau^0$  (see part 5),  $\widehat{\mathbf{W}} \xrightarrow{P} \boldsymbol{\Phi}_\tau^{-1}(\boldsymbol{\theta}_\tau^0)$  (see part 3), and  $\hat{\boldsymbol{\theta}}_\tau \xrightarrow{P} \boldsymbol{\theta}_\tau^0$  (see part 5), as  $N \rightarrow \infty$ . This completes the proof of first part of the lemma. Next, let us consider the second part. By Taylor expansion, we have

$$\sqrt{N} \hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau) = \sqrt{N} \hat{\boldsymbol{\mu}}_\tau(\boldsymbol{\theta}_\tau^0) + \hat{\mathbf{\Gamma}}_\tau(\bar{\boldsymbol{\theta}}_\tau) \sqrt{N} (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}_\tau^0), \quad (24)$$

where  $\bar{\theta}_\tau$  lies between  $\hat{\theta}_\tau$  and  $\theta_\tau^0$ . By substituting expression (24) into the first order condition  $\mathbf{0} = \sqrt{N}(\partial/\partial\theta'_\tau)\hat{Q}_\tau(\hat{\theta}_\tau) = 2\sqrt{N}\hat{\Gamma}'_\tau(\hat{\theta}_\tau)\widehat{\mathbf{W}}\hat{\mu}_\tau(\hat{\theta}_\tau)$  of the objective function  $\hat{Q}_\tau(\hat{\theta}_\tau)$ , we obtain

$$\hat{\Gamma}'_\tau(\hat{\theta}_\tau)\widehat{\mathbf{W}}\sqrt{N}\hat{\mu}_\tau(\theta_\tau^0) + \hat{\Gamma}'_\tau(\hat{\theta}_\tau)\widehat{\mathbf{W}}\hat{\Gamma}_\tau(\bar{\theta}_\tau)\sqrt{N}(\hat{\theta}_\tau - \theta_\tau^0) = \mathbf{0}. \quad (25)$$

Let  $\hat{\mathbf{C}} = \hat{\Gamma}'_\tau(\hat{\theta}_\tau)\widehat{\mathbf{W}}\hat{\Gamma}_\tau(\bar{\theta}_\tau)$ ,  $\mathbf{C} = \Gamma'_\tau(\theta_\tau^0)\Phi_\tau^{-1}(\theta_\tau^0)\Gamma_\tau(\theta_\tau^0)$ ,  $\hat{\mathbf{D}} = \hat{\Gamma}'_\tau(\hat{\theta}_\tau)\widehat{\mathbf{W}}$  and  $\mathbf{D} = \Gamma'_\tau(\theta_\tau^0)\Phi_\tau^{-1}(\theta_\tau^0)$ . Rearranging (25), we have

$$\hat{\mathbf{C}}[\sqrt{N}(\hat{\theta}_\tau - \theta_\tau^0)] = -\hat{\mathbf{D}}[\sqrt{N}\hat{\mu}_\tau(\theta_\tau^0)]. \quad (26)$$

Since  $\widehat{\mathbf{W}} \xrightarrow{p} \Phi_\tau^{-1}(\theta_\tau^0)$ , and  $\|\hat{\Gamma}_\tau(\hat{\theta}_\tau) - \Gamma_\tau(\theta_\tau^0)\| \xrightarrow{p} 0$ , as  $N \rightarrow \infty$ , we have  $\hat{\mathbf{C}} = \mathbf{C} + o_p(1)$  and  $\hat{\mathbf{D}} = \mathbf{D} + o_p(1)$ . Further since  $\mathbf{C}$  is nonsingular, applying Lemma 4.1 in Lehmann and Casella (1991, page 432), (26) leads to  $\sqrt{N}(\hat{\theta}_\tau - \theta_\tau^0) = -\mathbf{C}^{-1}\mathbf{D}\sqrt{N}\hat{\mu}_\tau(\theta_\tau^0) + o_p(1)$ . This completes the proof of part (ii) of the lemma.

Now, by Lemma A.4, we have that  $\sqrt{N}\hat{\mu}_\tau(\theta_\tau^0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Phi_\tau(\theta_\tau^0))$ , and

$$\sqrt{N}(\hat{\theta}_\tau - \theta_\tau^0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, [\Gamma_\tau(\theta_\tau^0)'\Phi_\tau^{-1}(\theta_\tau^0)\Gamma_\tau(\theta_\tau^0)]^{-1})$$

as  $N \rightarrow \infty$ . □

**Lemma A.6.** *Suppose that  $H_0$  holds. Let  $\tau \geq 2$ . Then  $\mathbf{M}_{\Phi_1^{-1/2}\Gamma_1} - \mathbf{M}_{\Phi_\tau^{-1/2}\Gamma_\tau}$  is a projection matrix with rank  $p = \dim(\eta_\tau) + \dim(\delta_\tau)$ .*

*Proof.* Write  $\Gamma_1 = \Gamma_1(\theta_1)$ ,  $\Phi_1 = \Phi_1(\theta_1)$ ,  $\Gamma_\tau = \Gamma_\tau(\theta_\tau)$  and  $\Phi_\tau = \Phi_\tau(\theta_\tau)$  for  $\tau \geq 2$ . Let  $(\gamma'_{11} | \gamma'_{12}) = [g_0, \dots, g_{T-1}, f_1, \dots, f_T | \rho, \beta] = \theta'_1$ , and  $(\gamma'_{\tau 1} | \gamma'_{\tau 2}) = [g_0, \dots, g_{T-1}, f_1, \dots, f_T | \rho, \beta, \eta_\tau, \delta_\tau] = \theta'_\tau$ . Since the null hypothesis holds, we have that  $\eta_\tau = \rho$  and  $\delta_\tau = \beta$ , and hence it follows from (4) and (6) that  $\mu_\tau(\theta_\tau) = \mu_1(\theta_1)$  and  $\Phi_\tau(\theta_\tau) = \Phi_1(\theta_1)$ . Recall that  $\Gamma_1$  and  $\Gamma_\tau$  are the Jacobian matrices of  $\mu_1(\theta_1)$  and  $\mu_\tau(\theta_\tau)$  respectively. Let us partition  $\Gamma_1$  as  $[\Gamma_{11} | \Gamma_{12}]$  to conform with the partition  $\theta'_1 = (\gamma'_{11} | \gamma'_{12})$ , where the left block  $\Gamma_{11}$  corresponds to the subvector  $(g_0, \dots, g_{T-1}, f_1, \dots, f_T)$  of  $\theta'_1$  which also appears in  $\theta'_\tau$ ;  $\Gamma_{12}$  corresponds to the subvector  $(\rho, \beta)$  of  $\theta'_1$ . Similarly,  $\Gamma_\tau$  is partitioned as  $[\Gamma_{\tau 1} | \Gamma_{\tau 2}]$  to conform with the partition  $\theta'_\tau = (\gamma'_{\tau 1} | \gamma'_{\tau 2})$ . Then, we have  $\Gamma_{\tau 1} = \Gamma_{11}$  since  $\gamma'_{\tau 1} = \gamma'_{11}$ , while  $\Gamma_{\tau 2}$  corresponds to the subvector  $(\rho, \beta, \eta_\tau, \delta_\tau)$  of  $\theta'_\tau$ . Further, partition  $\Gamma_{\tau 2}$  as  $[\Gamma_{\tau 2, (1)} | \Gamma_{\tau 2, (2)}]$  to conform with  $(\rho, \beta | \eta_\tau, \delta_\tau)$ . It follows that  $\Gamma_{12} = \Gamma_{\tau 2, (1)} + \Gamma_{\tau 2, (2)} = \Gamma_{\tau 2}[\mathbf{I}_q, \mathbf{I}_q]'$ , where  $\mathbf{I}_q$  is the identity matrix with  $q = \dim(\rho) + \dim(\beta) = 2$ .

For  $\tau \geq 2$ , by the notation introduced at the beginning of the Appendix, we have a projection matrix

$$\mathbf{M}_{\Phi_\tau^{-1/2}\Gamma_\tau} = \mathbf{I}_\zeta - \Phi_\tau^{-1/2}\Gamma_\tau(\Gamma'_\tau\Phi_\tau^{-1}\Gamma_\tau)^{-1}\Gamma'_\tau\Phi_\tau^{-1/2}.$$

Let  $\mathbf{E}_\tau = \Gamma'_{\tau 2}\Phi_\tau^{-1/2}\mathbf{M}_{\Phi_\tau^{-1/2}\Gamma_\tau}\Phi_\tau^{-1/2}\Gamma_{\tau 2}$ ,  $\mathbf{L}_\tau = \Phi_\tau^{-1/2}\Gamma_{\tau 2}\mathbf{E}_\tau^{-1}\Gamma'_{\tau 2}\Phi_\tau^{-1/2}$  and  $\mathbf{K}_{\Phi_\tau^{-1/2}\Gamma_\tau} = \Gamma'_{\tau 1}\Phi_\tau^{-1}\Gamma_{\tau 1}$ . By applying a formula for inverting a  $2 \times 2$  block matrix, we have

$$\begin{aligned}
(\Gamma'_\tau \Phi_\tau^{-1} \Gamma_\tau)^{-1} &= \begin{bmatrix} \Gamma'_{\tau 1} \Phi_\tau^{-1} \Gamma_{\tau 1}, & \Gamma'_{\tau 1} \Phi_\tau^{-1} \Gamma_{\tau 2} \\ \Gamma'_{\tau 2} \Phi_\tau^{-1} \Gamma_{\tau 1}, & \Gamma'_{\tau 2} \Phi_\tau^{-1} \Gamma_{\tau 2} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} K_{\Phi_\tau^{-1/2} \Gamma_{\tau 1}}^{-1} [I + \Gamma'_{\tau 1} \Phi_\tau^{-1} \Gamma_{\tau 2} E_\tau^{-1} \Gamma'_{\tau 2} \Phi_\tau^{-1} \Gamma_{\tau 1} K_{\Phi_\tau^{-1/2} \Gamma_{\tau 1}}^{-1}], & -K_{\Phi_\tau^{-1/2} \Gamma_{\tau 1}}^{-1} \Gamma'_{\tau 1} \Phi_\tau^{-1} \Gamma_{\tau 2} E_\tau^{-1} \\ -E_\tau^{-1} \Gamma'_{\tau 2} \Phi_\tau^{-1} \Gamma_{\tau 1} K_{\Phi_\tau^{-1/2} \Gamma_{\tau 1}}^{-1}, & E_\tau^{-1} \end{bmatrix}.
\end{aligned}$$

The projection matrix  $P_{\Phi_\tau^{-1/2} \Gamma_\tau}$  can be expressed as

$$\begin{aligned}
P_{\Phi_\tau^{-1/2} \Gamma_\tau} &= \Phi_\tau^{-1/2} \Gamma_\tau (\Gamma'_\tau \Phi_\tau^{-1} \Gamma_\tau)^{-1} \Gamma'_\tau \Phi_\tau^{-1/2} \\
&= \Phi_\tau^{-1/2} \Gamma_{\tau 1} K_{\Phi_\tau^{-1/2} \Gamma_{\tau 1}}^{-1} \Gamma'_{\tau 1} \Phi_\tau^{-1/2} + M_{\Phi_\tau^{-1/2} \Gamma_{\tau 1}} L_\tau M_{\Phi_\tau^{-1/2} \Gamma_{\tau 1}}.
\end{aligned}$$

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

$$\begin{aligned}
P_{\Phi_1^{-1/2} \Gamma_1} &= \Phi_1^{-1/2} \Gamma_1 (\Gamma'_1 \Phi_1^{-1} \Gamma_1)^{-1} \Gamma'_1 \Phi_1^{-1/2} \\
&= \Phi_1^{-1/2} \Gamma_{11} K_{\Phi_1^{-1/2} \Gamma_{11}}^{-1} \Gamma'_{11} \Phi_1^{-1/2} + M_{\Phi_1^{-1/2} \Gamma_{11}} L_1 M_{\Phi_1^{-1/2} \Gamma_{11}}.
\end{aligned}$$

Let

$$\begin{aligned}
J_1 &= M_{\Phi_1^{-1/2} \Gamma_{11}} \Phi_1^{-1/2} \Gamma_{\tau 2} (\Gamma'_{\tau 2} \Phi_1^{-1/2} M_{\Phi_1^{-1/2} \Gamma_{11}} \Phi_1^{-1/2} \Gamma_{\tau 2})^{-1} \Gamma'_{\tau 2} \Phi_1^{-1/2} M_{\Phi_1^{-1/2} \Gamma_{11}}, \\
J_2 &= M_{\Phi_1^{-1/2} \Gamma_{11}} \Phi_1^{-1/2} \Gamma_{12} (\Gamma'_{12} \Phi_1^{-1/2} M_{\Phi_1^{-1/2} \Gamma_{11}} \Phi_1^{-1/2} \Gamma_{12})^{-1} \Gamma'_{12} \Phi_1^{-1/2} M_{\Phi_1^{-1/2} \Gamma_{11}}.
\end{aligned}$$

Since  $\Phi_\tau = \Phi_1$  and  $\Gamma_{\tau 1} = \Gamma_{11}$ , under the null hypothesis, we have that  $V_\tau = M_{\Phi_1^{-1/2} \Gamma_1} - M_{\Phi_\tau^{-1/2} \Gamma_\tau} = J_1 - J_2$ . Recall that  $\Gamma_{12} = \Gamma_{\tau 2} [I_q, I_q]' = \Gamma_{\tau 2} B$ , where  $B$  is  $[I_q, I_q]'$  and  $q = \dim(\rho) + \dim(\beta)$ . It may be verified that  $J_1 J_2 = J_2$ ,  $J_1^2 = J_1$ , and  $J_2^2 = J_2$ . Combining these three equations, we have  $V_\tau^2 = V_\tau$ . Since  $V_\tau$  is also symmetric, it is a projection matrix of rank  $p = \text{rank}(M_{\Phi_1^{-1/2} \Gamma_1} - M_{\Phi_\tau^{-1/2} \Gamma_\tau}) = \text{tr}(P_{\Phi_\tau^{-1/2} \Gamma_\tau}) - \text{tr}(P_{\Phi_1^{-1/2} \Gamma_1}) = \dim(\eta_\tau) + \dim(\delta_\tau)$ . This completes the proof of Lemma A.6.  $\square$

**Proof of Theorem 1.** Let  $C$  and  $D$  be defined as in the proof of Lemma 1, namely  $C = \Gamma'_\tau(\theta_\tau^0) \Phi_\tau^{-1}(\theta_\tau^0) \Gamma_\tau(\theta_\tau^0)$ , and  $D = \Gamma'_\tau(\theta_\tau^0) \Phi_\tau^{-1}(\theta_\tau^0)$ . By part (ii) of Lemma 1 and the Mean Value Theorem applied to  $\sqrt{N} \hat{\mu}_\tau(\hat{\theta}_\tau)$ , we have, for  $\tau \geq 2$

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

and hence with  $\Gamma_\tau = \Gamma_\tau(\theta_\tau^0)$  and  $\Phi_\tau = \Phi_\tau(\theta_\tau^0)$  for simplicity

$$\begin{aligned}
\sqrt{N} \Phi_\tau^{-1/2} \hat{\mu}_\tau(\hat{\theta}_\tau) &= [I_\zeta - \Phi_\tau^{-1/2} \Gamma_\tau (\Gamma'_\tau \Phi_\tau^{-1} \Gamma_\tau)^{-1} \Gamma'_\tau \Phi_\tau^{-1/2}] \sqrt{N} \Phi_\tau^{-1/2} \hat{\mu}_\tau(\theta_\tau^0) + o_p(1) \\
&= M_{\Phi_\tau^{-1/2} \Gamma_\tau} \left( \sqrt{N} \Phi_\tau^{-1/2} \hat{\mu}_\tau(\theta_\tau^0) \right) + o_p(1). \tag{27}
\end{aligned}$$

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

$$\begin{aligned}\sqrt{N}\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) &= [\mathbf{I}_\zeta - \Phi_1^{-1/2}\Gamma_1(\Gamma_1'\Phi_1^{-1}\Gamma_1)^{-1}\Gamma_1'\Phi_1^{-1/2}]\sqrt{N}\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0) + o_p(1) \\ &= \mathbf{M}_{\Phi_1^{-1/2}\Gamma_1} \left( \sqrt{N}\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0) \right) + o_p(1),\end{aligned}\quad (28)$$

where  $\Gamma_1 = \Gamma_1(\boldsymbol{\theta}_1^0)$ ,  $\Phi_1 = \Phi_1(\boldsymbol{\theta}_1^0)$ .

Since the null hypothesis is assumed to hold, we also have,  $\eta_\tau = \rho$  and  $\delta_\tau = \beta$ , and hence  $\boldsymbol{\mu}_1(\boldsymbol{\theta}_1^0) = \boldsymbol{\mu}_\tau(\boldsymbol{\theta}_\tau^0)$  and  $\Phi_1(\boldsymbol{\theta}_1^0) = \Phi_\tau(\boldsymbol{\theta}_\tau^0)$ . The statistic for testing  $H_0$  against  $H_{(\tau)}$  is

$$\begin{aligned}\psi_\tau &= N \left[ \hat{Q}_1(\hat{\boldsymbol{\theta}}_1) - \hat{Q}_\tau(\hat{\boldsymbol{\theta}}_\tau) \right] = N \left[ \hat{\boldsymbol{\mu}}_1'(\hat{\boldsymbol{\theta}}_1)\hat{\Phi}_1^{-1}(\hat{\boldsymbol{\theta}}_1^{(1)})\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) - \hat{\boldsymbol{\mu}}_\tau'(\hat{\boldsymbol{\theta}}_\tau)\hat{\Phi}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)})\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau) \right] \\ &= N \left[ \hat{\boldsymbol{\mu}}_1'(\hat{\boldsymbol{\theta}}_1)\Phi_1^{-1}\hat{\boldsymbol{\mu}}_1(\hat{\boldsymbol{\theta}}_1) - \hat{\boldsymbol{\mu}}_\tau'(\hat{\boldsymbol{\theta}}_\tau)\Phi_\tau^{-1}\hat{\boldsymbol{\mu}}_\tau(\hat{\boldsymbol{\theta}}_\tau) \right] + o_p(1),\end{aligned}\quad (29)$$

as  $\hat{\Phi}_\tau^{-1}(\hat{\boldsymbol{\theta}}_\tau^{(1)}) \xrightarrow{p} \Phi_\tau^{-1}$  by parts 3 and 5 in Lemma A.5 for both  $\tau = 1$  and  $\tau \geq 2$ .

Substituting (27) and (28) into (29), we have

$$\psi_\tau = \sqrt{N}\hat{\boldsymbol{\mu}}_1'(\boldsymbol{\theta}_1^0)\Phi_1^{-1/2}\mathbf{V}_\tau\sqrt{N}\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0) + o_p(1),\quad (30)$$

where  $\mathbf{V}_\tau = \mathbf{M}_{\Phi_1^{-1/2}\Gamma_1} - \mathbf{M}_{\Phi_\tau^{-1/2}\Gamma_\tau}$ .

Let  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_\zeta)$ . Since  $\sqrt{N}\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0)$  converges to  $\mathbf{z}$ , in distribution by Lemma A.4, we have  $\psi_\tau \xrightarrow{d} \mathbf{z}'\mathbf{V}_\tau\mathbf{z}$ . Since  $\mathbf{V}_\tau$  is a projection matrix by Lemma A.6, it follows that  $\psi_\tau \xrightarrow{d} \chi_p^2$ , where  $p = \dim(\eta_\tau) + \dim(\delta_\tau)$ .  $\square$

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

$$\psi_{\tau_\ell} = \sqrt{N}\hat{\boldsymbol{\mu}}_1'(\boldsymbol{\theta}_1^0)\Phi_1^{-1/2}\mathbf{V}_{\tau_\ell}\sqrt{N}\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0) + o_p(1).$$

Therefore,

$$\begin{aligned}\boldsymbol{\psi} &= [\psi_{\tau_1}, \psi_{\tau_2}, \dots, \psi_{\tau_L}]' = \{\sqrt{N}[\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0)]'[\mathbf{V}_{\tau_1}, \dots, \mathbf{V}_{\tau_L}](\mathbf{I}_L \otimes \sqrt{N}\Phi_1^{-1/2}\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta}_1^0))\}' + o_p(1) \\ &\xrightarrow{d} \{\mathbf{z}'[\mathbf{V}_{\tau_1}, \dots, \mathbf{V}_{\tau_L}](\mathbf{I}_L \otimes \mathbf{z})\}'.\end{aligned}\quad (31)$$

The Distance type test statistic,  $\max_{l=1, \dots, L} \{\psi_{\tau_\ell}\}$ , is a continuous function of  $[\psi_{\tau_1}, \dots, \psi_{\tau_L}]'$ . Therefore, by the continuous mapping theorem, the asymptotic distribution of  $\max_{l=1, \dots, L} \{\psi_{\tau_\ell}\}$  is the distribution of the maximum of  $\{\mathbf{z}'[\mathbf{V}_{\tau_1}, \dots, \mathbf{V}_{\tau_L}](\mathbf{I}_L \otimes \mathbf{z})\}'$ .  $\square$