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Preliminary & Incomplete: Do Not Cite or Circulate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 distributions, 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 that are obtained from 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, at least in principle, using a pivotal statistic that is robust to identification 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: Bernard, Gavin, Khalaf and Voia (2015) discussed an empirical application of the environmental Kuznet curve. 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 Monte Carlo 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 (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 a well known 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 $N$ is large. Fieller’s method based on our likelihood based method works very well, even with unit roots, and interestingly, even when $N$ is large relative to $T$. 

3
This paper is structured as follows. Section 2 presents a general Fieller’s theorem for asymptotically mixed-normal estimators. Section 3 studies the problem of constructing Fieller’s confidence set for ratios of the parameters characterizing the long-run relationship in an error-correction representation of a dynamic heterogeneous polynomial panel. Asymptotic theory is derived for the case of fixed \( N \) and large \( T \). Section 4 summarizes our simulation findings, and Section 5 concludes the paper. Proofs of main theorems and lemmas as well as other materials of technical flavour can be found in four appendices at the end of this paper.

1.1 Notation

The following notation is used in the paper: \( X \) denotes a scalar, \( X \) is used to represent a vector or a matrix and \( C_0 \) is a generic constant that may vary from one context to another. For two random sequences, say \( a_T \) and \( b_T \), one often writes \( a_T \ll b_T \) a.s. if and only if \( P(\lim_{T\to\infty} |a_T/b_T| = \text{const.}) = 1 \), and \( a_T \ll b_T \) w.p. if and only if \( \lim_{T\to\infty} P(|a_T/b_T| < \xi) = 1 \), where \( \xi \) can be some almost-sure bounded random variable; \( o_p(\cdot) \) and \( O_p(\cdot) \) are standard symbols for stochastic orders of magnitude. \( \overset{p}{\longrightarrow} \) denotes convergence in probability and \( \overset{d}{\longrightarrow} \) denotes convergence in distribution. \( \| \cdot \| \) denotes the Euclidean norm of matrices and \( \lambda_1(X) \) represents the minimum eigenvalue of a square matrix, \( X \). \( I_n \) 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, \( \theta = (\theta_1, \ldots, \theta_p)^\top \). Let \( \theta_0 \in \Theta \subset \mathbb{R}^p \), where \( \Theta \) is a compact parameter space, represent the true parameters; and for a given data sample of size \( T \), one can estimate \( \theta_0 \) by \( \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)^\top \). We first make some assumptions about the asymptotic distribution of \( \hat{\theta} \). (Note that Assumptions 2.1 and 2.2 below are independent of each other, so are the notations.)

**Assumption 2.1.** \( \hat{\theta} \) is asymptotically normal as \( T \uparrow \infty \), in the sense that \( D_T(\hat{\theta} - \theta_0) \overset{d}{\longrightarrow} \)}
N \left( 0, \Sigma_0(\theta_0) \right) \text{ uniformly over } \Theta, \text{ where } D_T \text{ is a diagonal matrix consisting of normalizing factors that diverge to infinity with } T \text{ and may differ from one another; and } N \left( 0, \Sigma_0(\theta_0) \right) \text{ represents a tight family of Gaussian random variables with the asymptotic variance-covariance matrix, } \Sigma_0(\theta_0), \text{ which is the probability limit of a matrix of normalized sample statistics, } D_T^{-1}\Sigma_T D_T^{-1}.

**Assumption 2.2.** \( \hat{\theta} \) is asymptotically mixed normal as \( T \uparrow \infty \) such that

(a) \( D_T(\hat{\theta} - \theta_0) \overset{d}{\longrightarrow} \tilde{\Sigma}_0^{-1/2}(\theta_0) N(0, I_p) \) uniformly over \( \Theta \), where \( D_T \) is a diagonal matrix consisting of normalizing factors that diverge to infinity with \( T \) and may differ from one another; and \( \tilde{\Sigma}_0^{-1/2}(\theta_0) N(0, I_p) \) represents a tight family of Gaussian random variables with \( \tilde{\Sigma}_0^{-1} = \tilde{\Sigma}_0^{-1}(\theta_0) \) being some random asymptotic variance-covariance matrix that is independent of \( N(0, I_p) \);

(b) \( \tilde{\Sigma}_0 \) is the probability limit of a matrix of normalized sample statistics, \( D_T^{-1}\tilde{\Sigma}_T D_T^{-1} \), such that

\[
D_T^{-1}\tilde{\Sigma}_T D_T^{-1} \overset{p}{\longrightarrow} \tilde{\Sigma}_0.
\]

Our objects of interest are the ratios \( \rho = (\rho_1, \ldots, \rho_q) \) with \( \rho_i = \frac{L_i^\top D_T \theta_0}{K_i^\top D_T \theta_0} \) for \( i = 1, \ldots, q \), where \( L_1, \ldots, L_q \), and \( K \) 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 \( \rho_0 \), defined via the inverse relationship \( \lim_{T \uparrow \infty} \inf_{\theta_0 \in \Theta} P_{\theta_0}(\rho \in CS_T(\rho; \alpha)) \geq 1 - \alpha \), can be obtained by inverting the following Wald-type test statistic for the null hypothesis \( H_0: \ L_i^\top D_T \theta_0 - \rho_0,i K_i^\top D_T \theta_0 = 0 \) for all \( i = 1, \ldots, q \),

\[
W(\rho_0) = \hat{\theta}^\top D_T (L - K_\rho) \left( (L - K_\rho)^\top \left( D_T^{-1}\tilde{\Sigma}_T D_T^{-1} \right)^{-1} (L - K_\rho) \right)^{-1} (L - K_\rho)^\top D_T \hat{\theta} \overset{d}{\longrightarrow} \chi^2(q).
\]

(2.1)

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 \( \rho_0 \), defined via the inverse relationship \( \lim_{T \uparrow \infty} P_{\theta_0}(\rho \in CS_T(\rho; \alpha)) \geq 1 - \alpha \) for every \( \theta_0 \in \Theta \).

**Proof.** By defining \( p \times q \) matrices, \( L = (L_1, \ldots, L_q) \) and \( K_\rho = K \rho_0^\top \), the null hypothesis \( H_0 \) can also be written as \( (L^\top - K_\rho^\top) D_T \theta_0 = 0 \). An application of the uniform continuous mapping theorem
yields that

\[(L^\top - K_{\rho}^\top)D_T\hat{\theta} \xrightarrow{d} \left((L - K_{\rho})^\top\tilde{\Sigma}_0^{-1}(L - K_{\rho})\right)^{1/2} N(0, I_p) \text{ under } H_0.\]

By replacing the unknown \(\tilde{\Sigma}_0\) with \(D_T^{-1}\tilde{\Sigma}_T D_T^{-1}\), an application of Lemma 3 in Ogasawara and Takahishi (1951) yields the above Wald-type statistic. \(\square\)

**Remark 2.1.** One can then obtain the simultaneous confidence sets for \(\rho_0\):

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

where \(c_{q,\alpha}\) is the \((1 - \alpha)\) critical value of the \(\chi^2\) distribution with \(q\) degrees of freedom. Let \(R_\rho = \underbrace{R_\rho}_{q \times (q + 1)} = (I_q, -\rho_0), \quad H = \underbrace{(L, K)^\top}_{(q+1) \times p}\), and \(\hat{\Theta} = D_T\hat{\theta}\), we shall rewrite the above Wald-type test statistic as

\[W(\rho_0) = (R_\rho^\top H\hat{\Theta})^\top \left(R_\rho^\top H \left(D_T^{-1}\tilde{\Sigma}_T D_T^{-1}\right)^{-1} H^\top R_\rho^\top\right)^{-1} (R_\rho^\top H\hat{\Theta}).\]

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

**Remark 2.2.** Since the ratios \(\rho_i, i = 1, \ldots, q\), have a common denominator, the normalization matrices \(D_T\) in (2.1) can be cancelled out so that

\[W(\rho_0) = \hat{\theta}^\top(L - K_{\rho_0}^\top) \left(((L - K_{\rho_0}^\top)^\top\tilde{\Sigma}^{-1}(L - K_{\rho_0}^\top)\right)^{-1} ((L - K_{\rho_0}^\top)^\top \hat{\theta}.\]

To see this, notice that

\[\hat{\theta}^\top D_T^{-1}D_T^{-1}(L - K_{\rho_0}^\top) \left(((L - K_{\rho_0}^\top)^\top D_T^{-1}D_T^{-1}\tilde{\Sigma}^{-1} D_T D_T^{-1}(L - K_{\rho_0}^\top)\right)^{-1} ((L - K_{\rho_0}^\top)^\top D_T^{-1}D_T^{-1}\hat{\theta}\]
3 Dynamic Panel Polynomial Error Correction Models

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

$$\Delta y_{i,t} = \phi_i(y_{i,t-1} - \theta^\top W_{i,t}) + \sum_{j=1}^{p-1} \lambda_{i,j} \Delta y_{i,t-j} + \sum_{j=0}^{q_x-1} \gamma_{i,j} \Delta X_{i,t-j} + \sum_{j=0}^{q_z-1} \alpha_{i,j}^\top \Delta Z_{i,t-j} + \mu_i + \epsilon_{i,t}, \quad (3.1)$$

where $W_{i,t} = (Z_{i,t}^\top, X_{i,t}, X_{i,t}^2, \ldots, X_{i,t}^{k_x})^\top$, with $Z_{i,t}$ being of dimension $k_z \times 1$, represent vectors of explanatory variables; $\mu_i$ and $\epsilon_{i,t}$ are the fixed effects and the random errors respectively; $\lambda_{i,j}$, $\gamma_{i,j}$, and $\alpha_{i,j}$ denote the coefficients of the lagged explanatory variables; and $\theta$ represents the regression coefficients. Conditions imposed on the dynamics of the error process and of the covariates in the d.g.p. defined by (3.1) are summarized in Assumption 3.1.

**Assumption 3.1.** The innovations $\epsilon_i$ are orthogonal to both $W_i$ and $X_i$. 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, $\{\zeta_{i,t}, \ i = 1, \ldots, N, t = 1, \ldots, T\}$, which are independent across the individuals and stationary, are strongly mixing across the time periods with the mixing coefficient satisfying the condition stated in Lemma 2. The same assumptions about $X_{i,t}$ are also imposed on $Z_{i,t} = \sum_{s=1}^{t} \xi_{i,s}$.

In addition, Assumption 3.2 below allows (3.1) to have a long-run relationship, $y_{i,t} = \theta^\top W_{i,t} + \nu_{i,t}$, where $\nu_{i,t}$ is a stationary process.

**Assumption 3.2.** 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 $W_{i,t}$ in (3.1). Let’s denote by $\varphi = (\theta^\top, \phi^\top, \sigma^\top)^\top$, where $\phi = (\phi_1, \ldots, \phi_N)^\top$ and $\sigma = (\sigma_1^2, \ldots, \sigma_N^2)^\top$, 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, $\Delta y_{i,-j} = \cdots$
\((\Delta y_{i,p^* - j}, \ldots, \Delta y_{i,T - j})^\top\) and \(\Delta \mathbf{X}_{i,-j} = (\Delta X_{i,p^* - j}, \ldots, \Delta X_{i,T - j})^\top\), a \((T - p^*) \times k_z\) matrix, \(\Delta \mathbf{Z}_{i,-j} = (\Delta Z_{i,p^* - j}, \ldots, \Delta Z_{i,T - j})^\top\), and a \((T - p^*) \times 1\) vector of random errors, \(\mathbf{e}_i = (\epsilon_{i,p^*}, \ldots, \epsilon_{i,T})^\top\). Moreover, define

\[
\mathbf{U}_i = (\Delta \mathbf{y}_{i,-1}, \ldots, \Delta \mathbf{y}_{i,-p+1}, \Delta \mathbf{X}_i, \ldots, \Delta \mathbf{X}_{i,q_x+1}, \Delta \mathbf{Z}_i, \ldots, \Delta \mathbf{Z}_{i,q_x+1}, \mathbf{u}_T),
\]

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

\[
\Lambda_i = (\lambda_{i,1}, \ldots, \lambda_{i,p-1}, \gamma_{i,0}, \ldots, \gamma_{i,q_x-1}, \alpha_{i,0}^\top, \ldots, \alpha_{i,q_x-1}^\top, \mu_i)^\top.
\]

Note at this point that \(\mathbf{U}_i\) is of dimension \((T - p^*) \times k_u\) with \(k_u = p + q_x + q_zk_z\), and \(\Lambda_i\) is of dimension \(k_u \times 1\). One can now write (3.1) into the following matrix form:

\[
\mathbf{y}_i = \phi_i(\mathbf{y}_{i,-1} - \mathbf{W}_i \mathbf{\theta}) + \mathbf{U}_i \Lambda_i + \epsilon_i,
\]

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

\[
L_T(\varphi) = -\frac{T}{2} \sum_{i=1}^{N} \ln(2\pi \sigma_i^2) - \frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (\Delta \mathbf{y}_i - \phi_i \xi_i(\mathbf{\theta}))^\top \mathbf{P}_i (\Delta \mathbf{y}_i - \phi_i \xi_i(\mathbf{\theta})),
\]

where \(\xi_i(\mathbf{\theta}) = \mathbf{y}_{i,-1} - \mathbf{W}_i \mathbf{\theta}\) and \(\mathbf{P}_i = \mathbf{I}_T - \mathbf{U}_i (\mathbf{U}_i^\top \mathbf{U}_i)^{-1} \mathbf{U}_i^\top\) with \(\mathbf{I}_T\) being the \((T - p^*) \times (T - p^*)\) identity matrix. Note that by using the orthogonality of \(\mathbf{P}_i\) to \(\mathbf{U}_i\), one obtains for the true parameter \(\varphi_0\),

\[
\frac{1}{T} (L_T(\varphi_0) - L_T(\varphi)) = \frac{1}{2} \sum_{i=1}^{N} \frac{\sigma_{0,i}^2 - \sigma_i^2}{\sigma_i^2} \left(\frac{\epsilon_i^\top \mathbf{P}_i \epsilon_i}{T} - \frac{\sigma_{0,i}^2}{\sigma_i^2}\right) + \frac{1}{2} \sum_{i=1}^{N} \left(\frac{\sigma_{0,i}^2}{\sigma_i^2} - \ln\left(\frac{\sigma_{0,i}^2}{\sigma_i^2}\right) - 1\right)
\]

\[+
\frac{1}{2T} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} ((\Delta \mathbf{y}_i - \phi_i \xi_i(\mathbf{\theta}))^\top \mathbf{P}_i (\Delta \mathbf{y}_i - \phi_i \xi_i(\mathbf{\theta}))) - \epsilon_i^\top \mathbf{P}_i \epsilon_i
\]

\[= \frac{1}{2} \left(\mathcal{T}_1(\sigma, \sigma_0) + \mathcal{T}_2(\sigma, \sigma_0) + \mathcal{T}_3(\varphi)\right). \tag{3.2}\]
Since $\Delta y_i - \phi_i \xi_i(\theta) = U_i A_{0,i} + \epsilon_i + \phi_i W_i (\theta_0 - \theta) + (\phi_i - \phi_{0,i}) \xi_i(\theta_0)$, one can also write

$$T_{3,T}(\varphi) = (\Gamma - \Gamma_0)^T G_T (\Gamma - \Gamma_0) + 2(\Gamma - \Gamma_0)^T F_T,$$

where $\Gamma = (\theta^T, \phi^T)^T$ and $G_T = G_T(\phi, \sigma) = \begin{pmatrix} \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2} W_i^T P_i W_i & -\frac{\phi_i}{\sigma_i^2} W_i^T P_{i,0,1} & \cdots & -\frac{\phi_i}{\sigma_i^2} W_i^T P_{i,N,0,N} \\ -\frac{\phi_i}{\sigma_i^2} W_i^T P_{i,0,1} & \frac{\xi_{i,0,1}^T P_{i,0,1}}{\sigma_i} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ -\frac{\phi_i}{\sigma_i^2} W_i^T P_{i,N,0,N} & 0 & \cdots & \frac{\xi_{i,N}^T P_{i,N,0,N}}{\sigma_i} \end{pmatrix}$,

where $\xi_{0,i} = \xi_i(\theta_0)$ for $i = 1, \ldots, N$, and $F_T = \begin{pmatrix} \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2} W_i^T P_{i,i} \\ -\frac{1}{\sigma_i^2} \xi_{i,0,1}^T P_{i,0,1} \\ \vdots \\ -\frac{1}{\sigma_i^2} \xi_{i,N}^T P_{i,N,N} \end{pmatrix}$. To work out the probability limits for the random matrices $G_T$ and $F_T$, we need to state the following lemma:

**Lemma 1.** Suppose that Assumptions 3.1 and 3.2 hold. Let's denote by

$$D_{ww,T} = \text{diag}(T^{1/2} I_{k_1}, T^{1/2}, \ldots, T^{k_2/2}),$$

where $I_{k_1}$ is the $k_1 \times 1$ unit vector, the diagonal matrix of normalizing factors. Then,

$$D_{ww,T}^{-1} W_i^T P_{i} W_i D_{ww,T}^{-1} \xrightarrow{p} Q_{ww,i}, \quad (3.3)$$

$$D_{ww,T}^{-1} W_i^T P_{i} \xi_{0,i} D_{ww,T}^{-1} \xrightarrow{p} Q_{w\xi,i}, \quad (3.4)$$

$$D_{ww,T}^{-1} \xi_{0,i}^T P_{i} \xi_{0,i} D_{ww,T}^{-1} \xrightarrow{p} Q_{\xi\xi,i}, \quad (3.5)$$

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

**Theorem 2 (Consistency).** Suppose that Assumptions 3.1 and 3.2 hold. Let $D_{G,T} = \text{diag}(D_{ww,T}, I_N)$,
where $I_N$ is the $N \times N$ identity matrix, and
\[
Q_G = Q_G(\phi, \sigma) = 
\begin{pmatrix}
\sum_{i=1}^{N} \phi_i^2 Q_{w_{i,1}} & -\frac{\phi_1}{\sigma_1} Q_{w_{1,1}} & \cdots & -\frac{\phi_N}{\sigma_N} Q_{w_{N,1}} \\
-\frac{\phi_1}{\sigma_1} Q_{w_{1,1}} & \frac{1}{\sigma_1} Q_{w_{1,1}} & 0 & 0 \\
\vdots & 0 & \ddots & 0 \\
-\frac{\phi_N}{\sigma_N} Q_{w_{N,1}} & 0 & 0 & \frac{1}{\sigma_N} Q_{w_{N,N}}
\end{pmatrix}.
\]

Moreover, suppose that $E[|\epsilon_{i,t}|^{2+\delta}] < \infty$ for some $\delta > 0$, and $\inf_{\phi, \sigma} \lambda_1(Q_G) > 0$ a.s. Then,
\[
D_{ww,T}(\hat{\theta} - \theta_0) = o_p(1), \quad (\hat{\phi} - \phi_0) = o_p(1), \quad \text{and} \quad (\hat{\sigma} - \sigma_0) = o_p(1).
\]

**Theorem 3 (Asymptotic Mixed Normality).** Suppose that Assumptions 3.1 and 3.2 hold. Moreover, presume that $\lambda_1(Q_G(\phi_0, \sigma_0)) > 0$ a.s. and $E[|\epsilon_{i,t}|^{2+\delta}] < \infty$ for some $\delta > 0$. Define
\[
D_T = \text{diag}(T^{1/2} D_{ww,T}, T^{1/2} I_N) = \text{diag}(T \ell_k^\top, T, \ldots, T^{\ell_{i+1}}, \ldots, T^{\ell_{k_x+1}}, T^2 \ell_N^\top).
\]

Then,
\[
D_T(\hat{\Gamma} - \Gamma_0) \xrightarrow{d} \text{MN}(0, Q_G^{-1}(\phi_0, \sigma_0)).
\]

**Corollary 4.** Fieller’s confidence sets for the ratios $\delta_i = \frac{L_i^\top D_T \Gamma_0}{K^\top D_T \Gamma_0}$ for $i = 1, \ldots, m$, where $L_i$ and $K$ are some given column vectors, can be constructed by inverting a Wald’s statistics for testing $H_0 : L_i^\top D_T \Gamma_0 - \delta_i K^\top D_T \Gamma_0 = 0$ vs. $H_1 : L_i^\top D_T \Gamma_0 - \delta_i K^\top D_T \Gamma_0 \neq 0$. Theorem 3 suggests that this Wald’s statistics is given by
\[
W_T(\delta) = \hat{\Gamma}^\top D_T (L - \delta K_\delta) \left( (L^\top - \delta K_\delta^\top) (D_T^{-1} \hat{G}_T D_T^{-1}) (L - \delta K_\delta) \right)^{-1} \xrightarrow{d} \chi^2(m),
\]
where
\[
\hat{G}_T = \left( \sum_{i=1}^{N} \begin{pmatrix} \hat{\phi}_i W_i^\top P_i W_i & -\frac{1}{\hat{\phi}_i^2} W_i^\top P_i \hat{\xi}_i \\ -\frac{1}{\hat{\phi}_i^2} W_i^\top P_i \hat{\xi}_i & 0 \\ \vdots & \ddots & \ddots & \ddots \\ -\frac{1}{\hat{\phi}_N^2} W_N^\top P_N \hat{\xi}_N & 0 & 0 & \frac{1}{\hat{\phi}_N^2} \hat{\xi}_N P_N \hat{\xi}_N \end{pmatrix} \right) \text{ with } \hat{\xi}_i = y_{i,-1} - W_i \hat{\theta} \text{ for } i = 1, \ldots, N;
\]

\(D_T^{-1} \hat{G}_T D_T^{-1}\) is the estimate of \(Q_G(\phi_0, \sigma_0)\); \(\delta = \text{diag}(\delta_1, \ldots, \delta_m)\); \(L = (L_1, \ldots, L_m)\) and \(K_\delta = \iota_m \otimes K\) are \(k_{w n} \times m\) matrices.

4 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:

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

\(t = 1, \ldots, T, \quad i = 1, \ldots, N\)

\(W_{i,t} = \begin{pmatrix} Z_{i,t}^\top \end{pmatrix}_{k_x \times 1} \begin{pmatrix} X_{i,t} \end{pmatrix}_{k_{x} \times 1} \begin{pmatrix} X_{i,t-1}^2, \ldots, X_{i,t-1}^{k_z} \end{pmatrix}_{k_z}^\top \text{ polynomial}
\]

with \(\theta\) invariant across \(i\), \(y_{i,t}\), \(X_{i,t}\) and \(Z_{i,t}\) have unit roots and the long-run relation is stationary.

We consider that data generating process (DGP) is a finite-order \(ARDL(1,0)\) process as in Pesaran and Shin (1999), where the above model includes the ECM, a quadratic polynomial and 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 \((\epsilon_{it}, \eta_{it})\) are serially correlated and are generated according to the following bivariate
normal distribution:

\[
\begin{pmatrix}
\epsilon \\
\mu
\end{pmatrix} \sim N(0, \Omega)
\]

with

\[
\Omega = \begin{pmatrix}
1 & \omega_{12} \\
\omega_{12} & 1
\end{pmatrix}.
\]

The parameters \(\theta\) comprise of \(\beta_0\) (constant), \(\beta_1\) (of \(X_{it}\)), \(\beta_2\) (of \(X_{it}^2\)) and the covariance \(\omega_{12}\), 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 \(\beta_2\) was weakly identified. In the simulations we keep \(\beta_0, \beta_1, \beta_2\) and \(\omega_{12}\) fixed and we play with the degree of persistency for the \(y_{it}\) and \(X_{it}\) by changing the parameters \(\phi, \rho\) and \(\psi\). 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 \(X_{it}\) and high persistence of \(y_{it}\). 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.

5 Conclusion

When ratios of parameters are estimated and tested, it is important to obtain reliable confidence bounds especially when 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 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


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


Appendix A  Known Results

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

**Lemma 2.** 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 \( F^b_a \) represent the \( \sigma \)-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 \( A \in F^k_1 \), and \( B \in F_{k+n}^\infty \) 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 \( \Gamma \). 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 \( \Gamma \) such that

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

for some \( \lambda > 0 \) depending on \( \epsilon, \delta, \) and \( d \) only.

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

Appendix B  Proofs of Auxiliary Lemmas

**Proof of Lemma 1.** First, note that, in view of Assumptions 3.1 and 3.2, an application of Lemma 2 yields

\[
\Delta y_i^\top \Delta y_i = \sum_{t=1}^T \Delta y_{i,t}^2 \ll T \text{ w.p.},
\]

\[
U_i^\top U_i \ll T \mathbf{u}_{k_u} \mathbf{u}_{k_u}^\top \text{ w.p., where } \mathbf{u}_{k_u} \text{ is the } k_u \times 1 \text{ unit vector},
\]

\[
\Delta y_i^\top Z_i = \sum_{t=1}^T Z_{i,t} \Delta y_{i,t} \approx \int_0^1 Z_{i,[T\tau]}(y_{i,[T(\tau+dr)]} - y_{i,[T\tau]}) \ll T \mathbf{u}_{k_z} \text{ w.p.}
\]
as $y_{i,[T\tau]} - y_{i,[T(\tau + d\tau)]}$ can be approximated by a Brownian motion, $dW([T\tau]) = W([T(\tau + d\tau)]) - W([T\tau])$. And by the same argument, one also obtains

$$
\Delta X_i^\top Z_i \ll T k_z \text{ w.p.,}
$$

$$
\Delta Z_i^\top Z_i \ll T k_z \top qk_z \text{ w.p.,}
$$

$$
Z_i^\top i = \sum_{t=1}^{T} Z_{i,t} \ll T^{3/2} k_z \text{ w.p.,}
$$

$$
(X_i^\ell)^\top \Delta y_i = \sum_{t=1}^{T} X_{i,t}^\ell \Delta y_{i,t} \approx \int_0^1 X_{i,[T\tau]}^\ell (y_{i,[T(\tau + d\tau)]} - y_{i,[T\tau]}) \ll T^{\ell+1} \text{ w.p.,}
$$

$$
(X_i^\ell)^\top \iota_T \ll T^{\ell+2} \text{ w.p.,}
$$

$$
(X_i^\ell)^\top U_i \ll T^{\ell+1} k_u - 1, T^{\ell+2} 2 \text{ w.p.}
$$

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

$$
Z_i^\top U_i \ll (T k_z \times (k_u - 1), T^{3/2} k_z) \text{ w.p.,}
$$

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

$$
D_{uu,T}^{-1} W_i^\top U_i (U_i^\top U_i)^{-1} U_i^\top W_i D_{uu,T}^{-1} \xrightarrow{p} Q_{uu,i}^{(1)}
$$

$$
D_{uu,T}^{-1} W_i^\top W_i D_{uu,T}^{-1} \xrightarrow{p} Q_{uu,i}^{(2)}
$$
Hence, (3.3) immediately follows. In addition, note that
\[
Z_i^\top \xi_{0,i} \ll \sum_{t=1}^{T} t^{1/2} \approx T^{3/2} \text{ w.p.,}
\]
\[
(X_i^t)^\top \xi_{0,i} \ll \sum_{t=1}^{T} t^{\ell/2} = T^{\ell+2}, \text{ w.p.}
\]
\[
Z_i^\top \xi_{0,i} \ll T^{3/2} \text{ w.p.,}
\]
\[
(X_i^t)^\top \xi_{0,i} \ll T^{\ell+2} \text{ w.p.}
\]

One can immediately show (3.4) and (3.5). 

\[\square\]

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: 
\[B_T(\theta_0, \delta_\theta) = \{\theta \in \Theta_\theta \subset \mathbb{R}^{k_w} : \|D_{w,T}(\theta - \theta_0)\| < \delta_\theta\}, \]
where 
\[k_w = k_z + k_x \text{ and } \Theta_\theta \text{ is some compact parameter space of } \theta_0; \]
\[B(\phi_0, \delta_\phi) = \{\phi \in \Theta_\phi \subset \mathbb{R}^N : \|\phi - \phi_0\| < \delta_\phi\}, \]
where 
\[\Theta_\phi \text{ is some compact parameter space of } \phi_0; \]
\[B(\sigma_0, \delta_\sigma) = \{\sigma \in \Theta_\sigma \subset \mathbb{R}^N : \|\sigma - \sigma_0\| < \delta_\sigma\}, \]
where 
\[\Theta_\sigma \text{ is some compact parameter space of } \sigma_0. \]

Let \(B^c_T(\theta_0, \delta_\theta), B^c(\phi_0, \delta_\phi), \) and \(B^c(\sigma_0, \delta_\sigma)\) be the complements of \(B_T(\theta_0, \delta_\theta), B(\phi_0, \delta_\phi), \) and \(B(\sigma_0, \delta_\sigma)\) respectively. Define 
\[B_T(\varphi, \delta, \delta_\sigma) = \bigcup_{\delta_\theta, \delta_\phi : (\delta_\theta^2 + \delta_\phi^2)^{1/2} = \delta} \left\{B^c_T(\theta_0, \delta_\theta) \times B^c(\phi_0, \delta_\phi)\right\} \times B^c(\sigma_0, \delta_\sigma). \]

We need to show that 
\[
\lim_{T \to \infty} P \left( \inf_{\varphi \in B_T(\varphi, \delta, \delta_\sigma)} \frac{1}{T} (L_T(\varphi_0) - L_T(\varphi)) > 0 \right) = 1 \tag{C-1}
\]
for every \(\delta, \delta_\sigma > 0. \) In view of (3.2), one obtains
\[
\inf_{\varphi \in B_T(\varphi, \delta, \delta_\sigma)} \frac{1}{T} (L_T(\varphi_0) - L_T(\varphi)) \geq \frac{1}{2} \left\{ \inf_{\sigma \in B^c(\sigma_0, \delta_\sigma)} T_1(\sigma, \sigma_0) + \inf_{\sigma \in B^c(\sigma_0, \delta_\sigma)} T_2(\sigma, \sigma_0) + \inf_{\varphi \in B_T(\varphi, \delta, \delta_\sigma)} T_3(\varphi) \right\}. \]
It can immediately be shown that \( \inf_{\sigma \in B(\sigma_0, \delta_\sigma)} T_{1,T}(\sigma, \sigma_0) = o_p(1) \) and \( \inf_{\sigma \in B(\sigma_0, \delta_\sigma)} T_{2,T}(\sigma, \sigma_0) > 0 \). Furthermore,

\[
\inf_{\varphi \in B_T(\varphi, \delta, \delta_\varphi)} T_{3,T}(\varphi) = \inf_{\varphi \in B_T(\varphi, \delta, \delta_\varphi)} (\Gamma - \Gamma_0)^\top G_T (\Gamma - \Gamma_0) + 2 \inf_{\varphi \in B_T(\varphi, \delta, \delta_\varphi)} (\Gamma - \Gamma_0)^\top F_T = T_{3,a,T}(\varphi) + 2 T_{3,b,T}(\varphi).
\]

Note that, by an elementary matrix inequality and Lemma 1,

\[
T_{3,a,T}(\varphi) = \inf_{\varphi \in B_T(\varphi, \delta, \delta_\varphi)} (\Gamma - \Gamma_0)^\top D_{G,T}^{-1} D_{G,T} G_T (\Gamma - \Gamma_0) D_{G,T}^{-1} (\Gamma - \Gamma_0) \geq \inf_{\varphi \in B_T(\varphi, \delta, \delta_\varphi)} \| D_{G,T} (\Gamma - \Gamma_0) \|_2^2 \inf_{\varphi \in B_T(\varphi, \delta, \delta_\varphi)} \| D_{G,T}^{-1} G_T D_{G,T}^{-1} \|_2 \geq \delta^2 \inf_{\varphi \in B(\varphi_0, \delta_\varphi)} \lambda_1 (Q_G) \text{ w.p.}
\]

It then follows that \( T_{3,a,T}(\varphi) > 0 \) w.p. Moreover, by Lemma 1, one has

\[
D_{G,T}^{-1} W_i P_i E[\epsilon_i^\top \epsilon] P_i W_i D_{G,T}^{-1} T \xrightarrow{p} \sigma_i^2 Q_{ww,i}, \\
\frac{\xi_{0,i}^\top P_i E[\epsilon_i^\top \epsilon] P_i \xi_{0,i}}{T} \xrightarrow{p} \sigma_i^2 Q_{xx,i}.
\]

Conditioning \( F_T \) on \( W_i, P_i \) and \( \xi_i \), an application of the multivariate CLT to the sequence \( \epsilon_i \) yields \( D_{G,T}^{-1} F_T = O_p \left( T^{-1/2} \right) = o_p(1) \). Since, from the way \( B_T(\theta_0, \delta_\theta) \) is defined, the term \( \inf_{\varphi \in B_T(\varphi, \delta, \delta_\varphi)} (\Gamma - \Gamma_0)^\top D_{G,T}^{-1} T_{e_{k=+1}} N \) is bounded either above or below by a generic constant, which can be large but does not depend on \( T \), it immediately follows that \( T_{3,b,T} = o_p(1) \). Therefore, (C-1) has been verified.

**Proof of Theorem 3.** The gradient and Hessian matrices of \( L_T(\varphi) \) are given by

\[
\frac{\partial L_T(\varphi)}{\partial \Gamma} = \left( \frac{\partial L_T(\varphi)}{\partial \theta}, \frac{\partial L_T(\varphi)}{\partial \phi} \right)^\top
\]
and $\frac{\partial^2 L_T(\phi)}{\partial \Gamma^\top \partial \Gamma}$

\[
\left(\begin{array}{cc}
\frac{\partial^2 L_T(\phi)}{\partial \theta \partial \theta^\top} & \frac{\partial^2 L_T(\phi)}{\partial \Gamma \partial \theta^\top} \\
\frac{\partial^2 L_T(\phi)}{\partial \theta \partial \phi^\top} & \frac{\partial^2 L_T(\phi)}{\partial \phi \partial \phi^\top}
\end{array}\right),
\]

where

\[
\begin{aligned}
\frac{\partial L_T(\phi)}{\partial \phi_i} &= \frac{1}{\sigma_i^2} \xi_i(\theta)^\top P_i(\Delta y_i - \phi_i \xi_i(\theta)), \\
\frac{\partial L_T(\phi)}{\partial \theta} &= - \sum_{i=1}^N \frac{\phi_i}{\sigma_i^2} W_i^\top P_i(\Delta y_i - \phi_i \xi_i(\theta)), \\
\frac{\partial^2 L_T(\phi)}{\partial \phi_i \partial \phi_j} &= 0 \text{ for } i \neq j, \\
\frac{\partial^2 L_T(\phi)}{\partial \theta \partial \theta^\top} &= - \sum_{i=1}^N \frac{\phi_i^2}{\sigma_i^2} W_i^\top P_i W_i, \\
\frac{\partial^2 L_T(\phi)}{\partial \theta \partial \phi_i} &= - \frac{1}{\sigma_i^2} W_i^\top P_i (\Delta y_i - \phi_i \xi_i(\theta)) + \frac{\phi_i}{\sigma_i^2} W_i^\top P_i \xi_i(\theta).
\end{aligned}
\]

Since $\hat{\phi}$ is consistent by Theorem 2, an application of a first-order Taylor expansion of $\frac{\partial L_T(\hat{\phi})}{\partial \Gamma}$ about $\Gamma_0$ yields

\[
0 = \frac{\partial L_T(\hat{\phi})}{\partial \Gamma} = \frac{\partial L_T(\Gamma_0, \hat{\sigma})}{\partial \Gamma} + \frac{\partial^2 L_T(\Gamma_\star, \hat{\sigma})}{\partial \Gamma \partial \Gamma^\top}(\hat{\Gamma} - \Gamma_0),
\]

where $\Gamma_\star$ is some point lying on the line segment $L(\Gamma_0, \hat{\Gamma}) = \{s\Gamma_0 + (1 - s)\hat{\Gamma} : s \in (0, 1)\} \subset \Theta_\theta \times \Theta_\phi \subset \mathbb{R}^{k_{\text{wn}}}$, where $\Theta_\theta \times \Theta_\phi$ are the compact parameter spaces of $\Gamma_0$ (as defined in the proof of Theorem 2), and $k_{\text{wn}} = k_w + N$. One can then obtain

\[
D_T(\hat{\Gamma} - \Gamma_0) = - \left[D_T^{-1} \frac{\partial^2 L_T(\Gamma_\star, \hat{\sigma})}{\partial \Gamma \partial \Gamma^\top} D_T^{-1}\right]^{-1} D_T^{-1} \frac{\partial L_T(\Gamma_0, \hat{\sigma})}{\partial \Gamma}.
\]

For notational brevity, let $I_T(\Gamma^\star, \hat{\sigma}) = D_T^{-1} \frac{\partial^2 L_T(\Gamma^\star, \hat{\sigma})}{\partial \Gamma \partial \Gamma^\top} D_T^{-1}$. First, one needs to show that

\[
\lim_{T \uparrow \infty} P \left(\|I_T(\Gamma^\star, \hat{\sigma}) - I_T(\Gamma_0, \sigma_0)\| > \epsilon\right) = 0 \text{ given some arbitrarily small } \epsilon > 0.
\]
Note that

\[ P \left( \| \mathcal{I}_T(\Gamma^*, \hat{\sigma}) - \mathcal{I}_T(\Gamma_0, \sigma_0) \| > \epsilon \right) \]

\[ = P \left( \| \mathcal{I}_T(\Gamma^*, \hat{\sigma}) - \mathcal{I}_T(\Gamma_0, \sigma_0) \| > \epsilon \mid \Gamma^* \in B_T(\theta_0, \delta_\theta) \times B(\phi_0, \delta_\phi), \hat{\sigma} \in B(\sigma_0, \delta_\sigma) \right) \]

\[ + P \left( \| \mathcal{I}_T(\Gamma^*, \hat{\sigma}) - \mathcal{I}_T(\Gamma_0, \sigma_0) \| > \epsilon \mid \Gamma^* \in B_T^c(\theta_0, \delta_\theta) \times B^c(\phi_0, \delta_\phi), \hat{\sigma} \in B^c(\sigma_0, \delta_\sigma) \right) \]

where the balls \( B_T(\theta_0, \delta_\theta), B(\phi_0, \delta_\phi), \) and \( B(\sigma_0, \delta_\sigma) \) are defined in the proof of Theorem 2. Since \( \lim_{T \to \infty} P \left( \Gamma^* \in B_T^c(\theta_0, \delta_\theta) \times B^c(\phi_0, \delta_\phi), \hat{\sigma} \in B^c(\sigma_0, \delta_\sigma) \right) = 0 \) for every \( \Gamma^* \) lying on the line segment \( L(\Gamma_0, \hat{\Gamma}) \) by Theorem 2, one has

\[ \lim_{T \to \infty} P \left( \| \mathcal{I}_T(\Gamma^*, \hat{\sigma}) - \mathcal{I}_T(\Gamma_0, \sigma_0) \| > \epsilon \right) \leq \lim_{T \to \infty} P \left( \sup_{\Gamma \in B_T(\theta_0, \delta_\theta) \times B(\phi_0, \delta_\phi), \sigma \in B(\sigma_0, \delta_\sigma)} \| \mathcal{I}_T(\Gamma, \sigma) - \mathcal{I}_T(\Gamma_0, \sigma_0) \| > \epsilon \right) \]

for some arbitrarily small numbers, \( \delta_\theta, \delta_\phi \) and \( \delta_\sigma \). An application of Lemma 1 and some inequalities for matrices yields

\[ \sup_{\phi \in B(\phi_0, \delta_\phi), \sigma \in B(\sigma_0, \delta_\sigma)} \left\| D_{\gamma, k_w}^{-1} \left( \frac{\partial^2 L_T(\varphi)}{\partial \theta \partial \theta} - \frac{\partial^2 L_T(\varphi_0)}{\partial \theta \partial \theta} \right) D_{\gamma, k_w}^{-1} \right\| \leq C_0 \left( \delta_\phi^2 + \delta_\sigma^2 \right)^{\frac{1}{2}} \sum_{i=1}^{N} \| Q_{w w, i} \|, \]

\[ \sup_{\phi \in B(\phi_0, \delta_\phi), \sigma \in B(\sigma_0, \delta_\sigma)} \left\| D_{\gamma, k_w}^{-\frac{1}{2}} \left( \frac{\partial^2 L_T(\varphi)}{\partial \theta \partial \phi} - \frac{\partial^2 L_T(\varphi_0)}{\partial \theta \partial \phi} \right) \right\| I_{N T}^{1/2} \| \leq C_0 \left( \delta_\phi \sum_{i=1}^{N} \| Q_{w w, i} \| + \left( \delta_\phi^2 + \delta_\sigma^2 \right)^{\frac{1}{2}} \sum_{i=1}^{N} \| Q_{w \xi, i} \| \right), \]

\[ \sup_{\phi \in B(\phi_0, \delta_\phi), \sigma \in B(\sigma_0, \delta_\sigma)} \left\| \frac{\partial^2 L_T(\varphi)}{\partial \phi \partial \phi} - \frac{\partial^2 L_T(\varphi_0)}{\partial \phi \partial \phi} \right\| \leq \delta_\sigma \left( \sum_{i=1}^{N} \| Q_{\xi, \xi, i} \| \right)^{1/2}, \]

where \( C_0 \) is some finite generic constant that may differ from a line to another one. An application
of the matrix inequality: \(\| A C C D \| \leq \| A \|_2 + \sqrt{2} \| C \|_2 + \| D \|_2\) yields
\[
\sup_{\Gamma \in B_T(\theta_0, \delta_\theta) \times B(\phi_0, \delta_\phi) \atop \sigma \in B(\sigma_0, \delta_\sigma)} \| \mathcal{I}_T(\Gamma, \sigma) - \mathcal{I}_T(\Gamma_0, \sigma_0) \| 
\leq C_0 \left( \delta_\theta + (\delta_\phi^2 + \delta_\sigma^2)^{1/2} \right) \sum_{i=1}^{N} \| Q_{ww,i} \| + \left( \delta_\phi^2 + \delta_\sigma^2 \right)^{1/2} \sum_{i=1}^{N} \| Q_{w\xi,i} \| + \delta_\sigma \sum_{i=1}^{N} \| Q_{\xi \xi,i} \| \right). \tag{C-5}
\]

The consistency of \(\hat{\Gamma}\) allows one to make \(\delta_\theta, \delta_\phi, \) and \(\delta_\sigma\) in (C-5) arbitrarily small such that its RHS becomes less than \(\epsilon.\) In view of (C-4), (C-3) has been proved. Therefore, \(\| \mathcal{I}_T(\Gamma^*, \sigma) - \mathcal{I}_T(\Gamma_0, \sigma_0) \| = o_p(1).\) By the same argument, one can also show that
\[
\left\| D_T^{-1} \left( \frac{\partial L_T(\Gamma_0, \hat{\sigma})}{\partial \Gamma} - \frac{\partial L_T(\varphi_0)}{\partial \Gamma} \right) \right\| = o_p(1).
\]
Moreover, by Lemma 1, one has
\[
\mathcal{I}_T(\Gamma_0, \sigma_0) \xrightarrow{p} Q_G(\phi_0, \sigma_0),
\]
where \(Q_G\) is given in Theorem 2. Now, notice that for each
\[
E_\epsilon \left[ D_T^{-1} \frac{\partial L_T(\varphi_0)}{\partial \Gamma} \frac{\partial L_T(\varphi_0)}{\partial \Gamma^\top} D_T^{-1} \right] = Q_G(\phi_0, \sigma_0),
\]
where the expectation is taken with respect to the joint probability density of \(\epsilon_i.\) Therefore, conditioning \(D_T^{-1} \frac{\partial L_T(\varphi_0)}{\partial \Gamma}\) on \(W_i, P_i,\) and \(\xi_i(\theta_0),\) an application of the multivariate CLT to the sequence \(\epsilon_i\) yields
\[
D_T^{-1} \frac{\partial L_T(\varphi_0)}{\partial \Gamma} \xrightarrow{d} N(0, Q_G(\phi_0, \sigma_0)).
\]
The main theorem then follows from (C-2) and some marginal integration. \(\square\)
Appendix D  Confidence Set for Ratios of two Parameters

[Tipping Points]

Consider the general model \((\mathcal{Y}, \{P_\theta : \theta \in \Theta\}), \Theta \subset \mathbb{R}^p, p \geq 1\), where \(\mathcal{Y}\) is the sample space and \(P_\theta\) is a probability distribution over \(\mathcal{Y}\) indexed by \(\theta = (\theta_1, \theta_2, ..., \theta_p)'\). Our object of interest are functions of \(\theta\) 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 asymptotically normal estimator of \(\theta\) is available \(\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p)'\) \(\sim N(\theta, \Sigma_{\theta})\) where \(\Sigma_{\theta}\) is estimated consistently by \(\hat{\Sigma}_{\theta}\). The discontinuity set \(\{\theta \in \Theta : K'\theta = 0\}\) is clearly non-empty. In this context, the delta method exploits the following regular asymptotic result:

\[
h(\hat{\theta}) \sim N\left( h(\theta), \frac{\partial h(\hat{\theta})}{\partial \theta} \Sigma_{\theta} \frac{\partial h'(\hat{\theta})}{\partial \theta} \right). \tag{D-1}
\]

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 \(\delta_0\) values. For the ratio case presented in Section 2, Fieller’s method involves assembling all \(\delta_0\) values such that \(\theta_1 - \delta_0 \theta_2 = 0\) is not rejected at the \(\alpha\)% using the \(t\)-statistic \(\left(\hat{\theta}_1 - \delta_0 \hat{\theta}_2\right) / (\hat{\sigma}_2^2 \hat{\sigma}_1 + \hat{\sigma}_1^2 \hat{\sigma}_1^2)^{1/2}\) which is asymptotically standard normal under the null hypothesis. The confidence set is thus defined as solution to following inequality in \(\delta_0\)

\[
FCS(\delta; \alpha) = \left\{ \delta_0 : \left(\hat{\theta}_1 - \delta_0 \hat{\theta}_2\right)^2 \leq \frac{z_{\alpha/2}^2}{\hat{\sigma}_2^2 \hat{\sigma}_1^2} \left( \hat{\sigma}_1^2 \hat{\sigma}_1^2 - 2 \delta_0 \hat{\sigma}_1 \hat{\sigma}_1^2 \right) \right\}. \tag{D-2}
\]

This requires solving the following second-degree-polynomial inequality for \(\delta_0\):

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

\[
A = \hat{\theta}_2^2 - \frac{z_{\alpha/2}^2 \hat{\sigma}_2^2}{\hat{\sigma}_1^2}, \quad B = -\hat{\theta}_1 \hat{\theta}_2 + \frac{z_{\alpha/2}^2 \hat{\sigma}_1 \hat{\sigma}_1^2}{\hat{\sigma}_1^2}, \quad C = \hat{\theta}_1^2 - \frac{z_{\alpha/2}^2 \hat{\sigma}_1^2}{\hat{\sigma}_1^2}. \tag{D-4}
\]
for real solutions $\delta_0$. 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

$$\text{FCS} (\delta; \alpha) = \begin{cases} 
[\delta_{01}, \delta_{02}] & \text{if } A > 0 \\
(\delta_{01}, \infty) & \text{if } A < 0 
\end{cases}.$$  \hspace{1cm} (D-5)

Bolduc, Khalaf and Yelou (2010) further show that: (i) if $\Delta < 0$, then $A < 0$ and $\text{FCS} (\delta; \alpha) = \mathbb{R}$; (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

<table>
<thead>
<tr>
<th>Parameters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.2</td>
<td>0.8</td>
<td>0.9</td>
<td>0.99</td>
<td>0.2</td>
<td>0.8</td>
<td>0.9</td>
<td>0.99</td>
</tr>
<tr>
<td>$\psi$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\omega_{12}$</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-0.679</td>
<td>-0.679</td>
<td>-0.679</td>
<td>-0.679</td>
<td>-0.679</td>
<td>-0.679</td>
<td>-0.679</td>
<td>-0.679</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.619</td>
<td>0.619</td>
<td>0.619</td>
<td>0.619</td>
<td>0.619</td>
<td>0.619</td>
<td>0.619</td>
<td>0.619</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.007</td>
</tr>
</tbody>
</table>

Appendix E Figures
Figure 1: Size $\rho=0.2$, $\phi = 1$, $\psi = 1$
Figure 2: Power $\rho=0.2, \phi = 1, \psi = 1$
Figure 3: Size $\rho=0.8$, $\phi = 1$, $\psi = 1$
Figure 4: Power $\rho=0.8$, $\phi = 1$, $\psi = 1$
Figure 5: Size $\rho=0.9, \phi = 1, \psi = 1$
Figure 6: Power $\rho=0.9, \phi = 1, \psi = 1$
Figure 7: Size $\rho = 0.2$, $\phi = 0.2$, $\psi = 1$
Figure 8: Power $\rho=0.2, \phi = 0.2, \psi = 1$
Figure 9: Size $\rho=0.8$, $\phi = 0.2$, $\psi = 1$
Figure 10: Power $\rho=0.8, \phi = 0.2, \psi = 1$
Figure 11: Size $\rho=0.9$, $\phi = 0.2$, $\psi = 1$
Figure 12: Power $\rho=0.9$, $\phi = 0.2$, $\psi = 1$