Non-standard Confidence Sets for Ratios and Tipping Points with Applications to Dynamic Panel Data

Jean-Thomas Bernard [*] Ba Chu[†] Lynda Khalaf [‡]

Marcel Voia[§]

Preliminary & Incomplete: Do Not Cite or Circulate

Abstract

We study estimation uncertainty when the object of interest contains one or more ratios of parameters. The ratio of parameters is a discontinuous parameter transformation; it has been shown that traditional confidence intervals often fail to cover this true ratio with very high probability. Constructing confidence sets for ratios using Fieller's method is a viable solution as the method can avoid the discontinuity problem. This paper proposes an extension of the multivariate Fieller method beyond standard estimators, focusing on asymptotically mixed normal estimators that commonly arise in dynamic panel polynomial regression with persistent covariates. We discuss the cases where the underlying estimators converge to various distribution, depending on the persistence level of the covariates. We show that the asymptotic distribution of the pivotal statistic used for constructing a Fieller's confidence set remains a standard Chi-squared distribution regardless of rates of convergence, thus the rates are being 'self-normalized' and can be unknown. A simulation study illustrates the finite sample properties of the proposed method in a dynamic polynomial panel. Our method is demonstrated to work well in small samples, even when the persistence coefficient is unity.

[*] Department of Economics, University of Ottawa. Mailing address: 2292 Edwin Crescent, Ottawa, Ontario K2C 1H7, Canada. TEL: 613-562-5800 ext. 1374; FAX: 613-562-5999; e-mail: jbernard3@uottawa.ca .

[†] Department of Economics, Carleton University, Carleton University. Mailing address: Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel (613) 520-2600-3546; FAX: (613)-520-3906. email: Ba.Chu@carleton.ca.

[‡] Department of Economics, Carleton University, and Centre de Recherche de l'Environnement, de l'Agroalimentaire, des Transports et de l'Énergie (CREATE), Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel (613) 520-2600-8697; FAX: (613)-520-3906. email: Lynda.Khalaf@carleton.ca.

[§] Department of Economics, Carleton University and Centre for Monetary and Financial Economics, Carleton University. Mailing address: Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel (613) 520-2600-3546; FAX: (613)-520-3906. email: Marcel.Voia@carleton.ca.

1. Introduction

Estimating or testing parameter ratios is an important issue in statistics and econometrics. From a theoretical perspective, inference problems arising from non-linearity with such transformations have attracted a great deal of interest; for references in statistics, see, for example, Zerbe, Laska, Meisner and Kushner (1982), Read (1983), Buonaccorsi (2001), Frantz (2007), and Ulrike and Franz (2009). From an empirical perspective, and more specifically in economics, ratios are parameters of interest in various applications involving elasticities or tipping points, for example with familiar "U" or inverted "U" shaped curves: Kuznet, Laffer, Rahn, Engel, Beveridge curves, as well as statistical Phillips, Yield and wage curves. In this paper, we focus on parameter ratios in dynamic Panel data models.

In general, there are two basic approaches to estimating and assessing ratios. The first one employs a Wald-type approach, and is known as the "Delta" method [as explained in Appendix D]. This method suits asymptotically normal panel data estimators, provided of course underlying regularity conditions prevail. However, it is becoming increasingly clear from the literature that Wald-type methods raise identification problems.[1]

Even when a ratio's numerator and denominator are well identified, the ratio is not well defined when its denominator approaches zero. Consequently, the distribution of standard test statistics becomes irregular, so usual tests and confidence intervals are incorrectly sized, or (said differently) usual asymptotic standard errors understate sampling uncertainty. So even if standard errors estimated using usual methods are narrow, they still provide a spurious assessment of the true uncertainty. In fact, Bolduc, Khalaf and Yelou (2010) document coverage rates collapsing to zero, that is, estimated intervals missing the unknown true value in all Monte Carlo replications, for empirically relevant scenarios.

The second approach – which may be traced back to Fieller (1954) – avoids this problem,

[1]On identification problems, their consequences and possible corrections, see Dufour (2003), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz and Nelson (1998), Dufour and Jasiak (2001), Kleibergen (2002, 2005), Stock,Wright and Yogo (2002), Moreira (2003), Dufour and Taamouti (2005, 2007), Andrews, Moreira and Stock (2006), Antoine and Lavergne (2012).

at least in principle, using a pivotal statistic as an alternative to a Wald-type one that requires identification. To the best of our knowledge, applications of the Fieller method with Panel data are scarce. Furthermore, a formal analysis of the method with persistent data is unavailable even in univariate contexts. Bernard, Idoudi, Khalaf and Yelou (2007) are a notable exception, as evidence supporting the Fieller method is provided in a univariate dynamic regression, even close to the unit root boundary. In the absence of supportive theory, this result motivates further work. In time series there is work that deals with such discontinuities: Phillips (Econometrica, 2014), Mikusheva (2007, 2012), Gorodnichenko, Mikusheva & Ng (2012). We thus revisit dynamic contexts including panel data, which as is well known, pose more serious challenges than univariate regressions. In particular for dynamic panels we extend the work of Pesaran, Shin and Smith (1999) and consider polynomial panels that span a wide range of applications; from persistence to discontinuous limiting distributions (e.g. unit root or the far-from-unity case).

As the stationarity property of polynomial regressors is often not modeled, or checked adequately the analysis of polynomial panels is interesting in its own right. We propose a parsimonious set of assumptions that preserves the stability restriction of long run equations as in Pesaran, Shin and Smith (1999) and prove that the MLE estimators converge to mixed normality at different rates. We effectively extend the multivariate Fieller method beyond standard estimators; and in the context of dynamic polynomial panels, we show that the asymptotic distribution of Fieller's statistic still remains a standard Chi-squared distribution, regardless of the convergence rates of estimates, thus the rates are being 'self-normalized' and can be unknown.

Finally, we conduct an extensive simulation study, driving persistence parameters close to boundaries, with various choices for N and T using a design based on well know empirical example, the case of an environmental Kuznet curve. Results reveal that the delta method cannot be salvaged in dynamic Panels. The Fieller method works well with GMM methods when persistence is controlled and with large N. Fieller's method based on our method method work very well, even with unit roots, and interestingly, even when N is large relative to T.

This paper is structured as follows. Section 2 presents a general Fieller's theorem for asymptotically mixed-normal estimators. Section 3 focuses on a dynamic panel regression with polynomial covariates, where a covariate could potentially contain a unit root. Section 4 studies the problem of constructing Fieller's confidence set for ratios of the parameters characterizing long-run relationship in an error-correction representation of a dynamic polynomial panel. Section 5 summarizes our simulation findings, and Section 6 concludes the paper. Proofs of main theorems and lemmas as well as other materials of technical flavour are collected in four appendices at the end of this paper.

Some commonly-used notations are in the order: X is used to represent a scalar, X, and $\boldsymbol{X}$ is used to represent a vector or a matrix, $\boldsymbol{X}$. Let's denote by C_0 a generic constant that may vary from one context to another. Given two random sequences, say a_T and b_T, one often writes $a_T \ll b_T$ a.s. if and only if $P(\lim_{T\uparrow\infty} |a_T/b_T| = \text{const.}) = 1$, and $a_T \ll b_T$ w.p. if and only if $\lim_{T\uparrow\infty} P(|a_T/b_T| < \text{const.}\xi) = 1$, where ξ can be some almost-sure bounded random variable; $o_p(\cdot)$ and $O_p(\cdot)$ are standard symbols for stochastic orders of magnitude. $\xrightarrow{p}$ denotes convergence in probability and $\xrightarrow{d}$ denotes convergence in distribution. $\|\cdot\|$ denotes the Euclidean norm of matrices and $\lambda_1(\boldsymbol{X})$ represents the minimum eigenvalue of a square matrix, $\boldsymbol{X}$. $\boldsymbol{I}_n$ stands for the identity matrix of size n. $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

2. Mixed-Normality based Fieller's Theorem

Consider a parametric model with parameters of interest defined by a vector, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$. Let $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^p$, where Θ is a compact parameter space, represent the true parameters; and for a given data sample of size T, one can estimate $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \dots, \widehat{\theta}_p)^\top$. We first make some assumptions about the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$. (Note that Assumptions 2.1 and 2.2 below are independent of each other, so are the notations.)

Assumption 2.1. *$\widehat{\boldsymbol{\theta}}$ is asymptotically normal as $T \uparrow \infty$, in the sense that $\boldsymbol{D}_T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_0(\boldsymbol{\theta}_0))$ uniformly over Θ, where $\boldsymbol{D}_T$ is a diagonal matrix consisting of normalizing factors that diverge to infinity with T and may differ from one another; and $N(0, \boldsymbol{\Sigma}_0(\boldsymbol{\theta}_0))$ represents a tight family of Gaussian random variables with the asymptotic variance-covariance matrix, $\boldsymbol{\Sigma}_0(\boldsymbol{\theta}_0)$, being the probability limit of some matrix of normalized sample statistics, $\boldsymbol{D}_T^{-1}\boldsymbol{\Sigma}_T\boldsymbol{D}_T^{-1}$.*

Assumption 2.2. *$\widehat{\boldsymbol{\theta}}$ is asymptotically mixed normal as $T \uparrow \infty$ such that*

(a) $\boldsymbol{D}_T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \widetilde{\boldsymbol{\Sigma}}_0^{-1/2}(\boldsymbol{\theta}_0) N(0, \boldsymbol{I}_p)$ *uniformly over Θ, where $\boldsymbol{D}_T$ is a diagonal matrix consisting of normalizing factors that diverge to infinity with T and may differ from one another; and $\widetilde{\boldsymbol{\Sigma}}_0^{-1/2}(\boldsymbol{\theta}_0) N(0, \boldsymbol{I}_p)$ represents a tight family of Gaussian random variables with $\widetilde{\boldsymbol{\Sigma}}_0^{-1} \equiv \widetilde{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{\theta}_0)$ being some random asymptotic variance-covariance matrix that is independent of $N(0, \boldsymbol{I}_p)$;*

(b) *$\widetilde{\boldsymbol{\Sigma}}_0$ is the probability limit of a matrix of normalized sample statistics, $\boldsymbol{D}_T^{-1} \widehat{\boldsymbol{\Sigma}}_T \boldsymbol{D}_T^{-1}$, such that $\boldsymbol{D}_T^{-1} \widehat{\boldsymbol{\Sigma}}_T \boldsymbol{D}_T^{-1} \xrightarrow{p} \widetilde{\boldsymbol{\Sigma}}_0$.*

Objects of our interest are the ratios $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^\top$ with $\rho_i = \frac{\boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\theta}}{\boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\theta}}$ for $i = 1, \dots, q \leq p-1$, where $\boldsymbol{L}_1, \dots, \boldsymbol{L}_q$, and $\boldsymbol{K}$ are $q+1$ nonstochastic and linearly independent $p \times 1$ vectors.

Theorem 1. *Let either Assumption 2.1 or 2.2 hold. Then the $1-\alpha$ asymptotic uniform simultaneous confidence sets, $CS_T(\rho; \alpha)$, for $\boldsymbol{\rho}_0$, defined via the inverse relationship $\lim_{T\uparrow\infty} \inf_{\boldsymbol{\theta}_0 \in \Theta} P_{\theta_0}(\rho \in CS_T(\rho; \alpha)) \geq 1 - \alpha$, can be obtained by inverting the following Wald-type test statistic for the null hypothesis $H_0:\quad \boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\theta}_0 - \rho_{0,i} \boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\theta}_0 = 0$ for all $i = 1, \dots, q$.*

$$\mathcal{W}(\boldsymbol{\rho}_0) = \widehat{\boldsymbol{\theta}}^\top \boldsymbol{D}_T (\boldsymbol{L} - \boldsymbol{K}_\rho) \left((\boldsymbol{L} - \boldsymbol{K}_\rho)^\top \left(\boldsymbol{D}_T^{-1} \widehat{\boldsymbol{\Sigma}}_T \boldsymbol{D}_T^{-1} \right)^{-1} (\boldsymbol{L} - \boldsymbol{K}_\rho) \right)^{-1} (\boldsymbol{L} - \boldsymbol{K}_\rho)^\top \boldsymbol{D}_T \widehat{\boldsymbol{\theta}} \xrightarrow{d} \chi^2(q).$$

If the distributional convergence is not uniform in either Assumption 2.1 or 2.2, then one can only construct the $1-\alpha$ asymptotic pointwise *simultaneous confidence sets, $CS_T(\rho; \alpha)$, for $\boldsymbol{\rho}_0$, defined via the inverse relationship $\lim_{T\uparrow\infty} P_{\theta_0}(\rho \in CS_T(\rho; \alpha)) \geq 1 - \alpha$ for every $\boldsymbol{\theta}_0 \in \Theta$.*

In particular, if the rates of convergence are the same or there is one ratio, then the matrices, $\boldsymbol{D}_T$, of normalizing factors in the above equation can be canceled out so that

$$\mathcal{W}(\boldsymbol{\rho}_0) = \widehat{\boldsymbol{\theta}}^\top (\boldsymbol{L} - \boldsymbol{K}_\rho) \left((\boldsymbol{L} - \boldsymbol{K}_\rho)^\top \widehat{\boldsymbol{\Sigma}}_T^{-1} (\boldsymbol{L} - \boldsymbol{K}_\rho) \right)^{-1} (\boldsymbol{L} - \boldsymbol{K}_\rho)^\top \widehat{\boldsymbol{\theta}} \xrightarrow{d} \chi^2(q).$$

Proof. By defining $p \times q$ matrices, $\boldsymbol{L} = (\boldsymbol{L}_1, \dots, \boldsymbol{L}_q)$ and $\boldsymbol{K}_\rho = \boldsymbol{K} \boldsymbol{\rho}_0^\top$, the null hypothesis H_0 can also be written as $(\boldsymbol{L}^\top - \boldsymbol{K}_\rho^\top) \boldsymbol{D}_T \boldsymbol{\theta}_0 = \boldsymbol{0}$. An application of the uniform continuous mapping theorem yields that

$$(\boldsymbol{L}^\top - \boldsymbol{K}_\rho^\top)\boldsymbol{D}_T\widehat{\boldsymbol{\theta}} \xrightarrow{d} \left((\boldsymbol{L} - \boldsymbol{K}_\rho)^\top \widetilde{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{L} - \boldsymbol{K}_\rho)\right)^{1/2} N(0, \boldsymbol{I}_p) \text{ under } H_0.$$

By replacing the unknown $\widetilde{\boldsymbol{\Sigma}}_0$ with $\boldsymbol{D}_T^{-1}\widehat{\boldsymbol{\Sigma}}_T\boldsymbol{D}_T^{-1}$, an application of Lemma 3 in Ogasawara and Takahishi (1951) yields the above Wald-type statistic. □

Remark 2.1. *One can then obtain the simultaneous confidence sets for* $\boldsymbol{\rho}_0$*:*

$$CS(\boldsymbol{\rho}; \alpha) = \{\boldsymbol{\rho} \in \mathbb{R}^q : \ \mathcal{W}(\boldsymbol{\rho}) \leq c_{q,\alpha}\},$$

where $c_{q,\alpha}$ *is the* $(1-\alpha)$ *critical value of the* χ^2 *distribution with* q *degrees of freedom. Let* $\underbrace{\boldsymbol{R}_\rho}_{q\times(q+1)} = (\boldsymbol{I}_q, \boldsymbol{\rho}_0)$, $\underbrace{\boldsymbol{H}}_{(q+1)\times p} = (\boldsymbol{L}, \boldsymbol{K})^\top$, *and* $\widehat{\boldsymbol{\Theta}} = \boldsymbol{D}_T\widehat{\boldsymbol{\theta}}$, *we shall rewrite the above Wald-type test statistic as*

$$\mathcal{W}(\boldsymbol{\rho}_0) = (\boldsymbol{R}_\rho\boldsymbol{H}\widehat{\boldsymbol{\Theta}})^\top \left(\boldsymbol{R}_\rho\boldsymbol{H}\left(\boldsymbol{D}_T^{-1}\widehat{\boldsymbol{\Sigma}}_T\boldsymbol{D}_T^{-1}\right)^{-1}\boldsymbol{H}^\top\boldsymbol{R}_\rho^\top\right)^{-1}(\boldsymbol{R}_\rho\boldsymbol{H}\widehat{\boldsymbol{\Theta}}).$$

Therefore, a closed-form expression for $CS(\boldsymbol{\rho}; \alpha)$ *can be derived by utilizing the same argument as in Section 4 of Bolduc, Khalaf and Yelou (2010).*

3. Dynamic Polynomial Panel Data Models

Given observations available in time periods, $t = 1, \ldots, T$, and groups, $i = 1, \ldots, N$, we aim to do statistical inferences based on the following polynomial $ARDL(p, q_w, q_z)$ model:

$$y_{i,t} = \sum_{j=1}^{p} \lambda_{i,j} y_{i,t-j} + \sum_{j=0}^{q_w} \boldsymbol{\gamma}_{i,j}^\top \boldsymbol{W}_{i,t-j} + \sum_{j=0}^{q_z} \boldsymbol{\Gamma}_{i,j}^\top \boldsymbol{Z}_{i,t-j} + \mu_i + \epsilon_{i,t}, \tag{3.1}$$

where $\boldsymbol{W}_{i,t} = (X_{i,t}, X_{i,t}^2, \ldots, X_{i,t}^{k_w})^\top$ and the $k_z \times 1$ vector $\boldsymbol{Z}_{i,t} = (Z_{1,i,t}, \ldots, Z_{k_z,i,t})^\top$ are vectors of explanatory variables; μ_i for $i = 1, \ldots, N$ represent the fixed effects; $\lambda_{i,j}$, $\boldsymbol{\gamma}_{i,j}$ and $\boldsymbol{\Gamma}_{i,j}$ are the coefficients of the lagged explanatory variables; p, q_w and q_z denote the number of lags; and $\epsilon_{i,t}$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ represent the errors [with mean zero and variance $\sigma^2_{0,i}$] which are independent across time periods and groups. Throughout this section, we shall make the following assumption:

Assumption 3.1. *The roots of the lag polynomial $\sum_{j=1}^{p} \lambda_{i,j} = 1$ lie outside the unit circle.*

Assumption 3.1 above warrants that $y_{i,t}$ is stable so that its dynamics is entirely determined by those of $\boldsymbol{W}_{i,t}$ and $\boldsymbol{Z}_{i,t}$. Using some backward differences, $\sum_{j=1}^{p-1} \lambda_{i,j+1} y_{i,t-j-1} = \sum_{j=1}^{p-1} \lambda^*_{i,j} \Delta y_{i,t-j} + \phi_i y_{i,t-1}$, where $\lambda^*_{i,j} = -\sum_{k=j+1}^{p} \lambda_{i,k}$ and $\phi_i = \sum_{k=2}^{p} \lambda_{i,k}$; $\sum_{j=0}^{q_w} \boldsymbol{\gamma}_{i,j}^\top \boldsymbol{W}_{i,t-j} = \sum_{j=0}^{q_w-1} {\boldsymbol{\gamma}^*_{i,j}}^\top \Delta \boldsymbol{W}_{i,t-j} + \boldsymbol{\beta}_i^\top \boldsymbol{W}_{i,t}$, where $\boldsymbol{\gamma}^*_{i,j} = -\sum_{k=j+1}^{q_w} \boldsymbol{\gamma}_{i,k}$ and $\boldsymbol{\beta}_i = \sum_{j=0}^{q_w} \boldsymbol{\gamma}_{i,j}$; and $\sum_{j=0}^{q_z} \boldsymbol{\Gamma}_{i,j}^\top \boldsymbol{Z}_{i,t-j} = \sum_{j=0}^{q_z-1} {\boldsymbol{\Gamma}^*_{i,j}}^\top \Delta \boldsymbol{Z}_{i,t-j} + \boldsymbol{\theta}_i^\top \boldsymbol{Z}_{i,t}$, where $\boldsymbol{\Gamma}^*_{i,j} = -\sum_{k=j+1}^{q_z} \boldsymbol{\Gamma}_{i,k}$ and $\boldsymbol{\theta}_i = \sum_{j=0}^{q_z} \boldsymbol{\Gamma}_{i,j}$; one can immediately write (3.1) in the following form of short-run (changes) dynamics:

$$\Delta y_{i,t} = \alpha_i \left(y_{i,t-1} + \frac{\boldsymbol{\beta}_i^\top}{\alpha_i} \boldsymbol{W}_{i,t} + \frac{\boldsymbol{\theta}_i^\top}{\alpha_i} \boldsymbol{Z}_{i,t} \right) + \sum_{j=1}^{p-1} \lambda^*_{i,j} \Delta y_{i,t-j} + \sum_{j=0}^{q_w-1} {\boldsymbol{\gamma}^*_{i,j}}^\top \Delta \boldsymbol{W}_{i,t-j} + \sum_{j=0}^{q_z-1} {\boldsymbol{\Gamma}^*_{i,j}}^\top \Delta \boldsymbol{Z}_{i,t-j} + \mu_i + \epsilon_{i,t}, \tag{3.2}$$

where $\alpha_i = \sum_{j=1}^{p} \lambda_{i,j} - 1$.

Remark 3.1. *Let $X_{i,t}$ and $\boldsymbol{Z}_{i,t}$ have unit roots. Then the first term on the RHS of (3.2) does not capture the adjustment towards an equilibrium state, thus there does not exist any long-run relationship in (3.1). To see this point, suppose that, on the contrary, there is a long-run relationship, then $y_{i,t-1} + \frac{\boldsymbol{\beta}_i^\top}{\alpha_i} \boldsymbol{W}_{i,t} + \frac{\boldsymbol{\theta}_i^\top}{\alpha_i} \boldsymbol{Z}_{i,t} = \eta_{i,t}$ with $\eta_{i,t}$ is a stationary process. Since $\Delta X^\ell_{i,t} = X^\ell_{i,t} - X^\ell_{i,t-1} = (X_{i,t} - X_{i,t-1})(X^{\ell-1}_{i,t} + X_{i,t-1} X^{\ell-2}_{i,t} + \cdots + X^{\ell-1}_{i,t-1})$, where ℓ is some integer greater than 1, the changes $\Delta \boldsymbol{W}_{i,t-j}$ for $j = 0, \ldots, q_w - 1$ still behave like nonstationary random chaos. In view of Assumption 3.1 the short-run dynamics of $y_{i,t}$ in (3.2) then remains unstable over time, thus no equilibrium state can be reached ultimately.*

Therefore, we shall now focus on the instantaneous relationship between $y_{i,t}$ and $\boldsymbol{W}_{i,t}$ in (3.1). Let the parameters of interest be denoted by $\boldsymbol{\phi} = \boldsymbol{\gamma}_{i,0}$, which are assumed to be invariant across the groups. Let $p^* = \max(p, q_w, q_z)$, define a $(T - p^*) \times 1$ vector, $\boldsymbol{y}_{i,-j} = (y_{i,p^*-j}, \ldots, y_{i,T-j})^\top$, a $(T - p^*) \times k_w$ matrix, $\boldsymbol{W}_{i,-j} = (\boldsymbol{W}_{i,p^*-j}, \ldots, \boldsymbol{W}_{i,T-j})^\top$, a $(T - p^*) \times k_z$ matrix, $\boldsymbol{Z}_{i,-j} = (\boldsymbol{Z}_{i,p^*-j}, \ldots, \boldsymbol{Z}_{i,T-j})^\top$, and a $(T-p^*)\times 1$ vector, $\boldsymbol{\epsilon}_i = (\epsilon_{i,p^*},\ldots,\epsilon_{i,T})^\top$. Moreover, let

$$\boldsymbol{X}_i = (\boldsymbol{y}_{i,-1},\ldots,\boldsymbol{y}_{i,-p},\boldsymbol{W}_{i,-1},\ldots,\boldsymbol{W}_{i,-q_w},\boldsymbol{Z}_i,\ldots,\boldsymbol{Z}_{i,-q_z},\boldsymbol{\iota}_T),$$

where $\boldsymbol{\iota}_T$ is the $(T-p^*)\times 1$ unit vector, and

$$\boldsymbol{\rho}_i = (\lambda_{i,1},\ldots,\lambda_{i,p},\boldsymbol{\gamma}_{i,1}^\top,\ldots,\boldsymbol{\gamma}_{i,q_w}^\top,\boldsymbol{\Gamma}_{i,0}^\top,\ldots,\boldsymbol{\Gamma}_{i,q_z}^\top,\mu_i)^\top.$$

One can then cast (3.1) into a matrix form,

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{\phi} + \boldsymbol{X}_i\boldsymbol{\rho}_i + \boldsymbol{\epsilon}_i.$$

Define $\boldsymbol{\varphi}_0 = (\boldsymbol{\phi}_0^\top,\boldsymbol{\sigma}_0^\top)^\top$ with $\boldsymbol{\sigma}_0 = (\sigma_{1,0}^2,\ldots,\sigma_{N,0}^2)^\top$ as the vector of the true parameters to be estimated. Suppose throughout that $\boldsymbol{\varphi}_0$ lies in the interior of some compact domain, $\Theta_\phi\times\Theta_\sigma\subset\mathbb{R}^{k_w}\times\mathbb{R}^N$. The concentrated log-likelihood function is given by

$$\ell_T(\boldsymbol{\varphi}) = -\frac{T}{2}\sum_{i=1}^N \ln(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=1}^N \frac{1}{\sigma_i^2}(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{\phi})^\top \boldsymbol{P}_i(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{\phi}),$$

where $\boldsymbol{P}_i = \boldsymbol{I}_T - \boldsymbol{X}_i(\boldsymbol{X}_i^\top\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^\top$ and $\boldsymbol{I}_T$ is the $(T-p^*)\times(T-p^*)$ identity matrix. The maximum likelihood estimate $\widehat{\boldsymbol{\varphi}}_T$ of $\boldsymbol{\varphi}_0$ is given by

$$\widehat{\boldsymbol{\varphi}}_T = \text{argmax}_{\boldsymbol{\varphi}\in\Theta_\phi\times\Theta_\sigma}\ell_T(\boldsymbol{\varphi}).$$

By the orthogonality between $\boldsymbol{P}_i$ and $\boldsymbol{X}_i$, one can also obtain that

$$\ell_T(\boldsymbol{\varphi}_0) = -\frac{T}{2}\sum_{i=1}^N \ln(2\pi\sigma_{0,i}^2) - \frac{1}{2}\sum_{i=1}^N \frac{1}{\sigma_{0,i}^2}\boldsymbol{\epsilon}_i^\top\boldsymbol{P}_i\boldsymbol{\epsilon}_i.$$

Some algebra manipulations yields

$$\begin{aligned}
\frac{1}{T}\{\ell_T(\boldsymbol{\varphi}_0)-\ell_T(\boldsymbol{\varphi})\} &= \frac{1}{2}\sum_{i=1}^N \frac{\sigma_{0,i}^2-\sigma_i^2}{\sigma_i^2\sigma_{0,i}^2}\left(\frac{\boldsymbol{\epsilon}_i^\top \boldsymbol{P}_i\boldsymbol{\epsilon}_i}{T}-\sigma_{0,i}^2\right)+\frac{1}{2}\sum_{i=1}^N\left(\frac{\sigma_{0,i}^2}{\sigma_i^2}-\ln\left(\frac{\sigma_{0,i}^2}{\sigma_i^2}\right)-1\right) \\
&+ \frac{1}{2}\frac{1}{T}\sum_{i=1}^N\frac{1}{\sigma_i^2}\left((\boldsymbol{y}_i-\boldsymbol{W}_i\boldsymbol{\phi})^\top\boldsymbol{P}_i(\boldsymbol{y}_i-\boldsymbol{W}_i\boldsymbol{\phi})-\boldsymbol{\epsilon}_i^\top\boldsymbol{P}_i\boldsymbol{\epsilon}_i\right) \\
&= \frac{1}{2}(\mathcal{T}_{1,T}+\mathcal{T}_{2,T}+\mathcal{T}_{3,T}), \qquad (3.3)
\end{aligned}$$

where $\mathcal{T}_{3,T}=(\boldsymbol{\phi}-\boldsymbol{\phi}_0)^\top\left(\sum_{i=1}^N\frac{1}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top\boldsymbol{P}_i\boldsymbol{W}_i}{T}\right)(\boldsymbol{\phi}-\boldsymbol{\phi}_0)-2(\boldsymbol{\phi}-\boldsymbol{\phi}_0)^\top\left(\sum_{i=1}^N\frac{1}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top\boldsymbol{P}_i\boldsymbol{\epsilon}_i}{T}\right)$. For asymptotic theory, some further assumptions are required.

Assumption 3.2. *The innovations $\boldsymbol{\epsilon}_i$ are orthogonal to both $\boldsymbol{W}_i$ and $\boldsymbol{X}_i$. In addition, given i, $X_{i,t}$ is $I(1)$ and can be represented as $X_{i,t}=\sum_{s=1}^t\zeta_{i,s}$ for some zero-mean innovations, $\{\zeta_{i,t},\ i=1,\dots,N,\ t=1,\dots,T\}$, independent across the groups and stationary, strongly mixing across the time periods with the mixing coefficient satisfying the condition stated in Lemma 3. The same assumptions about $X_{i,t}$ are also imposed on $\boldsymbol{Z}_{i,t}=\sum_{s=1}^t\boldsymbol{\xi}_{i,s}$.*

Lemma 1. *Let Assumptions 3.1 and 3.2 hold. Throughout this section (Section 2), we denote by*

$$\boldsymbol{D}_T = diag\left(T,\dots,T^{\frac{\ell+1}{2}},\dots,T^{\frac{k_w+1}{2}}\right)$$

the diagonal matrix of normalizing factors. We then have

$$\boldsymbol{D}_T^{-1}\boldsymbol{W}_i^\top\boldsymbol{P}_i\boldsymbol{W}_i\boldsymbol{D}_T^{-1} \xrightarrow{p} \boldsymbol{Q}_{ww,i}^*, \qquad (3.4)$$

$$\boldsymbol{D}_T^{-1}\boldsymbol{W}_i^\top\boldsymbol{P}_i\boldsymbol{\epsilon}_i \xrightarrow{p} \boldsymbol{Q}_{w\epsilon,i}^*, \qquad (3.5)$$

where $\boldsymbol{Q}_{ww,i}^*$ and $\boldsymbol{Q}_{w\epsilon,i}^*$ are some random matrices.

Theorem 2 (Consistency)**.** *Suppose that Assumptions 3.1 and 3.2 hold and $E[|\epsilon_{i,t}|^{2+\delta}]<\infty$ for* *some $\delta > 0$. In addition, let*

$$\inf_{\boldsymbol{\sigma}\in\Theta_\sigma} \lambda_1 \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \boldsymbol{Q}_{ww,i}^{*} \right) > C_{ww} \ \ a.s.$$

for some $C_{ww} > 0$. Then, the elements of $\widehat{\varphi}_T$ are consistent, i.e., $T^{\frac{\ell}{2}}(\widehat{\phi}_\ell - \phi_{0,\ell}) = o_p(1)$ for $\ell = 1, \ldots, k_w$ and $\widehat{\sigma}_\ell - \sigma_{0,\ell} = o_p(1)$ for $\ell = 1, \ldots, N$.

Theorem 3 (Asymptotic Mixed Normality)**.** *Suppose that Assumptions 3.1 and 3.2 and the condition in Theorem 2 hold. Moreover, let the innovations satisfies $E[|\epsilon_{i,t}|^{2+\delta}] < \infty$ for some $\delta > 0$. Then,*

$$\boldsymbol{D}_T(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \xrightarrow{d} MN\left(0, \left(\boldsymbol{Q}_{ww,N}^{*}\right)^{-1}\right),$$

where MN stands for the mixed normal distribution and $\boldsymbol{Q}_{ww,N}^{*} = \sum_{i=1}^{N} \frac{1}{\sigma_{0,i}^2} \boldsymbol{Q}_{ww,i}^{*}$.

Corollary 4. *Fieller's confidence sets for the ratios $\delta_i = \frac{\boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\phi}_0}{\boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\phi}_0}$ for $i = 1, \ldots, m$, where $\boldsymbol{L}_i$ and $\boldsymbol{K}$ are some given column vectors, can be constructed by inverting a Wald's statistics for testing $H_0: \ \boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\phi}_0 - \delta_i \boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\phi}_0 = 0$ vs. $H_1: \ \boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\phi}_0 - \delta_i \boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\phi}_0 \neq 0$. Theorem 3 suggests that this Wald's statistics is given by*

$$\begin{aligned} &\mathcal{W}_T(\boldsymbol{\delta}) \\ &= \widehat{\boldsymbol{\phi}}^\top \boldsymbol{D}_T(\boldsymbol{L} - \boldsymbol{\delta}\boldsymbol{K}_\delta) \left((\boldsymbol{L}^\top - \boldsymbol{\delta}\boldsymbol{K}_\delta^\top) \left(\sum_{i=1}^{N} \frac{1}{\widehat{\sigma}_i^2} \boldsymbol{D}_T^{-1} \boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i \boldsymbol{D}_T^{-1} \right)^{-1} (\boldsymbol{L} - \boldsymbol{\delta}\boldsymbol{K}_\delta) \right)^{-1} (\boldsymbol{L}^\top - \boldsymbol{\delta}\boldsymbol{K}_\delta^\top) \boldsymbol{D}_T \widehat{\boldsymbol{\phi}} \\ &\xrightarrow{d} \chi^2(m). \end{aligned}$$

where $\boldsymbol{\delta} = diag(\delta_1, \ldots, \delta_m)$; and $\boldsymbol{L} = (\boldsymbol{L}_1, \ldots, \boldsymbol{L}_m)$ and $\boldsymbol{K}_\delta = \boldsymbol{\iota}_m^\top \otimes \boldsymbol{K}$ are $k_w \times m$ matrices.

4. Dynamic Panel Polynomial Error Correction Models

Suppose that we have observations of some random variables, $y_{i,t}$, $X_{i,t}$ and $\boldsymbol{Z}_{i,t}$, across time periods, $t = 1, \ldots, T$, and groups, $i = 1, \ldots, N$. Let the observations be generated from the following error correction model:

$$\Delta y_{i,t} = \phi_i(y_{i,t-1} - \boldsymbol{\theta}^\top \boldsymbol{W}_{i,t}) + \sum_{j=1}^{p-1} \lambda_{i,j}\Delta y_{i,t-j} + \sum_{j=0}^{q_x-1} \gamma_{i,j}\Delta X_{i,t-j} + \sum_{j=0}^{q_z-1} \boldsymbol{\alpha}_{i,j}^\top \Delta \boldsymbol{Z}_{i,t-j} + \mu_i + \epsilon_{i,t}, \quad (4.1)$$

where $\boldsymbol{W}_{i,t} = (\boldsymbol{Z}_{i,t}^\top, X_{i,t}, X_{i,t}^2, \ldots, X_{i,t}^{k_x})^\top$, with $\boldsymbol{Z}_{i,t}$ being of dimension $k_z \times 1$, represent vectors of explanatory variables satisfying Assumption 3.2. As in the above section, μ_i and $\epsilon_{i,t}$ are the fixed effects and the random errors respectively; $\lambda_{i,j}$, $\gamma_{i,j}$, and $\boldsymbol{\alpha}_{i,j}$ denote the coefficients of the lagged explanatory variables; and $\boldsymbol{\theta}$ represents the regression coefficients. The following assumption allows (4.1) to have a long-run relationship, $y_{i,t} = \boldsymbol{\theta}^\top \boldsymbol{W}_{i,t} + \nu_{i,t}$, where $\nu_{i,t}$ is a stationary process.

Assumption 4.1. *The process $y_{i,t}$ has a unit root for each i, and the lag polynomial $\sum_{j=1}^{p-1} \lambda_{i,j} z^j = 1$ has roots outside the unit circle.*

We then focus on the long-run relationship between $y_{i,t}$ and $\boldsymbol{W}_{i,t}$ in (4.1). Let's denote by $\boldsymbol{\varphi} = (\boldsymbol{\theta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\sigma})^\top$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)^\top$ and $\boldsymbol{\sigma} = (\sigma_1^2, \ldots, \sigma_N^2)^\top$, the parameters of interest, which are assumed to lie in the interior of some parameter spaces. We shall assume throughout this section that all the parameter spaces are compact, and the log-likelihood maximization is carried out on these compact spaces. Let $p^* = \max(p, q_x, q_z)$, define $(T - p^*) \times 1$ vectors, $\Delta \boldsymbol{y}_{i,-j} = (\Delta y_{i,p^*-j}, \ldots, \Delta y_{i,T-j})^\top$ and $\Delta \boldsymbol{X}_{i,-j} = (\Delta X_{i,p^*-j}, \ldots, \Delta X_{i,T-j})^\top$, a $(T - p^*) \times k_z$ matrix, $\Delta \boldsymbol{Z}_{i,-j} = (\Delta \boldsymbol{Z}_{i,p^*-j}, \ldots, \Delta \boldsymbol{Z}_{i,T-j})^\top$, and a $(T-p^*)\times 1$ vector of random errors, $\boldsymbol{\epsilon}_i = (\epsilon_{i,p^*}, \ldots, \epsilon_{i,T})^\top$. Moreover, define

$$\boldsymbol{U}_i = (\Delta \boldsymbol{y}_{i,-1}, \ldots, \Delta \boldsymbol{y}_{i,-p+1}, \Delta \boldsymbol{X}_i, \ldots, \Delta \boldsymbol{X}_{i,-q_x+1}, \Delta \boldsymbol{Z}_i, \ldots, \Delta \boldsymbol{Z}_{i,-q_z+1}, \boldsymbol{\iota}_T),$$

where $\boldsymbol{\iota}_T$ is the $(T - p^*) \times 1$ unit vector, and

$$\boldsymbol{\Lambda}_i = (\lambda_{i,1}, \ldots, \lambda_{i,p-1}, \gamma_{i,0}, \ldots, \gamma_{i,q_x-1}, \boldsymbol{\alpha}_{i,0}^\top, \ldots, \boldsymbol{\alpha}_{i,q_z-1}^\top, \mu_i)^\top.$$

Note at this point that $\boldsymbol{U}_i$ is of dimension $(T - p^*) \times k_u$ with $k_u = p + q_x + q_z k_z$, and $\boldsymbol{\Lambda}_i$ is of dimension $k_u \times 1$. One can now cast (4.1) into the following matrix form:

$$\boldsymbol{y}_i = \phi_i(\boldsymbol{y}_{i,-1} - \boldsymbol{W}_i\boldsymbol{\theta}) + \boldsymbol{U}_i\boldsymbol{\Lambda}_i + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{W}_i = (\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{X}_i^2, \ldots, \boldsymbol{X}_i^{k_x})$ with $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i,1}, \ldots, \boldsymbol{Z}_{i,T-p^*})^\top$ and $\boldsymbol{X}_i^\ell = (X_{i,1}^\ell, \ldots, X_{i,T-p^*}^\ell)^\top$ for $\ell = 1, \ldots, k_x$. The log-likelihood function is given by

$$\ell_T(\boldsymbol{\varphi}) = -\frac{T}{2}\sum_{i=1}^N \ln(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=1}^N \frac{1}{\sigma_i^2}(\Delta\boldsymbol{y}_i - \phi_i\boldsymbol{\xi}_i(\boldsymbol{\theta}))^\top \boldsymbol{P}_i(\Delta\boldsymbol{y}_i - \phi_i\boldsymbol{\xi}_i(\boldsymbol{\theta})),$$

where $\boldsymbol{\xi}_i(\boldsymbol{\theta}) = \boldsymbol{y}_{i,-1} - \boldsymbol{W}_i\boldsymbol{\theta}$ and $\boldsymbol{P}_i = \boldsymbol{I}_T - \boldsymbol{U}_i(\boldsymbol{U}_i^\top\boldsymbol{U}_i)^{-1}\boldsymbol{U}_i^\top$ with $\boldsymbol{I}_T$ being the $(T-p^*)\times(T-p^*)$ identity matrix. Note the orthogonality of $\boldsymbol{P}_i$ to $\boldsymbol{U}_i$, one obtains that, for the true parameter $\boldsymbol{\varphi}_0$,

$$\begin{aligned}\frac{1}{T}(\ell_T(\boldsymbol{\varphi}_0) - \ell_T(\boldsymbol{\varphi})) &= \frac{1}{2}\sum_{i=1}^N \frac{\sigma_{0,i}^2 - \sigma_i^2}{\sigma_i^2\sigma_{0,i}^2}\left(\frac{\boldsymbol{\epsilon}^\top\boldsymbol{P}_i\boldsymbol{\epsilon}}{T} - \sigma_{0,i}^2\right) + \frac{1}{2}\sum_{i=1}^N\left(\frac{\sigma_{0,i}^2}{\sigma_i^2} - \ln\left(\frac{\sigma_{0,i}^2}{\sigma_i^2}\right) - 1\right)\\ &\quad + \frac{1}{2}\frac{1}{T}\sum_{i=1}^N \frac{1}{\sigma_i^2}\left((\Delta\boldsymbol{y}_i - \phi_i\boldsymbol{\xi}_i(\boldsymbol{\theta}))^\top\boldsymbol{P}_i(\Delta\boldsymbol{y}_i - \phi_i\boldsymbol{\xi}_i(\boldsymbol{\theta})) - \boldsymbol{\epsilon}_i^\top\boldsymbol{P}_i\boldsymbol{\epsilon}_i\right)\\ &= \frac{1}{2}\left(\mathcal{T}_{1,T}(\boldsymbol{\sigma},\boldsymbol{\sigma}_0) + \mathcal{T}_{2,T}(\boldsymbol{\sigma},\boldsymbol{\sigma}_0) + \mathcal{T}_{3,T}(\boldsymbol{\varphi})\right). \quad (4.2)\end{aligned}$$

Since $\Delta\boldsymbol{y}_i - \phi_i\boldsymbol{\xi}_i(\boldsymbol{\theta}) = \boldsymbol{U}_i\boldsymbol{\Lambda}_{0,i} + \boldsymbol{\epsilon}_i + \phi_i\boldsymbol{W}_i(\boldsymbol{\theta}_0 - \boldsymbol{\theta}) + (\phi_i - \phi_{0,i})\boldsymbol{\xi}_i(\boldsymbol{\theta}_0)$, one can also write

$$\mathcal{T}_{3,T}(\boldsymbol{\varphi}) = (\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^\top\boldsymbol{G}_T(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0) + 2(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^\top\boldsymbol{F}_T,$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\theta}^\top, \boldsymbol{\phi}^\top)^\top$ and $\boldsymbol{G}_T = \boldsymbol{G}_T(\boldsymbol{\phi},\boldsymbol{\sigma}) = \begin{pmatrix} \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top\boldsymbol{P}_i\boldsymbol{W}_i}{T} & -\frac{\phi_1}{\sigma_1^2}\frac{\boldsymbol{W}_1^\top\boldsymbol{P}_1\boldsymbol{\xi}_{0,1}}{T} & \cdots & -\frac{\phi_N}{\sigma_N^2}\frac{\boldsymbol{W}_N^\top\boldsymbol{P}_N\boldsymbol{\xi}_{0,N}}{T} \\ -\frac{\phi_1}{\sigma_1^2}\frac{\boldsymbol{W}_1^\top\boldsymbol{P}_1\boldsymbol{\xi}_{0,1}}{T} & \frac{1}{\sigma_1^2}\frac{\boldsymbol{\xi}_{0,1}^\top\boldsymbol{P}_1\boldsymbol{\xi}_{0,1}}{T} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \boldsymbol{0} & \ddots & \boldsymbol{0} \\ -\frac{\phi_N}{\sigma_N^2}\frac{\boldsymbol{W}_N^\top\boldsymbol{P}_N\boldsymbol{\xi}_{0,N}}{T} & \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\sigma_N^2}\frac{\boldsymbol{\xi}_{0,N}^\top\boldsymbol{P}_N\boldsymbol{\xi}_{0,N}}{T} \end{pmatrix}$,

where $\boldsymbol{\xi}_{0,i} = \boldsymbol{\xi}_i(\boldsymbol{\theta}_0)$ for $i = 1, \ldots, N$, and $\boldsymbol{F}_T = \begin{pmatrix} \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top\boldsymbol{P}_i\boldsymbol{\epsilon}_i}{T} \\ -\frac{1}{\sigma_1^2}\frac{\boldsymbol{\xi}_{0,1}^\top\boldsymbol{P}_1\boldsymbol{\epsilon}_1}{T} \\ \vdots \\ -\frac{1}{\sigma_N^2}\frac{\boldsymbol{\xi}_{0,N}^\top\boldsymbol{P}_N\boldsymbol{\epsilon}_N}{T} \end{pmatrix}$. To work out the probability limits for the random matrices $\boldsymbol{G}_T$ and $\boldsymbol{F}_T$, we need to state the following lemma:

Lemma 2. *Suppose that Assumptions 3.2 and 4.1 hold. Let's denote by*

$$\boldsymbol{D}_{ww,T} = diag\left(T^{1/2}\boldsymbol{\iota}_{k_z}^\top, T^{1/2}, \ldots, T^{\ell/2}, \ldots, T^{k_x/2}\right),$$

where $\boldsymbol{\iota}_{k_z}$ is the $k_z \times 1$ unit vector, the diagonal matrix of normalizing factors. Then,

$$\boldsymbol{D}_{ww,T}^{-1}\frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i}{T}\boldsymbol{D}_{ww,T}^{-1} \xrightarrow{p} \boldsymbol{Q}_{ww,i}, \tag{4.3}$$

$$\boldsymbol{D}_{ww,T}^{-1}\frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{\xi}_{0,i}}{T} \xrightarrow{p} \boldsymbol{Q}_{w\xi,i}, \tag{4.4}$$

$$\frac{\boldsymbol{\xi}_{0,i}^\top \boldsymbol{P}_i \boldsymbol{\xi}_{0,i}}{T} \xrightarrow{p} \boldsymbol{Q}_{\xi\xi,i}, \tag{4.5}$$

where $\boldsymbol{Q}_{ww,i}$, $\boldsymbol{Q}_{w\xi,i}$, and $\boldsymbol{Q}_{\xi\xi,i}$ are some random matrices.

Theorem 5 (Consistency)**.** *Suppose that Assumptions 3.1-4.1 hold. Let $\boldsymbol{D}_{G,T} = diag(\boldsymbol{D}_{ww,T}, \boldsymbol{I}_N)$, where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and*

$$\boldsymbol{Q}_G = \boldsymbol{Q}_G(\boldsymbol{\phi}, \boldsymbol{\sigma}) = \begin{pmatrix} \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2}\boldsymbol{Q}_{ww,i} & -\frac{\phi_1}{\sigma_1^2}\boldsymbol{Q}_{w\xi,1} & \cdots & -\frac{\phi_N}{\sigma_N^2}\boldsymbol{Q}_{w\xi,N} \\ -\frac{\phi_1}{\sigma_1^2}\boldsymbol{Q}_{w\xi,1} & \frac{1}{\sigma_1^2}\boldsymbol{Q}_{\xi\xi,1} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \boldsymbol{0} & \ddots & \boldsymbol{0} \\ -\frac{\phi_N}{\sigma_N^2}\boldsymbol{Q}_{w\xi,N} & \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\sigma_N^2}\boldsymbol{Q}_{\xi\xi,N} \end{pmatrix}.$$

Moreover, suppose that $E[|\epsilon_{i,t}|^{2+\delta}] < \infty$ for some $\delta > 0$, and $\inf_{\boldsymbol{\phi},\boldsymbol{\sigma}} \lambda_1(\boldsymbol{Q}_G) > 0$ a.s. Then,

$$\boldsymbol{D}_{ww,T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = o_p(1), \;\; (\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = o_p(1), \;\; and \; (\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0) = o_p(1).$$

Theorem 6 (Asymptotic Mixed Normality)**.** *Suppose that Assumptions 3.1-4.1 hold. Moreover, presume that $\lambda_1(\boldsymbol{Q}_G(\boldsymbol{\phi}_0, \boldsymbol{\sigma}_0)) > 0$ a.s. and $E[|\epsilon_{i,t}|^{2+\delta}] < \infty$ for some $\delta > 0$. Define*

$$\boldsymbol{D}_T = diag\left(T^{1/2}\boldsymbol{D}_{ww,T}, T^{1/2}\boldsymbol{I}_N\right) = diag\left(T\boldsymbol{\iota}_{k_z}^\top, T, \ldots, T^{\frac{\ell+1}{2}}, \ldots, T^{\frac{k_x+1}{2}}, T^{\frac{1}{2}}\boldsymbol{\iota}_N^\top\right).$$

Then,

$$\boldsymbol{D}_T(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0) \xrightarrow{d} MN\left(\boldsymbol{0}, \boldsymbol{Q}_G^{-1}(\boldsymbol{\phi}_0, \boldsymbol{\sigma}_0)\right).$$

Corollary 7. *Fieller's confidence sets for the ratios* $\delta_i = \frac{\boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\Gamma}_0}{\boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\Gamma}_0}$ *for* $i = 1, \ldots, m$, *where* $\boldsymbol{L}_i$ *and* $\boldsymbol{K}$ *are some given column vectors, can be constructed by inverting a Wald's statistics for testing* $H_0: \ \boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\Gamma}_0 - \delta_i \boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\Gamma}_0 = 0$ *vs.* $H_1: \ \boldsymbol{L}_i^\top \boldsymbol{D}_T \boldsymbol{\Gamma}_0 - \delta_i \boldsymbol{K}^\top \boldsymbol{D}_T \boldsymbol{\Gamma}_0 \neq 0$. *Theorem 6 suggests that this Wald's statistics is given by*

$$\mathcal{W}_T(\boldsymbol{\delta}) = \widehat{\boldsymbol{\Gamma}}^\top \boldsymbol{D}_T(\boldsymbol{L} - \boldsymbol{\delta}\boldsymbol{K}_\delta)\left((\boldsymbol{L}^\top - \boldsymbol{\delta}\boldsymbol{K}_\delta^\top)\left(\boldsymbol{D}_T^{-1}\widehat{\boldsymbol{G}}_T\boldsymbol{D}_T^{-1}\right)^{-1}(\boldsymbol{L} - \boldsymbol{\delta}\boldsymbol{K}_\delta)\right)^{-1}(\boldsymbol{L}^\top - \boldsymbol{\delta}\boldsymbol{K}_\delta^\top)\boldsymbol{D}_T\widehat{\boldsymbol{\Gamma}} \xrightarrow{d} \chi^2(m),$$

where $\widehat{\boldsymbol{G}}_T = \begin{pmatrix} \sum_{i=1}^N \frac{\widehat{\phi}_i^2}{\widehat{\sigma}_i^2}\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i & -\frac{\widehat{\phi}_1}{\widehat{\sigma}_1^2}\boldsymbol{W}_1^\top \boldsymbol{P}_1 \widehat{\boldsymbol{\xi}}_1 & \cdots & -\frac{\widehat{\phi}_N}{\widehat{\sigma}_N^2}\boldsymbol{W}_N^\top \boldsymbol{P}_N \widehat{\boldsymbol{\xi}}_N \\ -\frac{\widehat{\phi}_1}{\widehat{\sigma}_1^2}\boldsymbol{W}_1^\top \boldsymbol{P}_1 \widehat{\boldsymbol{\xi}}_1 & \frac{1}{\widehat{\sigma}_1^2}\widehat{\boldsymbol{\xi}}_1^\top \boldsymbol{P}_1 \widehat{\boldsymbol{\xi}}_1 & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \boldsymbol{0} & \ddots & \boldsymbol{0} \\ -\frac{\widehat{\phi}_N}{\widehat{\sigma}_N^2}\boldsymbol{W}_N^\top \boldsymbol{P}_N \widehat{\boldsymbol{\xi}}_N & \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\widehat{\sigma}_N^2}\widehat{\boldsymbol{\xi}}_N^\top \boldsymbol{P}_N \widehat{\boldsymbol{\xi}}_N \end{pmatrix}$ *with* $\widehat{\boldsymbol{\xi}}_i = \boldsymbol{y}_{i,-1} - \boldsymbol{W}_i\widehat{\boldsymbol{\theta}}$ *for* $i = 1, \ldots, N$; $\boldsymbol{D}_T^{-1}\widehat{\boldsymbol{G}}_T\boldsymbol{D}_T^{-1}$ *is the estimate of* $\boldsymbol{Q}_G(\boldsymbol{\phi}_0, \boldsymbol{\sigma}_0)$; $\boldsymbol{\delta} = diag(\delta_1, \ldots, \delta_m)$; $\boldsymbol{L} = (\boldsymbol{L}_1, \ldots, \boldsymbol{L}_m)$ *and* $\boldsymbol{K}_\delta = \boldsymbol{\iota}_m^\top \otimes \boldsymbol{K}$ *are* $k_{wn} \times m$ *matrices.*

5. Monte-Carlo Results

This section contains a Monte Carlo simulation to demonstrate the finite-sample performance of the proposed method. Specifically, we calculate the size and power of the Fieller-based test involving a ratio of two estimated parameters. To be precise the Monte-Carlo design is based on the following transformed dynamic polynomial panel:

$$\Delta y_{i,t} = \tilde{\phi}_i(y_{i,t-1} - \boxed{\theta^\top} W_{i,t}) + \sum_{j=1}^{p-1} \lambda_{i,j}\Delta y_{i,t-j} + \sum_{j=0}^{q_x-1} \gamma_{i,j}\Delta X_{i,t-j} + \sum_{j=0}^{q_z-1} \tilde{\alpha}_{i,j}^\top \Delta Z_{i,t-j} + \mu_i + \epsilon_{i,t},$$

$$t = 1, \ldots, T, \qquad i = 1, \ldots, N$$

$$W_{i,t} = (\underbrace{Z_{i,t}^\top}_{k_z \times 1}, \boxed{X_{i,t}}, \qquad \underbrace{X_{i,t}^2, \ldots, X_{i,t}^{k_x}}_{polynomial})^\top$$

θ invariant across i: stability

$y_{i,t}$, $X_{i,t}$ & var. in $Z_{i,t}$ have unit roots

long-run relation stationary

We consider that data generating process (DGP) is a finite-order $ARDL(1,0)$ process as in Pesaran and Shin (1999):

The above model including ECM, quadratic polynomial, no further lags and no $Z_{i,t-j}$

$$\tilde{\phi}_i = \phi$$

$$X_{it} - \psi X_{i,t-1} = \rho(X_{it} - \psi X_{i,t-1}) + \eta_{it}$$

where the errors (ϵ_{it}, η_{it}) are serially correlated and are generated according to the following bivariate normal distribution:

$$\begin{pmatrix} \epsilon \\ \mu \end{pmatrix} \rightarrow N(0, \Omega)$$

with

$$\Omega = \begin{pmatrix} 1 & \omega_{12} \\ \omega_{12} & 1 \end{pmatrix}$$

The parameters θ comprise of β_0 (constant),β_1 (of X_{it}),β_2 (of X_{it}^2) and the covariance ω_{12} were obtained from a real data exercise done by Khalaf et al. (2011), where an empirical estimation and inference of the Environmental Kuznets Curve (EKC) for carbon dioxide and sulfur were proposed. The y_{it} in our simulations were obtained using the data on annual per capita CO_2 emission and X_{it} was measuring per capita income, in 1000s of 2000 USD. The parameters of the DGP were obtained by employing a Dynamic Panel Polynomial Error Correction Model with fixed effects (we use a DFE abbreviation as in the Figures presented in the Appendix) on CO_2 data. The above example was considered to conduct our simulations because the original data was highly persistent and β_2 was weakly identified. In the simulations we keep $\beta_0, \beta_1, \beta_2$ and ω_{12} fixed and we play with the degree of persistency for the y_{it} and X_{it} by changing the parameters ϕ, ρ and ψ. In Appendix D we present a detailed description on how to construct confidence sets for ratio of two parameters using Delta and Fieller methods.

We use different levels of persistence for both y_{it} and X_{it} starting from low persistence for both y_{it} and X_{it} to non-stationarity of y_{it} and high persistence of X_{it}. The parameters considered in the simulations are listed in the Table 1 (see Appendix).

The results of the simulation study show how poorly the Delta Method works compared to the Fieller method when we test the existence of a ratio of two parameters. In particular we show that in presence of persistent outcome variables, combining the DFE method that estimates the parameters of the model with the Fieller method used to test the existence of a ratio of two parameters, outperform any other combination of estimation and testing considered in the simulation exercise. As an alternative case we consider the Arrellano-Bond (AB) estimator that is widely used for fixed T dynamic panels. We report size and power of the test underlying both Fieller and delta-method for all the cases from Table 1, however for the exposition in the paper we present few relevant cases (all the other cases are available in a separate appendix). The results show that the combination of DFE-Fieller achieves the correct level even in finite samples, while DFE-Delta fails for any sample size. Interesting, for this combination of parameters, both the combination of AB - Delta and AB -Fieller achieves the correct level for micro panels (large N, not highly persistent data), however the combination AB-Fieller is much stable for different sample sizes than AB-Delta. From Figure 1 we can conclude that the combination of DFE-Fieller outperforms all other combinations for any sample sizes.

Figure 2 completes the picture of the performance of the combination DFE-Fieller by showing how powerful this combination is when compared to any other combination. The results also show that AB - Delta is more powerful than AB - Fieller, but much less powerful than DFE-Fieller or DFE-Delta.

In all the other cases of this Monte-Carlo study, we observe a similar behaviour for both size and power (see Figures: 3,4,5,6 for example). Therefore, we find that DFE-Fieller proposed method works in all cases where data can be highly persistent [with nonstationary covariates] while the other methods such as DFE-Delta, AB-Fieller and AB-Delta do not.

6. Conclusion

When ratios of parameters are estimated and tested, it is important to obtain reliable confidence bounds especiallywhen one deals with longitudinal and possible nonstationary data.

As theoretical contributions, we prove that the MLE estimators for persistent dynamic panel data models converge to mixed normality at different rates, we extend the multivariate Fieller method beyond standard estimators and apply it to ratios of parameters obtained in dynamic polynomial panels and we show that the asymptotic distribution of Fieller's statistic remains a standard Chi-squared distribution regardless of the convergence rates of estimates.

A comprehensive Monte Carlo exercise suggest that highly persistent data require adequate estimation methods coupled with appropriate testing procedures. Using a long-run estimation approach based on Pesaran, Shin and Smith (1999) holds promise - in the sense that it provides reliable estimates for curvatures with nonstationary data. In addition, To answer the question whether data supports a plausible tipping point, statistical methods that account for a weakly identified tipping point should be preferred. Consequently, combining the appropriate estimation method with Fieller method to construct confidence sets for ratios of parameters of interest provides a powerful tool to a researcher because the constructed confidence sets remain valid with both persistent and less persistent data.

References

Arellano M. and S. Bond (1991). *Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.* Review of Economic Studies 58, 277-297.

Beaulieu M.-C., Dufour J.-M. and L. Khalaf (2011). *Identification-Robust Estimation and Testing of the Zero-Beta CAPM*, revised and resubmitted to: The Review of Economic Studies.

Bernard J.-T., Idoudi N., Khalaf L. and C. Yélou (2007). *Finite Sample Inference Methods for Dynamic Energy Demand Models*, Journal of Applied Econometrics 22, 1211-1226.

Baltagi B. 1995. Econometric Analysis of Panel Data (Chichester: Wiley).

Blundell R. and S. Bond (1998). *Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.* Journal of Econometrics 87, 115-143.

Bolduc D., Khalaf L. and C. Yelou (2010). *Identification Robust Confidence Set Methods for Inference on Parameter Ratios with Applications to Discrete Choice Models.* Journal of Econometrics 157, 317–327.

Bruno G. S. F. (2005). Estimation and Inference in Dynamic Unbalanced Panel Data Models with a Small Number of Individuals. The Stata Journal 5, 473-500.

Buonaccorsi, J. P. (2001). *Fieller's theorem.* In A. H. El-Shaarawi

Dufour J.-M. (1997). *Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models.* Econometrica 65, 1365-1389.

Dufour J.-M. (2003). *Identification, Weak Instruments and Statistical Inference in Econometrics.* Canadian Journal of Economics 36, 767-808.

Fieller E. C. (1940). *The Biological Standardization of Insulin.* Journal of the Royal Statistical Society (Supplement) 7, 1-64.

Fieller E. C. (1954). *Some Problems in Interval Estimation.* Journal of the Royal Statistical Society B 16, 175-185.

Franz, V. H. (2007). *Ratios: A short guide to confidence limits and proper use.* (Preprint available at http://arxiv.org/abs/0710.2024)

Khalaf, L., Bernard, Jean-Thomas, Gavin, M.l,and Voia, M.C. (2011). *The Environmental Kuznets Curve: Tipping Points, Uncertainty and Weak Identification* . CREATE Working Paper No. 2011-4. Available at SSRN: http://ssrn.com/abstract=1974622 or http://dx.doi.org/10.2139/ssrn.1974622

Kiviet J. F. (1995). *On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models.* Journal of Econometrics 68: 53-78.

Nickell, S. J. 1981. Biases in Dynamic Models with Fixed Effects. Econometrica 49: 1417-1426.

Pesaran, M. H. and Smith R. 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics, Volume 68, Issue 1, July 1995, Pages 79-113.

Pesaran, M. H.and Shin, Y.1999, "An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis," in Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium, chapter 11, (ed.) S. Strom, Cambridge University Press, Cambridge.

Pesaran, M.H., Shin, Y., Smith, R.P. 1999. Pooled mean group estimation of dynamic heterogeneous panels. Journal of the American Statistical Association, 94, 621 - 634.

Read, C. B. (1983). "Fieller's theorem." In Encyclopedia of Statistical Sciences (pp. 8688). Wiley, New York.

Stock J. H. (2010). The Other Transformation in Econometric Practice: Robust Tools for Inference. Journal of Economic Perspectives 24, 83-94.

Stock J. H., J. H. Wright and M. Yogo (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. Journal of Business and Economic Statistics 20, 518-529.

Ulrike von L. and Franz, V. H. (2009). A Geometric Approach to Confidence Sets dor Ratios: Fieller's Theorem, Generalizations and Bootstrap. Statistica Sinica 19, 1095-1117.

Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics, 126(1), 25-51.

Zerbe, G.O., Laska, E., Meisner, M. and Kushner, A.B. (1982). On Multivariate Confidence Regions and Simultaneous Confidence Limits for Ratios. Communications in Statistics, Theory and Methods 11, 2401 - 2425.

Andrews, D. W. K (2000), 'Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space', *Econometrica* 68, 399-405.

Andrews, D. W. K., Moreira, M. J. and Stock, J. H. (2006), 'Optimal two-sided invariant similar tests for instrumental variables regression', *Econometric*a 74, 715–752.

Antoine B. and P. Lavergne (2012). 'Conditional moment models under weak identification', Working paper, Simon Fraser University.

Beaulieu M.-C., Dufour J.-M. and L. Khalaf (2012). 'Testing portfolio efficiency with an unobservable zero-beta rate and non-Gaussian distributions: an exact identification-robust approach', *Review of Economic Studies*, forthcoming.

Bolduc, D., Khalaf, L. and Yelou, C. (2010), 'Identification robust confidence sets methods for inference on parameter ratios with application to discrete choice models', *Journal of Econometrics,* 157, 317-327.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* 65, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* 36(4), 767–808.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* 73, 1351–1365.

Dufour, J.-M. and Taamouti, M. (2007), 'Further results on projection-based inference in IV regressions with weak, collinear or missing instruments', *Journal of Econometrics* 139, 133–153.

Kleibergen, F. (2002), ‘Pivotal statistics for testing structural parameters in instrumental variables regression’, *Econometrica* 70, 1781–1803.

Kleibergen, F. (2005), ‘Testing parameters in gmm without assuming that they are identificated’, *Econometrica* 73, 1103–1123.

Moreira, M. J. (2003), ‘A conditional likelihood ratio test for structural models’, *Econometrica* 71(4), 1027–1048.

Staiger, D. and Stock, J. H. (1997), ‘Instrumental variables regression with weak instruments’, *Econometrica* 65(3), 557–586.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), ‘A survey of weak instruments and weak identification in generalized method of moments’, J*ournal of Business and Economic Statistics* 20(4), 518–529.

Wang, J. and Zivot, E. (1998), ‘Inference on structural parameters in instrumental variable regression with weak instruments’, *Econometrica* 66, 1389–1404.

Wright, J. H. (2000), ’Confidence set for cointegrating coefficients based on stationarity tests’, *Jorunal of Business and Economic Statistics* 18, 211-222.

Zivot, E., Startz, R. and Nelson, C. R. (1998), ‘Valid confidence intervals and inference in the presence of weak instruments’, *International Economic Review* 39, 1119–1144.

Appendix A. Known Results

The following lemma contains an almost sure invariance principle for sums of mixing random vectors.

Lemma 3. *Let $\{\xi_n,\ n \geq 1\}$ be a weak sense stationary sequence of $\mathbb{R}^d$-valued random vectors, centered at expectations and having $(2+\delta)$-th moments with $0 < \delta \leq 1$, uniformly bounded by 1; and let $\mathcal{F}_a^b$ represent the σ-field generated by the random vectors $\xi_a, \xi_{a+1}, \ldots, \xi_b$. Suppose that $\{\xi_n,\ n \geq 1\}$ satisfies the following strong-mixing condition:*

$$|P(AB) - P(A)P(B)| \leq \alpha(n)$$

for all $n,\ k \geq 1$, all $A \in \mathcal{F}_1^k$, and $B \in \mathcal{F}_{k+n}^{\infty}$ such that $\alpha(n) = C_0 n^{-(1+\epsilon)(1+2/\delta)}$ for some $\epsilon > 0$. Write $\xi_n = (\xi_{n,1}, \ldots, \xi_{n,d})$. Then the two series in $\gamma_{i,j} = E[\xi_{1,i}\xi_{1,j}] + \sum_{k\geq 2} E[\xi_{1,i}\xi k, j] + \sum_{k \geq 2} E[\xi_{k,i}\xi_{1,j}]$ converge absolutely. Denote the matrix $(\gamma_{i,j},\ 1 \leq i,\ j \leq d)$ by Γ. Then, we can redefine the sequence $\{\xi_n,\ n \geq 1\}$ on a new probability space together with Brownian motion $W(t)$ with covariance matrix Γ such that

$$\sum_{n \leq t} \xi_n - W(t) \ll t^{1/2-\lambda} \ a.s.$$

for some $\lambda > 0$ depending on ϵ, δ, and d only.

Proof. See Theorem 4 in Kuelbs and Philipp (1980). □

Appendix B. Proofs of Auxiliary Lemmas

Proof of Lemma 1. We only need to show (3.4) as (3.5) can be obtained in the same way. First, define $\boldsymbol{X}_i^{*m} = (X_{i,1}^m, \ldots, X_{i,T}^m)^\top$. By applying Lemma 3 and the formula $\sum_{k=1}^n k^p \approx \frac{n^{p+1}}{p+1}$ for every

$p > -1$, one can immediately show that

$$\begin{aligned}
\boldsymbol{y}_i^\top \boldsymbol{y}_i &\ll \sum_{t=1}^{T} X_{i,t}^{2k_w} \ll \sum_{t=1}^{T} t^{k_w} \approx \frac{T^{k_w+1}}{k_w+1} \text{ w.p.,} \\
(\boldsymbol{X}_i^{*m})^\top \boldsymbol{X}_i^{*\ell} &\ll \sum_{t=1}^{T} t^{\frac{m+\ell}{2}} \approx \frac{T^{1+\frac{m+\ell}{2}}}{1+\frac{m+\ell}{2}} \text{ w.p.,} \\
\boldsymbol{Z}_i^\top \boldsymbol{Z}_i &\ll \frac{T^2}{2} \text{ w.p.}
\end{aligned}$$

It then follows that

$$\boldsymbol{D}_{xx,T}^{-1}(\boldsymbol{X}_i^\top \boldsymbol{X}_i)\boldsymbol{D}_{xx,T}^{-1} \xrightarrow{p} \boldsymbol{Q}_{xx}, \tag{B-1}$$

where $\boldsymbol{Q}_{xx}$ is a random matrix, $\boldsymbol{D}_{xx,T} = \text{diag}\left(T^{\frac{k_w+1}{2}}\boldsymbol{\iota}_p^\top, \boldsymbol{I}_{q_w} \otimes \boldsymbol{D}_{x,T}, \boldsymbol{I}_{q_z} \otimes \boldsymbol{D}_z, T^{1/2}\right)$ with $\boldsymbol{D}_{x,T} = \text{diag}(T^1, T^{3/2}, \ldots, T^{\frac{\ell+1}{2}}, \ldots, T^{\frac{k_w+1}{2}})$, $\boldsymbol{D}_z = T\boldsymbol{I}_{k_z}$, $\boldsymbol{\iota}_p$ is the $p \times 1$ unit vector, and $\boldsymbol{I}_q$ represents the $q \times q$ identity matrix. Moreover, by the same argument, one also has

$$\begin{aligned}
(\boldsymbol{X}_i^{*1})^\top \boldsymbol{y}_i &\ll \sum_{t=1}^{T} X_{i,t}^{k_w+1} \ll \sum_{t=1}^{T} t^{\frac{k_w+1}{2}} \approx \frac{T^{\frac{k_w+3}{2}}}{\frac{k_w+3}{2}} \text{ w.p.,} \\
(\boldsymbol{X}_i^{*1})^\top \boldsymbol{W}_i &\ll (T^2, \ldots, T^{1+\frac{\ell+1}{2}}, \ldots, T^{\frac{k_w+3}{2}}) \text{ w.p.,} \\
(\boldsymbol{X}_i^{*1})^\top \boldsymbol{Z}_i &\ll T^2 \boldsymbol{\iota}_{k_z}^\top \text{ w.p.,} \\
(\boldsymbol{X}_i^{*\ell})^\top \boldsymbol{y}_i &\ll \sum_{t=1}^{T} X_{i,t}^{\ell+k_w} \ll \sum_{t=1}^{T} t^{\frac{k_w+\ell}{2}} \approx \frac{T^{\frac{k_w+\ell+2}{2}}}{\frac{k_w+\ell+2}{2}} \text{ w.p.,} \\
(\boldsymbol{X}_i^{*\ell})^\top \boldsymbol{W}_i &\ll \left(T^{\frac{\ell+3}{2}}, \ldots, T^{\frac{\ell+k_w+2}{2}}\right) \text{ w.p.,} \\
(\boldsymbol{X}_i^{*\ell})^\top \boldsymbol{Z}_i &\ll T^{\frac{\ell+3}{2}} \boldsymbol{\iota}_{k_z}^\top \text{ w.p.}
\end{aligned}$$

Therefore, by collecting all the relevant convergence rates, one can end up with

$$\boldsymbol{W}_i^\top \boldsymbol{X}_i \boldsymbol{D}_{xx,T}^{-1} \ll \left(T, \ldots, T^{\frac{\ell+1}{2}}, \ldots, T^{\frac{k_w+1}{2}}\right) \otimes \boldsymbol{\iota}_{k_x}^\top \text{ w.p.,} \tag{B-2}$$

where $k_x = p + k_w q_w + k_z(q_z+1) + 1$. It then follows from (B-1) and (B-2) that

$$\boldsymbol{D}_T^{-1}\boldsymbol{W}_i^\top \boldsymbol{X}_i(\boldsymbol{X}_i^\top \boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^\top \boldsymbol{W}_i\boldsymbol{D}_T^{-1} \xrightarrow{p} \boldsymbol{Q}_{1,ww},$$

where $\boldsymbol{Q}_{1,ww}$ denotes some random matrix. Using the same arguments as above, one can also prove that

$$\boldsymbol{D}_T^{-1}\boldsymbol{W}_i^\top \boldsymbol{W}_i\boldsymbol{D}_T^{-1} \xrightarrow{p} \boldsymbol{Q}_{2,ww},$$

where $\boldsymbol{Q}_{2,ww}$ represents some random matrix. Therefore, we obtain (3.4). □

Proof of Lemma 2. First, note that, in view of Assumptions 3.2 and 4.1, an application of Lemma 3 yields

$$\begin{aligned}
\Delta\boldsymbol{y}_i^\top \Delta\boldsymbol{y}_i &= \sum_{t=1}^T \Delta y_{i,t}^2 \ll T \text{ w.p.},\\
\boldsymbol{U}_i^\top \boldsymbol{U}_i &\ll T\boldsymbol{\iota}_{k_u}\boldsymbol{\iota}_{k_u}^\top \text{ w.p., where } \boldsymbol{\iota}_{k_u} \text{ is the } k_u \times 1 \text{ unit vector},\\
\Delta\boldsymbol{y}_i^\top \boldsymbol{Z}_i &= \sum_{t=1}^T \boldsymbol{Z}_{i,t}\Delta y_{i,t} \approx \int_0^1 \boldsymbol{Z}_{i,\lfloor T\tau\rfloor}(y_{i,\lfloor T(\tau+d\tau)\rfloor} - y_{i,\lfloor T\tau\rfloor}) \ll T\boldsymbol{\iota}_{k_z} \text{ w.p.}
\end{aligned}$$

as $y_{i,\lfloor T\tau\rfloor} - y_{i,\lfloor T(\tau+d\tau)\rfloor}$ can be approximated by a Brownian motion, $dW(\lfloor T\tau\rfloor) = W(\lfloor T(\tau+d\tau)\rfloor) - W(\lfloor T\tau\rfloor)$. And by the same argument, one also obtains

$$\begin{aligned}
\Delta\boldsymbol{X}_i^\top \boldsymbol{Z}_i &\ll T\boldsymbol{\iota}_{k_z} \text{ w.p.},\\
\Delta\boldsymbol{Z}_i^\top \boldsymbol{Z}_i &\ll T\boldsymbol{\iota}_{k_z}\boldsymbol{\iota}_{k_z}^\top \text{ w.p.},\\
\boldsymbol{Z}_i^\top \boldsymbol{i} &= \sum_{t=1}^T \boldsymbol{Z}_{i,t} \ll T^{3/2}\boldsymbol{\iota}_{k_z} \text{ w.p.},\\
(\boldsymbol{X}_i^\ell)^\top \Delta\boldsymbol{y}_i &= \sum_{t=1}^T X_{i,t}^\ell \Delta y_{i,t} \approx \int_0^1 X_{i,\lfloor T\tau\rfloor}^\ell (y_{i,\lfloor T(\tau+d\tau)\rfloor} - y_{i,\lfloor T\tau\rfloor}) \ll T^{\frac{\ell+1}{2}} \text{ w.p.},\\
(\boldsymbol{X}_i^\ell)^\top \boldsymbol{\iota}_T &\ll T^{\frac{\ell+2}{2}} \text{ w.p.},\\
(\boldsymbol{X}_i^\ell)^\top \boldsymbol{U}_i &\ll \left(T^{\frac{\ell+1}{2}}\boldsymbol{\iota}_{k_u-1}^\top, T^{\frac{\ell+2}{2}}\right) \text{ w.p.}
\end{aligned}$$

Collecting all the above-derived rates of divergence, one can immediately show that

$$\boldsymbol{Z}_i^\top \boldsymbol{U}_i \ll \left(T\boldsymbol{\iota}_{k_z\times(k_u-1)}, T^{3/2}\boldsymbol{\iota}_{k_z}\right) \text{ w.p.},$$

where $\boldsymbol{\iota}_{k_z\times(k_u-1)}$ represents the $k_z \times (k_u-1)$ unit matrix. Some matrix manipulations then yield

$$\begin{aligned} \boldsymbol{D}_{ww,T}^{-1}\frac{\boldsymbol{W}_i^\top\boldsymbol{U}_i(\boldsymbol{U}_i^\top\boldsymbol{U}_i)^{-1}\boldsymbol{U}_i^\top\boldsymbol{W}_i}{T}\boldsymbol{D}_{ww,T}^{-1} &\xrightarrow{p} \boldsymbol{Q}_{ww,i}^{(1)}, \\ \boldsymbol{D}_{ww,T}^{-1}\frac{\boldsymbol{W}_i^\top\boldsymbol{W}_i}{T}\boldsymbol{D}_{ww,T}^{-1} &\xrightarrow{p} \boldsymbol{Q}_{ww,i}^{(2)}. \end{aligned}$$

Hence, (4.3) immediately follows. In addition, note that

$$\begin{aligned} \boldsymbol{Z}_i^\top\boldsymbol{\xi}_{0,i} &\ll \sum_{t=1}^T t^{1/2} \approx T^{3/2} \text{ w.p.}, \\ (\boldsymbol{X}_i^\ell)^\top\boldsymbol{\xi}_{0,i} &\ll \sum_{t=1}^T t^{\ell/2} = T^{\frac{\ell+2}{2}}, \text{ w.p.} \\ \boldsymbol{Z}_i^\top\boldsymbol{\xi}_{0,i} &\ll T^{3/2} \text{ w.p.}, \\ (\boldsymbol{X}_i^\ell)^\top\boldsymbol{\xi}_{0,i} &\ll T^{\frac{\ell+2}{2}} \text{ w.p.}. \end{aligned}$$

One can immediately show (4.4) and (4.5). □

Appendix C. Proofs of Main Theorems

Proof of Theorem 2. We adopt the strategy used in Saikkonen (1995) and Pesaran, Shin and Smith (1998). First, define the open shrinking balls: $B_T(\phi_\ell, \delta_{\phi_\ell}) = \{\phi_\ell \in \Theta_{\phi_\ell} : T^{\frac{\ell}{2}}|\phi_\ell - \phi_{0,\ell}| < \delta_{\phi_\ell}\}$ for $\ell = 1, \ldots, k_w$, where ϕ_ℓ in the ℓ-th element of $\boldsymbol{\phi}$ and Θ_{ϕ_ℓ} is the marginal subset of $\Theta_\phi = \Theta_{\phi_1} \times \cdots \times \Theta_{\phi_{k_w}}$, and $B(\boldsymbol{\sigma}_0, \delta_\sigma) = \{\boldsymbol{\sigma} \in \Theta_\sigma : \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_0\| < \delta_\sigma\}$; and let $B_T^c(\phi_\ell, \delta_{\phi_\ell})$ and $B^c(\boldsymbol{\sigma}_0, \delta_\sigma)$ be their complements in Θ_{ϕ_ℓ} and Θ_σ respectively. Define $C_T(\boldsymbol{\varphi}, \delta, \delta_\sigma) = \bigcup_{\left\{\delta_{\phi_\ell},\ \ell=1,\ldots,k_w:\ \left(\sum_{\ell=1}^{k_w}\delta_{\phi_\ell}^2\right)^{1/2}=\delta\right\}} B_T^c(\phi_\ell, \delta_{\phi_\ell}) \times$

$B^c(\boldsymbol{\sigma}_0, \delta_\sigma)$. We need to show that

$$\lim_{T\uparrow\infty} P\left(\inf_{\boldsymbol{\varphi}\in C_T(\boldsymbol{\varphi},\delta,\delta_\sigma)} \frac{1}{T}(\ell_T(\boldsymbol{\varphi}_0) - \ell_T(\boldsymbol{\varphi})) > 0\right) = 1 \tag{C-1}$$

for every $\delta,\ \delta_\sigma > 0$. In view of (3.3), we have

$$\begin{aligned}\inf_{\boldsymbol{\varphi}\in C_T(\boldsymbol{\varphi},\delta,\delta_\sigma)} \frac{1}{T}(\ell_T(\boldsymbol{\varphi}_0) - \ell_T(\boldsymbol{\varphi})) &\geq \frac{1}{2}\left(\inf_{\boldsymbol{\sigma}\in B^c(\boldsymbol{\sigma}_0,\delta_\sigma)} \mathcal{T}_{1,T} + \inf_{\boldsymbol{\sigma}\in B^c(\boldsymbol{\sigma}_0,\delta_\sigma)} \mathcal{T}_{2,T}\right) \\ &+ \frac{1}{2}\inf_{\boldsymbol{\varphi}\in C_T(\boldsymbol{\varphi},\delta,\delta_\sigma)} (\boldsymbol{\phi}-\boldsymbol{\phi}_0)^\top \left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i}{T}\right)(\boldsymbol{\phi}-\boldsymbol{\phi}_0) \\ &+ \inf_{\boldsymbol{\varphi}\in C_T(\boldsymbol{\varphi},\delta,\delta_\sigma)} (\boldsymbol{\phi}_0-\boldsymbol{\phi})^\top \left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{\epsilon}_i}{T}\right).\end{aligned}$$

By Lemma 1, one can show that $\frac{\boldsymbol{\epsilon}_i^\top \boldsymbol{P}_i \boldsymbol{\epsilon}_i}{T} - \sigma_{0,i}^2 = o_p(1)$. As $\boldsymbol{\sigma} \in \Theta_\sigma$ are bounded, then $\mathcal{T}_{1,T} \xrightarrow{p} 0$. Therefore, $\inf_{\boldsymbol{\sigma}\in B^c(\boldsymbol{\sigma}_0,\delta_\sigma)} \mathcal{T}_{1,T} = o_p(1)$. One can also immediately show that $\inf_{\boldsymbol{\sigma}\in B^c(\boldsymbol{\sigma}_0,\delta_\sigma)} \mathcal{T}_{2,T} > 0$. Since, by Lemma 1,

$$\begin{aligned}(\boldsymbol{\phi}_0-\boldsymbol{\phi})^\top \left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{\epsilon}_i}{T}\right) &= (\boldsymbol{\phi}_0-\boldsymbol{\phi})^\top \frac{\boldsymbol{D}_T}{\sqrt{T}}\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\frac{\boldsymbol{D}_T^{-1}\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{\epsilon}_i}{\sqrt{T}}\right) \\ &= O_P(T^{-1/2})(\boldsymbol{\phi}_0-\boldsymbol{\phi})^\top \frac{\boldsymbol{D}_T}{\sqrt{T}}\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\boldsymbol{Q}^*_{w\epsilon,i}\right),\end{aligned}$$

one has $\inf_{\boldsymbol{\varphi}\in C_T(\boldsymbol{\varphi},\delta,\delta_\sigma)}(\boldsymbol{\phi}_0-\boldsymbol{\phi})^\top \left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{\epsilon}_i}{T}\right) = o_P(1)$ because, for each $\boldsymbol{\phi} \in \bigcup_{\left\{\delta_{\phi_\ell},\ \ell=1,\dots,k_w:\ \left(\sum_{\ell=1}^{k_w}\delta_{\phi_\ell}^2\right)^{1/2}=\delta\right\}} B_T^c(\phi_\ell, \delta_{\phi_\ell})$ for every $\delta > 0$, $(\boldsymbol{\phi}_0-\boldsymbol{\phi})^\top \frac{\boldsymbol{D}_T}{\sqrt{T}}\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\boldsymbol{Q}^*_{w\epsilon,i}\right)$ is stochastically bounded either above or below by some constant not depending on T. Finally, by the inequality for quadratic forms: $\boldsymbol{m}^\top \boldsymbol{W}\boldsymbol{m} \geq \|\boldsymbol{m}\|^2\lambda_1(\boldsymbol{W})$ and Lemma 1, one can show that

$$\begin{aligned}&\inf_{\boldsymbol{\varphi}\in C_T(\boldsymbol{\varphi},\delta,\delta_\sigma)} (\boldsymbol{\phi}-\boldsymbol{\phi}_0)^\top \left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i}{T}\right)(\boldsymbol{\phi}-\boldsymbol{\phi}_0) \\ &\qquad \geq \inf_{\boldsymbol{\varphi}\in C_T(\boldsymbol{\varphi},\delta,\delta_\sigma)} \left\|\frac{\boldsymbol{D}_T}{\sqrt{T}}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)\right\|^2 \inf_{\boldsymbol{\sigma}\in B^c(\boldsymbol{\sigma}_0,\delta_\sigma)} \lambda_1\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\boldsymbol{Q}^*_{ww,i}\right) \geq \delta^2 C_{ww} > 0.\end{aligned}$$

Collecting all the terms above, the theorem has been proved. □

Proof of Theorem 3. One has the first and second order derivatives of $\ell_T(\boldsymbol{\varphi})$:

$$\begin{aligned}\frac{\partial \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\phi}} &= \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \boldsymbol{W}_i^\top \boldsymbol{P}_i (\boldsymbol{y}_i - \boldsymbol{W}_i \boldsymbol{\phi}),\\ \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= -\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i.\end{aligned}$$

Doing a first-order Taylor expansion of $\frac{\partial \ell_T(\widehat{\boldsymbol{\varphi}})}{\partial \boldsymbol{\phi}}$ about $\boldsymbol{\phi}_0$ yields

$$\frac{\partial^2 \ell_T(\boldsymbol{\phi}^*, \widehat{\boldsymbol{\sigma}})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} (\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = -\frac{\partial \ell_T(\boldsymbol{\phi}_0, \widehat{\boldsymbol{\sigma}})}{\partial \boldsymbol{\phi}},$$

where $\boldsymbol{\phi}^*$ is some point on the line segment $L(\boldsymbol{\phi}_0, \widehat{\boldsymbol{\phi}}) = \{\boldsymbol{\phi} = s\boldsymbol{\phi}_0 + (1-s)\widehat{\boldsymbol{\phi}} : \ s \in (0,1)\}$. One has

$$\boldsymbol{D}_T(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) = \left(\sum_{i=1}^{N} \frac{1}{\widehat{\sigma}_i^2} \boldsymbol{D}_T^{-1} \boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i \boldsymbol{D}_T^{-1} \right)^{-1} \left(\sum_{i=1}^{N} \frac{1}{\widehat{\sigma}_i^2} \boldsymbol{D}_T^{-1} \boldsymbol{W}_i \boldsymbol{P}_i \epsilon_i \right) = I_T(\widehat{\boldsymbol{\sigma}})^{-1} J_T(\widehat{\boldsymbol{\sigma}}). \quad \text{(C-2)}$$

Note that, by Lemma 1 and Theorem 2, it follows that

$$I_T(\widehat{\boldsymbol{\sigma}}) \xrightarrow{p} \sum_{i=1}^{N} \frac{1}{\sigma_{0,i}^2} \boldsymbol{Q}_{ww,i}^*.$$

Moreover, since $J_T(\widehat{\boldsymbol{\sigma}}) = J_T(\boldsymbol{\sigma}_0) + o_p(1)$. By conditioning on $\boldsymbol{W}_i$ and $\boldsymbol{P}_i$, in view of Assumption 3.2, an application of the multivariate CLT yields

$$J_T(\boldsymbol{\sigma}_0) \xrightarrow{d} N\left(0, \sum_{i=1}^{N} \frac{1}{\sigma_{0,i}^2} \boldsymbol{Q}_{ww,i}^*\right) \quad \text{(C-3)}$$

conditional on $\boldsymbol{W}_i$ and $\boldsymbol{P}_i$. Therefore, it follows from (C-2) and marginal integration that

$$\boldsymbol{D}_T(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \xrightarrow{d} MN\left(0, \left(\sum_{i=1}^{N} \frac{1}{\sigma_{0,i}^2} \boldsymbol{Q}_{ww,i}^*\right)^{-1}\right).$$

□

Proof of Theorem 5. Following the argument used in the proof of Theorem 2, one needs to define some open neighborhoods, $B_T(\boldsymbol{\theta}_0, \delta_\theta) = \{\boldsymbol{\theta} \in \Theta_\theta \subset \mathbb{R}^{k_w} : \|\boldsymbol{D}_{ww,T}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| < \delta_\theta\}$, where $k_w = k_z + k_x$ and Θ_θ is some compact parameter space of $\boldsymbol{\theta}_0$; $B(\boldsymbol{\phi}_0, \delta_\phi) = \{\boldsymbol{\phi} \in \Theta_\phi \subset \mathbb{R}^N : \|\boldsymbol{\phi} - \boldsymbol{\phi}_0\| < \delta_\phi\}$, where Θ_ϕ is some compact parameter space of $\boldsymbol{\phi}_0$; and $B(\boldsymbol{\sigma}_0, \delta_\sigma) = \{\boldsymbol{\sigma} \in \Theta_\sigma \subset \mathbb{R}^N : \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_0\| < \delta_\sigma\}$, where Θ_σ is some compact parameter space of $\boldsymbol{\sigma}_0$. Let $B_T^c(\boldsymbol{\theta}_0, \delta_\theta)$, $B^c(\boldsymbol{\phi}_0, \delta_\phi)$, and $B^c(\boldsymbol{\sigma}_0, \delta_\sigma)$ be the complements of $B_T(\boldsymbol{\theta}_0, \delta_\theta)$, $B(\boldsymbol{\phi}_0, \delta_\phi)$, and $B(\boldsymbol{\sigma}_0, \delta_\sigma)$ respectively. Define $\mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma) = \left\{\bigcup_{\{\delta_\theta, \delta_\phi:\ (\delta_\theta^2 + \delta_\phi^2)^{1/2} = \delta\}} B_T^c(\boldsymbol{\theta}_0, \delta_\theta) \times B^c(\boldsymbol{\phi}_0, \delta_\phi)\right\} \times B^c(\boldsymbol{\sigma}_0, \delta_\sigma)$. As in the proof of Theorem 2, we need to show that

$$\lim_{T\uparrow\infty} P\left(\inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} \frac{1}{T}(\ell_T(\boldsymbol{\varphi}_0) - \ell_T(\boldsymbol{\varphi})) > 0\right) = 1 \tag{C-4}$$

for every $\delta,\ \delta_\sigma > 0$. In view of (4.2), one obtains

$$\inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} \frac{1}{T}(\ell_T(\boldsymbol{\varphi}_0) - \ell_T(\boldsymbol{\varphi})) \geq \frac{1}{2}\left\{\inf_{\boldsymbol{\sigma} \in B^c(\boldsymbol{\sigma}_0, \delta_\sigma)} \mathcal{T}_{1,T}(\boldsymbol{\sigma}, \boldsymbol{\sigma}_0) + \inf_{\boldsymbol{\sigma} \in B^c(\boldsymbol{\sigma}_0, \delta_\sigma)} \mathcal{T}_{2,T}(\boldsymbol{\sigma}, \boldsymbol{\sigma}_0) + \inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} \mathcal{T}_{3,T}(\boldsymbol{\varphi})\right\}.$$

By the same argument as the proof of Theorem 2, it can be shown that $\inf_{\boldsymbol{\sigma} \in B^c(\boldsymbol{\sigma}_0, \delta_\sigma)} \mathcal{T}_{1,T}(\boldsymbol{\sigma}, \boldsymbol{\sigma}_0) = o_p(1)$ and $\inf_{\boldsymbol{\sigma} \in B^c(\boldsymbol{\sigma}_0, \delta_\sigma)} \mathcal{T}_{2,T}(\boldsymbol{\sigma}, \boldsymbol{\sigma}_0) > 0$. Furthermore,

$$\inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} \mathcal{T}_{3,T}(\boldsymbol{\varphi}) \geq \inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} (\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^\top \boldsymbol{G}_T(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0) + 2\inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} (\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^\top \boldsymbol{F}_T = \mathcal{T}_{3,a,T}(\boldsymbol{\varphi}) + 2\mathcal{T}_{3,b,T}(\boldsymbol{\varphi}).$$

Note that, by an elementary matrix inequality and Lemma 2,

$$\begin{aligned}
\mathcal{T}_{3,a,T}(\boldsymbol{\varphi}) &= \inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} (\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^\top \boldsymbol{D}_{G,T}[\boldsymbol{D}_{G,T}^{-1} \boldsymbol{G}_T \boldsymbol{D}_{G,T}^{-1}] \boldsymbol{D}_{G,T}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0) \\
&\geq \inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} \|\boldsymbol{D}_{G,T}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)\|^2 \inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} \lambda_1\left(\boldsymbol{D}_{G,T}^{-1} \boldsymbol{G}_T \boldsymbol{D}_{G,T}^{-1}\right) \\
&\geq \delta^2 \inf_{\substack{\boldsymbol{\phi} \in B(\boldsymbol{\phi}_0, \delta_\phi) \\ \boldsymbol{\sigma} \in B(\boldsymbol{\sigma}_0, \delta_\sigma)}} \lambda_1(\boldsymbol{Q}_G) \text{ w.p.}
\end{aligned}$$

It then follows that $\mathcal{T}_{3,a,T}(\boldsymbol{\varphi}) > 0$ w.p. Moreover, by Lemma 2, one has

$$\begin{aligned} \boldsymbol{D}_{ww,T}^{-1} \frac{\boldsymbol{W}_i^\top \boldsymbol{P}_i E[\boldsymbol{\epsilon}_i^\top \boldsymbol{\epsilon}] \boldsymbol{P}_i \boldsymbol{W}_i}{T} \boldsymbol{D}_{ww,T}^{-1} &\xrightarrow{p} \sigma_i^2 \boldsymbol{Q}_{ww,i}, \\ \frac{\boldsymbol{\xi}_{0,i}^\top \boldsymbol{P}_i E[\boldsymbol{\epsilon}_i^\top \boldsymbol{\epsilon}] \boldsymbol{P}_i \boldsymbol{\xi}_{0,i}}{T} &\xrightarrow{p} \sigma_i^2 \boldsymbol{Q}_{\xi\xi,i}. \end{aligned}$$

Conditioning $\boldsymbol{F}_T$ on $\boldsymbol{W}_i$, $\boldsymbol{P}_i$ and $\boldsymbol{\xi}_i$, an application of the multivariate CLT to the sequence $\boldsymbol{\epsilon}_i$ yields $\boldsymbol{D}_{G,T}^{-1} \boldsymbol{F}_T = O_p\left(T^{-1/2}\right) = o_p(1)$. Since, from the way $\mathcal{B}_T(\boldsymbol{\theta}_0, \delta_\theta)$ is defined, the term $\inf_{\boldsymbol{\varphi} \in \mathcal{B}_T(\boldsymbol{\varphi}, \delta, \delta_\sigma)} (\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0)^\top \boldsymbol{D}_{G,T}^{-1} \boldsymbol{\iota}_{k_w+N}$ is bounded either above or below by a generic constant, which can be large but does not depend on T, it immediately follows that $\mathcal{T}_{3,b,T} = o_p(1)$. Therefore, (C-4) has been verified. □

Proof of Theorem 6. The gradient and Hessian matrices of $\ell_T(\boldsymbol{\varphi})$ are given by $\frac{\partial \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\Gamma}} = \left(\frac{\partial \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\theta}^\top}, \frac{\partial \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\phi}^\top} \right)^\top$ and $\frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}^\top} = \begin{pmatrix} \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} & \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\phi}} \\ \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\phi}} & \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} \end{pmatrix}$, where

$$\begin{aligned} \frac{\partial \ell_T(\boldsymbol{\varphi})}{\partial \phi_i} &= \frac{1}{\sigma_i^2} \boldsymbol{\xi}_i(\boldsymbol{\theta})^\top \boldsymbol{P}_i (\Delta \boldsymbol{y}_i - \phi_i \boldsymbol{\xi}_i(\boldsymbol{\theta})), \\ \frac{\partial \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\theta}} &= -\sum_{i=1}^N \frac{\phi_i}{\sigma_i^2} \boldsymbol{W}_i^\top \boldsymbol{P}_i (\Delta \boldsymbol{y}_i - \phi_i \boldsymbol{\xi}_i(\boldsymbol{\theta})), \\ \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\phi_i \phi_j} &= 0 \text{ for } i \neq j, \\ \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\phi_i^2} &= -\frac{1}{\sigma_i^2} \boldsymbol{\xi}_i(\boldsymbol{\theta})^\top \boldsymbol{P}_i \boldsymbol{\xi}_i(\boldsymbol{\theta}), \\ \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= -\sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2} \boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{W}_i, \\ \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial \boldsymbol{\theta} \partial \phi_i} &= -\frac{1}{\sigma_i^2} \boldsymbol{W}_i^\top \boldsymbol{P}_i (\Delta \boldsymbol{y}_i - \phi_i \boldsymbol{\xi}_i(\boldsymbol{\theta})) + \frac{\phi_i}{\sigma_i^2} \boldsymbol{W}_i^\top \boldsymbol{P}_i \boldsymbol{\xi}_i(\boldsymbol{\theta}). \end{aligned}$$

Since $\widehat{\boldsymbol{\varphi}}$ is consistent by Theorem 5, an application of a first-order Taylor expansion of $\frac{\partial \ell_T(\widehat{\boldsymbol{\varphi}})}{\partial \boldsymbol{\Gamma}}$ about $\boldsymbol{\Gamma}_0$ yields

$$0 = \frac{\partial \ell_T(\widehat{\boldsymbol{\varphi}})}{\partial \boldsymbol{\Gamma}} = \frac{\partial \ell_T(\boldsymbol{\Gamma}_0, \widehat{\boldsymbol{\sigma}})}{\partial \boldsymbol{\Gamma}} + \frac{\partial^2 \ell_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}})}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}^\top} (\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0),$$

where $\boldsymbol{\Gamma}^*$ is some point lying on the line segment $L(\boldsymbol{\Gamma}_0, \widehat{\boldsymbol{\Gamma}}) = \{s\boldsymbol{\Gamma}_0 + (1-s)\widehat{\boldsymbol{\Gamma}} : \ s \in (0,1)\} \subset \Theta_\theta \times \Theta_\phi \subset \mathbb{R}^{k_{wn}}$, where $\Theta_\theta \times \Theta_\phi$ are the compact parameter spaces of $\boldsymbol{\Gamma}_0$ (as defined in the proof of Theorem 5), and $k_{wn} = k_w + N$. One can then obtain

$$\boldsymbol{D}_T(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0) = -\left[\boldsymbol{D}_T^{-1}\frac{\partial^2 \ell_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}})}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}^\top}\boldsymbol{D}_T^{-1}\right]^{-1}\boldsymbol{D}_T^{-1}\frac{\partial \ell_T(\boldsymbol{\Gamma}_0, \widehat{\boldsymbol{\sigma}})}{\partial\boldsymbol{\Gamma}}. \tag{C-5}$$

For notational brevity, let $\mathcal{I}_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}}) = \boldsymbol{D}_T^{-1}\frac{\partial^2 \ell_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}})}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}^\top}\boldsymbol{D}_T^{-1}$. First, one needs to show that

$$\lim_{T\uparrow\infty} P\left(\|\mathcal{I}_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0, \boldsymbol{\sigma}_0)\| > \epsilon\right) = 0 \text{ given some arbitrarily small } \epsilon > 0. \tag{C-6}$$

Note that

$$\begin{aligned}
&P\left(\|\mathcal{I}_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0, \boldsymbol{\sigma}_0)\| > \epsilon\right) \\
&\quad = P\left(\|\mathcal{I}_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0, \boldsymbol{\sigma}_0)\| > \epsilon \,|\boldsymbol{\Gamma}^* \in B_T(\boldsymbol{\theta}_0, \delta_\theta) \times B(\boldsymbol{\phi}_0, \delta_\phi), \widehat{\boldsymbol{\sigma}} \in B(\boldsymbol{\sigma}_0, \delta_\sigma)\right) \\
&\qquad\qquad P\left(\boldsymbol{\Gamma}^* \in B_T(\boldsymbol{\theta}_0, \delta_\theta) \times B(\boldsymbol{\phi}_0, \delta_\phi), \widehat{\boldsymbol{\sigma}} \in B(\boldsymbol{\sigma}_0, \delta_\sigma)\right) \\
&\quad + P\left(\|\mathcal{I}_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0, \boldsymbol{\sigma}_0)\| > \epsilon \,|\boldsymbol{\Gamma}^* \in B_T^c(\boldsymbol{\theta}_0, \delta_\theta) \times B^c(\boldsymbol{\phi}_0, \delta_\phi), \widehat{\boldsymbol{\sigma}} \in B^c(\boldsymbol{\sigma}_0, \delta_\sigma)\right) \\
&\qquad\qquad P\left(\boldsymbol{\Gamma}^* \in B_T^c(\boldsymbol{\theta}_0, \delta_\theta) \times B^c(\boldsymbol{\phi}_0, \delta_\phi), \widehat{\boldsymbol{\sigma}} \in B^c(\boldsymbol{\sigma}_0, \delta_\sigma)\right),
\end{aligned}$$

where the balls $B_T(\boldsymbol{\theta}_0, \delta_\theta)$, $B(\boldsymbol{\phi}_0, \delta_\phi)$, and $B(\boldsymbol{\sigma}_0, \delta_\sigma)$ are defined in the proof of Theorem 5. Since $\lim_{T\uparrow\infty} P\left(\boldsymbol{\Gamma}^* \in B_T^c(\boldsymbol{\theta}_0, \delta_\theta) \times B^c(\boldsymbol{\phi}_0, \delta_\phi), \widehat{\boldsymbol{\sigma}} \in B^c(\boldsymbol{\sigma}_0, \delta_\sigma)\right) = 0$ for every $\boldsymbol{\Gamma}^*$ lying on the line segment $L(\boldsymbol{\Gamma}_0, \widehat{\boldsymbol{\Gamma}})$ by Theorem 5, one has

$$\lim_{T\uparrow\infty} P\left(\|\mathcal{I}_T(\boldsymbol{\Gamma}^*, \widehat{\boldsymbol{\sigma}}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0, \boldsymbol{\sigma}_0)\| > \epsilon\right) \leq \lim_{T\uparrow\infty} P\left(\sup_{\substack{\boldsymbol{\Gamma} \in B_T(\boldsymbol{\theta}_0, \delta_\theta) \times B(\boldsymbol{\phi}_0, \delta_\phi) \\ \boldsymbol{\sigma} \in B(\boldsymbol{\sigma}_0, \delta_\sigma)}} \|\mathcal{I}_T(\boldsymbol{\Gamma}, \boldsymbol{\sigma}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0, \boldsymbol{\sigma}_0)\| > \epsilon\right) \tag{C-7}$$

for some arbitrarily small numbers, δ_θ, δ_ϕ and δ_σ. An application of Lemma 2 and some inequalities for matrices yields

$$\begin{aligned}
\sup_{\substack{\boldsymbol{\phi}\in B(\boldsymbol{\phi}_0,\delta_\phi)\\ \boldsymbol{\sigma}\in B(\boldsymbol{\sigma}_0,\delta_\sigma)}} \left\| \boldsymbol{D}_{\gamma,k_w}^{-1} \left(\frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2 \ell_T(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \right) \boldsymbol{D}_{\gamma,k_w}^{-1} \right\| &\leq C_0 \left(\delta_\phi^2 + \delta_\sigma^2\right)^{\frac{1}{2}} \sum_{i=1}^N \|\boldsymbol{Q}_{ww,i}\|, \\
\sup_{\substack{\boldsymbol{\phi}\in B(\boldsymbol{\phi}_0,\delta_\phi)\\ \boldsymbol{\sigma}\in B(\boldsymbol{\sigma}_0,\delta_\sigma)}} \left\| \boldsymbol{D}_{\gamma,k_w}^{-1} \left(\frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\phi}^\top} - \frac{\partial^2 \ell_T(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\phi}^\top} \right) \boldsymbol{I}_N T^{1/2} \right\| &\leq C_0 \left(\delta_\theta \sum_{i=1}^N \|\boldsymbol{Q}_{ww,i}\| + \left(\delta_\phi^2 + \delta_\sigma^2\right)^{\frac{1}{2}} \sum_{i=1}^N \|\boldsymbol{Q}_{w\xi,i}\| \right), \\
\sup_{\substack{\boldsymbol{\phi}\in B(\boldsymbol{\phi}_0,\delta_\phi)\\ \boldsymbol{\sigma}\in B(\boldsymbol{\sigma}_0,\delta_\sigma)}} \left\| \frac{\partial^2 \ell_T(\boldsymbol{\varphi})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} - \frac{\partial^2 \ell_T(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} \right\| &\leq \delta_\sigma \left(\sum_{i=1}^N \|\boldsymbol{Q}_{\xi\xi,i}\| \right)^{1/2},
\end{aligned}$$

where C_0 is some finite generic constant that may differ from a line to another one. An application of the matrix inequality: $\| \begin{smallmatrix} A & C \\ C & D \end{smallmatrix} \| \leq \|A\|_2 + \sqrt{2}\|C\|_2 + \|D\|_2$ yields

$$\begin{aligned}
&\sup_{\substack{\boldsymbol{\Gamma}\in B_T(\boldsymbol{\theta}_0,\delta_\theta)\times B(\boldsymbol{\phi}_0,\delta_\phi)\\ \boldsymbol{\sigma}\in B(\boldsymbol{\sigma}_0,\delta_\sigma)}} \|\mathcal{I}_T(\boldsymbol{\Gamma},\boldsymbol{\sigma}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0,\boldsymbol{\sigma}_0)\| \\
&\quad \leq C_0 \left(\left(\delta_\theta + (\delta_\phi^2 + \delta_\sigma^2)^{1/2}\right) \sum_{i=1}^N \|\boldsymbol{Q}_{ww,i}\| + (\delta_\phi^2 + \delta_\sigma^2)^{1/2} \sum_{i=1}^N \|\boldsymbol{Q}_{w\xi,i}\| + \delta_\sigma \sum_{i=1}^N \|\boldsymbol{Q}_{\xi\xi,i}\| \right). \quad \text{(C-8)}
\end{aligned}$$

The consistency of $\widehat{\boldsymbol{\Gamma}}$ allows one to make δ_θ, δ_ϕ, and δ_σ in (C-8) arbitrarily small such that its RHS becomes less than ϵ. In view of (C-7), (C-6) has been proved. Therefore, $\|\mathcal{I}_T(\boldsymbol{\Gamma}^*,\widehat{\boldsymbol{\sigma}}) - \mathcal{I}_T(\boldsymbol{\Gamma}_0,\boldsymbol{\sigma}_0)\| = o_p(1)$. By the same argument, one can also show that

$$\left\| \boldsymbol{D}_T^{-1} \left(\frac{\partial \ell_T(\boldsymbol{\Gamma}_0,\widehat{\boldsymbol{\sigma}})}{\partial\boldsymbol{\Gamma}} - \frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\Gamma}} \right) \right\| = o_p(1).$$

Moreover, by Lemma 2, one has

$$\mathcal{I}_T(\boldsymbol{\Gamma}_0,\boldsymbol{\sigma}_0) \xrightarrow{p} \boldsymbol{Q}_G(\boldsymbol{\phi}_0,\boldsymbol{\sigma}_0),$$

where $\boldsymbol{Q}_G$ is given in Theorem 5. Now, notice that for each

$$E_\epsilon \left[\boldsymbol{D}_T^{-1} \frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\Gamma}} \frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\Gamma}^\top} \boldsymbol{D}_T^{-1} \right] = \boldsymbol{Q}_G(\boldsymbol{\phi}_0,\boldsymbol{\sigma}_0),$$

where the expectation is taken with respect to the joint probability density of $\boldsymbol{\epsilon}_i$. Therefore, conditioning $\boldsymbol{D}_T^{-1}\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\Gamma}}$ on $\boldsymbol{W}_i$, $\boldsymbol{P}_i$, and $\boldsymbol{\xi}_i(\boldsymbol{\theta}_0)$, an application of the multivariate CLT to the sequence $\boldsymbol{\epsilon}_i$ yields

$$\boldsymbol{D}_T^{-1}\frac{\partial \ell_T(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\Gamma}} \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{Q}_G(\boldsymbol{\phi}_0, \boldsymbol{\sigma}_0)\right).$$

The main theorem then follows from (C-5) and some marginal integration. □

Appendix D. Confidence Set for Ratios of two Parameters [Tipping Points]

Consider the general model $(\mathcal{Y}, \{P_\theta : \theta \in \Theta\})$, $\Theta \subset R^p$, $p \geq 1$, where $\mathcal{Y}$ is the sample space and P_θ is a probability distribution over $\mathcal{Y}$ indexed by $\theta = (\theta_1, \theta_2, ..., \theta_p)'$. Our object of interest are functions of θ of the form $h(\theta) = L'\theta/K'\theta$ where L and K are nonstochastic $p \times 1$ vectors. Given a sample of size T, assume a consistent and asympotically normal estimator of θ is available $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p)' \overset{asy}{\sim} \mathrm{N}(\theta, \Sigma_\theta)$ where Σ_θ is estimated consistently by $\widehat{\Sigma}_\theta$. The discontinuity set $\{\theta \in \Theta : K'\theta = 0\}$ is clearly non-empty. In this context, the *delta* method expoits the following regualr asympttoic result:

$$h(\hat{\theta}) \overset{asy}{\sim} N\left(h(\theta), \frac{\partial h\left(\hat{\theta}\right)}{\partial \theta'}\hat{\Sigma}_\theta \frac{\partial h'\left(\hat{\theta}\right)}{\partial \theta}\right). \tag{D-1}$$

For the same problem, Fieller's method inverts a Wald-type test associated with the hypothesis $L'\theta - \delta_0 K'\theta = 0$ for a collection of fixed δ_0 values. For the ratio case presented in Section 2, Fieller's method involves assembling all δ_0 values such that $\theta_1 - \delta_0\theta_2 = 0$ is not rejected at the $\alpha\%$ using the $\mathbf{t}$-statistic $\left(\hat{\theta}_1 - \delta_0\hat{\theta}_2\right) / \left(\delta^2\hat{v}_2 - 2\delta_0\hat{v}_{12} + \hat{v}_1\right)^{1/2}$ which is asymptotically standard normal under the null hypothesis. The confidence set is thus defined as solution to following inequality in δ_0

$$\mathrm{FCS}(\delta; \alpha) = \left\{\delta_0 : \left(\hat{\theta}_1 - \delta_0\hat{\theta}_2\right)^2 \leq z_{\alpha/2}^2\left(\hat{v}_1 + \delta_0^2\hat{v}_2 - 2\delta_0\hat{v}_{12}\right)\right\}. \tag{D-2}$$

This requires solving the following second-degree-polynomial inequality for δ_0:

$$A\delta_0^2 + 2B\delta_0 + C \leq 0 \tag{D-3}$$

$$A = \hat{\theta}_2^2 - z_{\alpha/2}^2\hat{v}_2, \quad B = -\hat{\theta}_1\hat{\theta}_2 + z_{\alpha/2}^2\hat{v}_{12}, \quad C = \hat{\theta}_1^2 - z_{\alpha/2}^2\hat{v}_1. \tag{D-4}$$

for real solutions δ_0. Except for a set of measure zero, $A \neq 0$. Similarly, except for a set of measure zero, $\Delta = B^2 - AC \neq 0$. Real roots

$$\delta_{01} = \frac{-B - \sqrt{\Delta}}{A}, \quad \delta_{02} = \frac{-B + \sqrt{\Delta}}{A}$$

exist if and only if $\Delta > 0$, so

$$\text{FCS}\,(\delta;\alpha) = \begin{cases} [\delta_{01}, \quad \delta_{02}] & if \quad A > 0 \\ ]-\infty, \quad \delta_{01}] \cup [\delta_{02}, \quad +\infty[& if \quad A < 0 \end{cases}. \tag{D-5}$$

Bolduc, Khalaf and Yelou (2010) further show that: (i) if $\Delta < 0$, then $A < 0$ and $\text{FCS}\,(\delta;\alpha) = R$; (ii) $\text{FCS}\,(\delta;\alpha)$ contains the point estimate $\hat{\delta} = \hat{\theta}_1/\hat{\theta}_2$ and thus cannot be empty, and (iii) asymptotically, Fieller's solution and the *delta* method give similar results when the former leads to an interval, *i.e.* when the denominator is far from zero.

Table 1: Parameters - for Monte Carlo Simulation

Parameters	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
ρ	0.2	0.8	0.9	0.99	0.2	0.8	0.9	0.99
ψ	1	1	1	1	0.2	0.2	0.2	0.2
ϕ	1	1	1	1	1	1	1	1
ω_{12}	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
β_0	-0.679	-0.679	-0.679	-0.679	-0.679	-0.679	-0.679	-0.679
β_1	0.619	0.619	0.619	0.619	0.619	0.619	0.619	0.619
β_2	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007

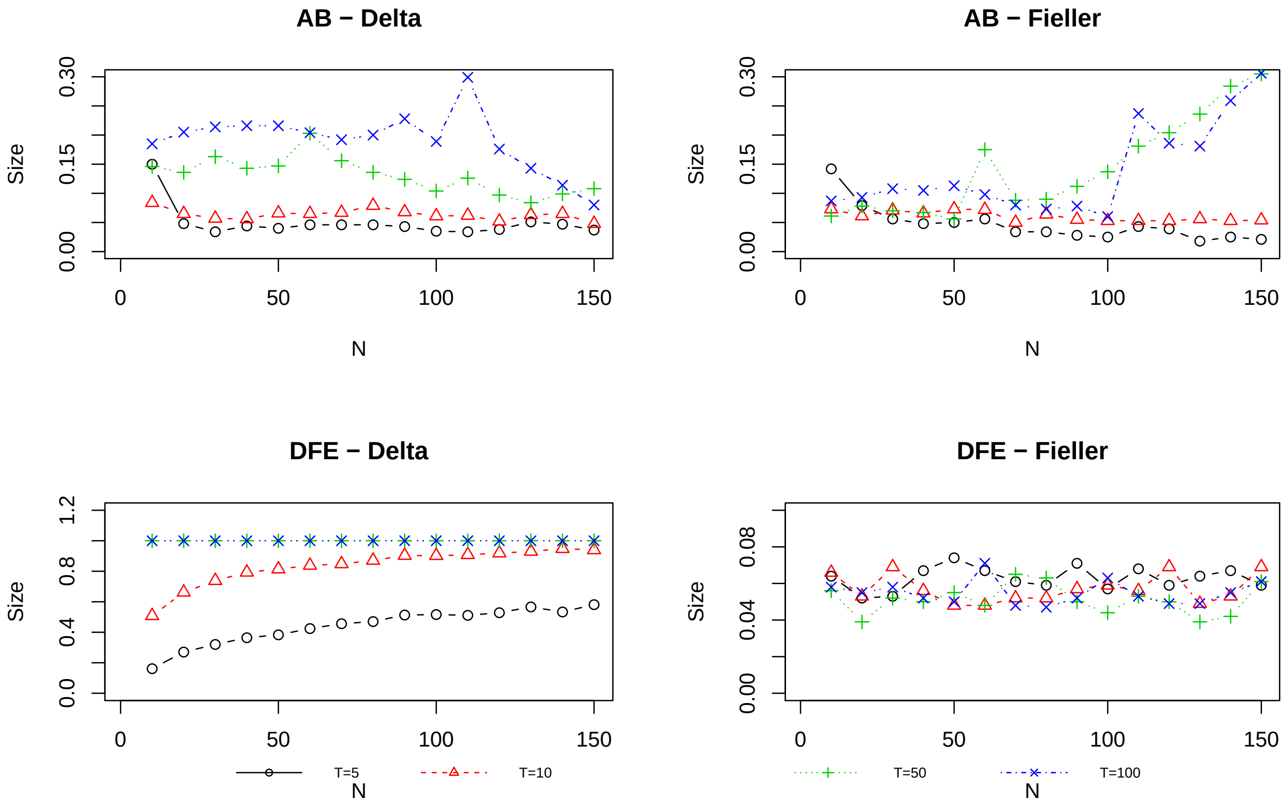

Figure 1: Size ρ=0.2, $\phi = 0.2$, $\psi = 1$

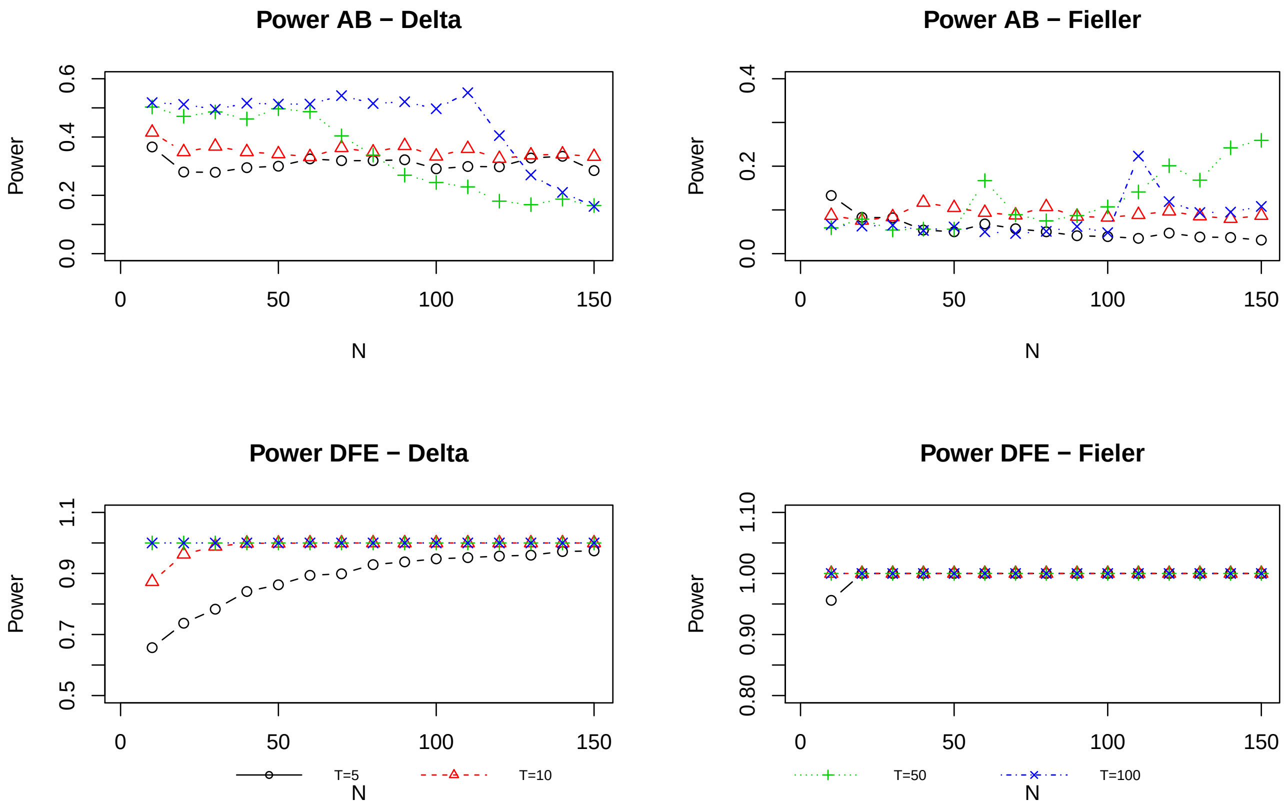

Figure 2: Power ρ=0.2, $\phi = 0.2$, $\psi = 1$

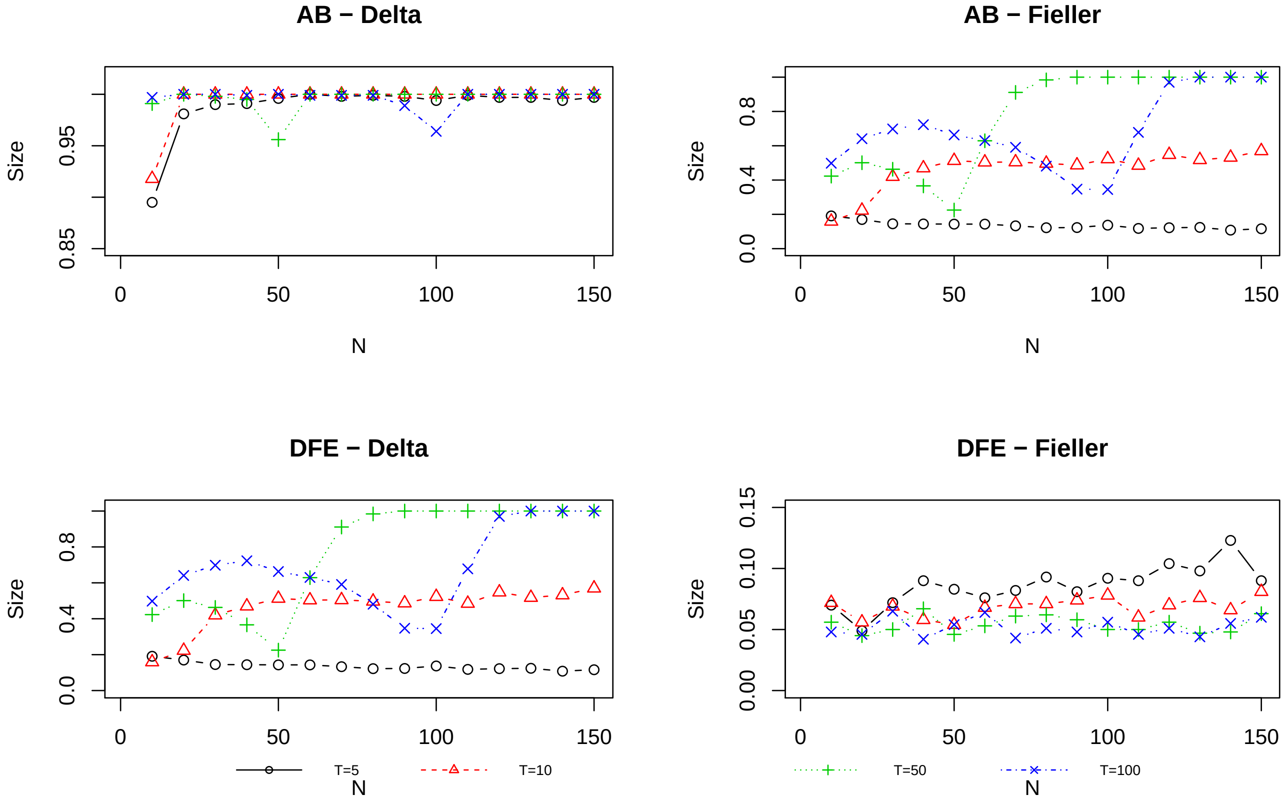

Figure 3: Size ρ=0.9, $\phi = 1$, $\psi = 0.2$

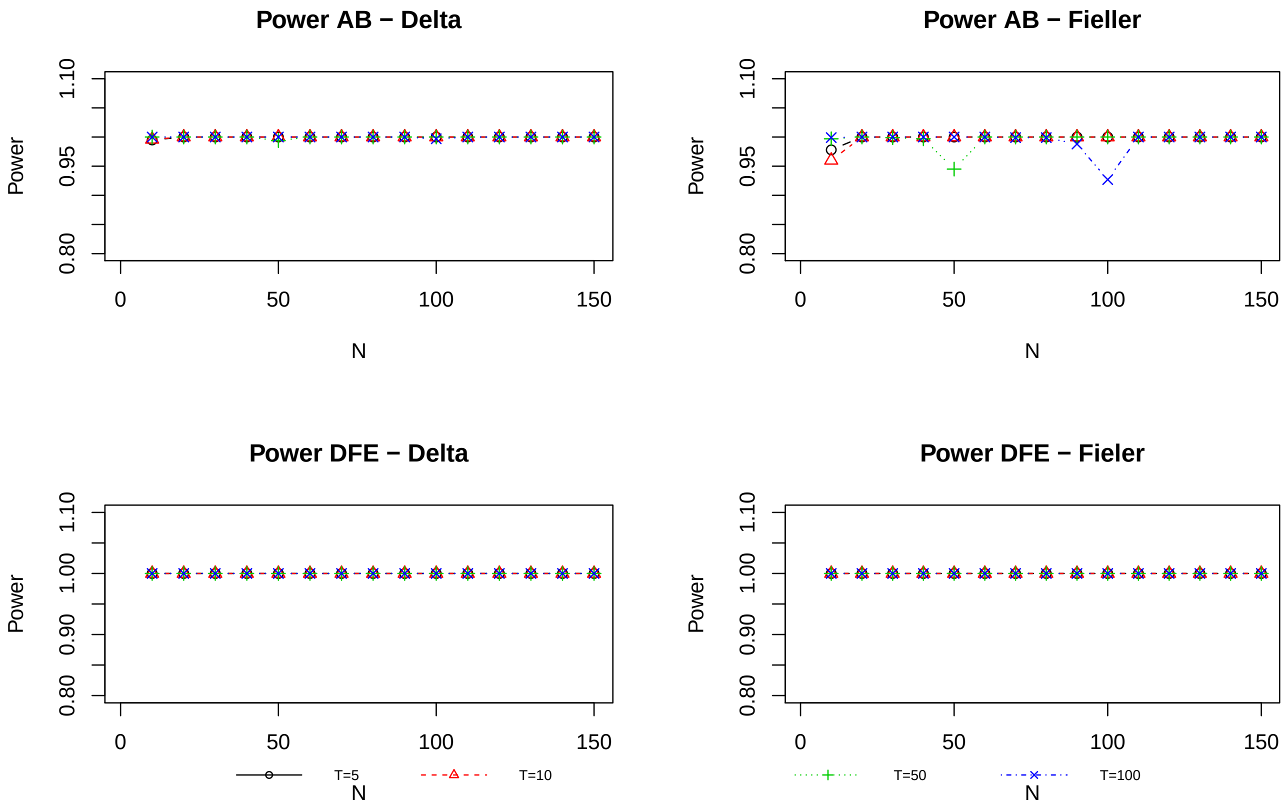

Figure 4: Power ρ=0.9, $\phi = 1$, $\psi = 0.2$

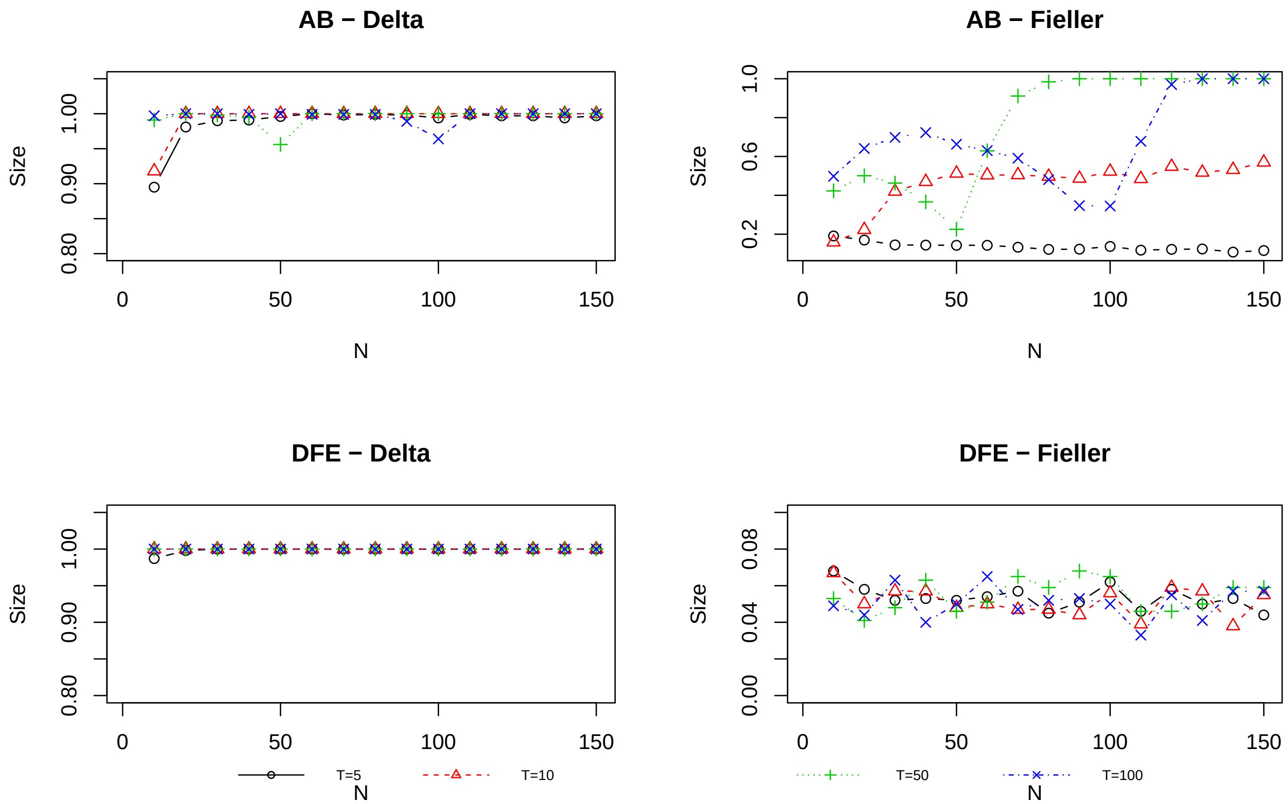

Figure 5: Size ρ=0.9, $\phi = 1$, $\psi = 1$

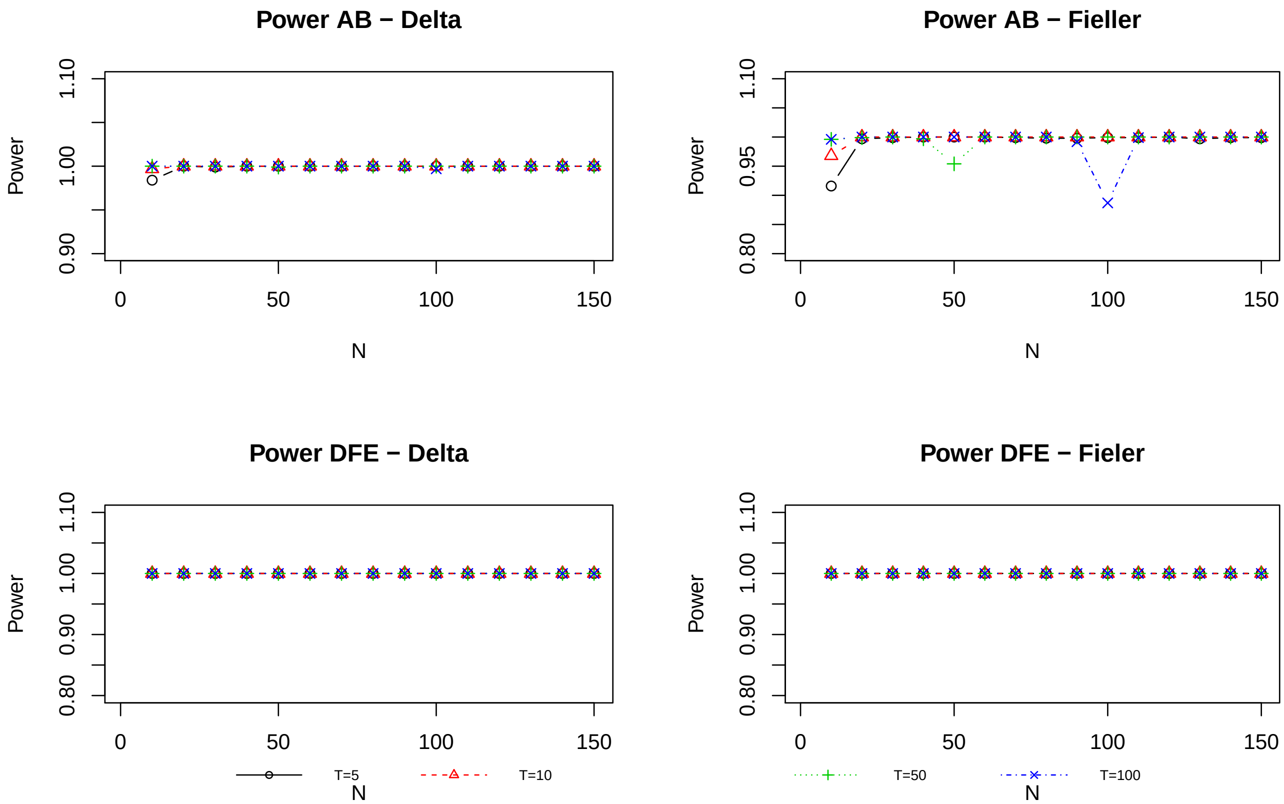

Figure 6: Power ρ=0.9, $\phi = 1$, $\psi = 1$