On the Behavior of the GMM estimator in Persistent Dynamic Panel Data Models with Unrestricted Initial Conditions

Kazuhiko Hayakawa
Department of Economics, Hiroshima University*

February 23, 2010

Abstract

This paper investigates the behavior of the first-difference GMM estimator for dynamic panel data models when persistency of data is (moderately) strong and initial conditions are unrestricted. We show that initial conditions affect the rate of convergence of the GMM estimator under a local to unity system where autoregressive parameter is modeled as $\alpha_N = 1 - c/N^p$ where N is the cross-sectional sample size and $0 < p \leq 1$. Specifically, we show that when initial conditions that renders the data to be mean-stationary are assumed, the GMM estimator is inconsistent when $1/2 \leq p \leq 1$ while $N^{1/2-p}$-consistent when $0 < p < 1/2$. If initial conditions that renders the data to be mean-nonstationary are assumed, the GMM estimator is shown to be $\sqrt{N}$-consistent for $0 < p \leq 1$. We also introduce a notion of "near mean-stationary" to take the closeness to mean-stationarity into consideration, and derive asymptotic distributions. A Monte Carlo simulation is conducted to assess the theoretical results.

*I acknowledge the financial support from the JSPS Fellowship and the Grant-in-Aid for Scientific Research (KAKENHI 20830056) provided by the JSPS.

1 Introduction

In empirical studies, the first-difference generalized method of moments (FD-GMM) estimator has been widely used to estimate dynamic panel data models since the works of Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991). However, when persistency of the data is strong, it has been recognized that the FD-GMM estimator may not work well due to the weak instruments problem, e.g. Blundell and Bond (1998) and Blundell, Bond and Windmeijer (2000). To overcome the weak instruments problem, Arellano and Bover (1995) and Blundell and Bond (1998) proposed a system GMM estimator which is obtained by imposing a restriction on initial conditions[1]. Since then, the system GMM estimator has been widely used in empirical studies since it is considered that the system GMM estimator is free from the weak instruments problem and more efficient that the FD-GMM estimator[2].

Although the system GMM estimator is widely used, its consistency depends on the form of initial conditions. Specifically, the consistency of the system GMM estimator is established only when initial conditions satisfies mean-stationarity, in other words, the data are mean-stationary in the sense that the mean of data does not depend on time t. However, mean-stationarity of the data is quite strong and may not hold in practice[3]. In such a case where the assumption of mean-stationarity is doubtful, the FD-GMM estimator is still consistent since its consistency does not depend on initial conditions. However, as explained, it is considered that the FD-GMM estimator suffers from the weak instruments problem when persistency of data is strong. Recently, Hayakawa (2009) shows that the finite sample behavior of the FD-GMM estimator in persistent dynamic panel models changes dramatically depending on the initial conditions; the FD-GMM estimator has very small bias in some cases, and has a large bias in other cases. The purpose of this paper is to theoretically demonstrate this behavior by using a local to unity approach.

There are several studies that employ a local to unity approach in dynamic panel setting. These include Hahn, Hausman and Kuersteiner (2007), Kruiniger (2009),

[1] Initial conditions have a significant effect on the behavior of estimators and tests in dynamic econometric models. For example, Müller and Elliott (2003), Elliott and Müller (2006), Harvey and Leybourne (2005, 2006), Harvey, Leybourne and Taylor (2009) discuss the problem of initial conditions associated with a unit root test in a time series context. In panel models, Harris, Harvey, Leybourne and Sakkas (2010) investigate the impact of initial conditions on the power of panel unit root test by Im, Pesaran and Shin (2003). For stable dynamic panel models, Anderson and Hsiao (1981, 1982), Hsiao (2003, Chap. 4), Arellano (2003b, pp. 96-107), Kiviet (2007) discuss the problems associated with initial conditions.

[2] Recently, Bun and Windmeijer (2010) show that the system GMM estimator does suffer from the weak instruments problem when the variance of individual effects are larger than that of disturbances.

[3] For empirical examples and results of mean-nonstationarity, see Arellano (2003b, p.96) and Arellano (2003a)

Madsen (2010) and Moon and Phillips (2000, 2004). The first three papers consider a parameter sequence such that as N, sample size of cross section, gets larger, the autoregressive parameter approaches one, while Moon and Phillips (2000, 2004) consider a parameter sequence such that as T, sample size of time series, gets larger, the autoregressive parameter approaches one. Since this paper focus on small T and large N panel data, the first approach is relevant to ours. However, in these three papers, the situation of Hahn, Hausman and Kuersteiner (2007) and that of Kruiniger (2009) and Madsen (2010) is different. Hahn, Hausman and Kuersteiner (2007) investigate the behavior of several instrumental variables and GMM estimators and proposed a long-difference IV estimator. They use a local to unity approach to approximate the behavior of these estimators near unit root and the unit root case is excluded. Contrary to Hahn, Hausman and Kuersteiner (2007), Kruiniger (2009) and Madsen (2010) consider both stable and unit root cases in a unified model. Therefore, they assume that individual effects vanish as the autoregressive parameter approaches one. Since this paper is concerned with the behavior of the FD-GMM estimator near unit root and unit root case is ruled out, among the papers above, Hahn, Hausman and Kuersteiner (2007) is most related to ours. The major difference between Hahn, Hausman and Kuersteiner (2007) and this paper is that we allow unrestricted initial conditions. Hence, this paper can be seen as an extension of Hahn, Hausman and Kuersteiner (2007) to allow for unrestricted initial conditions. Although this extension seems to be minor, we show that the assumption on initial conditions affects the rate of convergence. Specifically, under a local to unity system where an autoregressive parameter is given by $\alpha_N = 1 - c/N^p, (0 < p \leq 1)$, we show that the GMM estimator is $N^{1/2-p}$ consistent when $0 < p < 1/2$, while inconsistent when $1/2 \leq p \leq 1$ if data are mean-stationary. However, if data are mean-nonstationary, the GMM estimator is $\sqrt{N}$-consistent for $0 < p \leq 1$. Furthermore, to account for the closeness of the data to mean-stationarity, we introduce a notion of "near mean-stationarity" where data are deviated from mean-stationarity by $O(1/N^q), (0 < q \leq 1)$. Under a local to unity and near mean-stationary systems, we derive asymptotic distributions of the GMM estimator. Specifically, we show that depending on the values of p and q, the rate of convergence differs. When either p and q is smaller than 1/2, the GMM estimator is $N^{1/2-p}-$ or $N^{1/2-q}$ consistent and when both p and q are in the region $1/2 \leq p, q \leq 1$, the GMM estimator is inconsistent. Monte Carlo simulation shows that the derived asymptotic distributions explain the behavior of the GMM estimator well.

The rest of paper is organized as follows. In section 2, we introduce the model, assumptions and estimators. In section 3, the main results of this paper is provided. In section 4, we extend a model to the case where an exogenous variable is included and show that a similar result to section 3 could occur. In section 5, we conduct a

Monte Carlo simulation. Finally, section 6 concludes.

2 Setup

2.1 Model and assumptions

Let us consider the following model:

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}, \qquad (i = 1, ..., N;\ t = 1, ..., T) \tag{1}$$

where α is the parameter of interest with $|\alpha| < 1$ and η_i is unobserved individual effects. We make the following assumptions:

Assumption 1. *$v_{it} \sim iid(0, \sigma_v^2)$ over $i = 1, ..., N$ and $t = 1, ..., T$.*

Assumption 2. *$\eta_i \sim iid(\eta, \sigma_\eta^2)$ over $i = 1, ..., N$.*

Assumption 3. *For initial conditions, we assume*

$$y_{i0} = \delta\mu_i + \varepsilon_{i0} \tag{2}$$

where $\mu_i = \eta_i/(1-\alpha)$ and $\delta \neq 0$. For ε_{i0}, we assume that $\varepsilon_{i0} \sim iid(0, \sigma_0^2)$ over $i = 1, ..., N$ where $\sigma_0^2 = \lambda\sigma_v^2/(1-\alpha^2)$ with $\lambda > 0$. Further, assume that $E(\varepsilon_{i0}^4) = \kappa\sigma_0^4$.

Assumption 4. *v_{it}, η_i and ε_{i0} are mutually independent.*

Assumptions 1, 2, and 4 are widely used in previous studies, e.g. Alvarez and Arellano (2003). Assumption 3 plays a very important role in deriving the asymptotic distribution of the FD-GMM estimator. Hence, we discuss Assumption 3 in some detail below. We do not consider the case $\delta = 0$ since it is natural to consider that initial conditions contain individual effects[4].

With initial conditions (2), we have the following expression:

$$\begin{aligned} y_{it} &= \left[1-(1-\delta)\alpha^t\right]\mu_i + \sum_{j=0}^{t-1}\alpha^j v_{i,t-j} + \alpha^t\varepsilon_{i0} \\ &= h_t\mu_i + \xi_{it} + \alpha^t\varepsilon_{i0} \end{aligned} \tag{3}$$

where $h_t = 1-(1-\delta)\alpha^t$ and $\xi_{it} = \sum_{j=0}^{t-1}\alpha^j v_{i,t-j}$. Expectation and covariance of y_{it} given in (3) are given by

$$\begin{aligned} E(y_{it}) &= h_t\mu \\ cov(y_{is}, y_{it}) &= h_t h_s\sigma_\mu^2 + \sigma_v^2\alpha^{t-s}\left(\frac{1-(1-\lambda)\alpha^{2s}}{1-\alpha^2}\right), \quad (s \le t) \end{aligned}$$

where $\mu = E(\mu_i)$.

[4]When $\delta = 0$, a different result would be obtained in the next section.

By imposing restrictions on δ and λ, we have several schemes for data property. For example, if we set $\delta = 1, \lambda = 1$, then y_{it} is a covariance stationary process. If we set $\delta = 1$ and λ be unrestricted, $E(y_{it})$ does not depend on t; hence we call the case $\delta = 1$ as "mean-stationary". If we assume $\delta \neq 1$, y_{it} is mean-nonstationary since $E(y_{it})$ depends on time t[5]. Note that most of the previous studies including Hahn, Hausman and Kuersteiner (2007) assume $\delta = \lambda = 1$, while we do not assume $\delta = 1$ in this paper, i.e., δ can be both $\delta = 1$ and $\delta \neq 1$. Thus, assuming a general initial condition (2) is useful in characterizing the data property.

2.2 GMM estimator

We consider a first-differenced model:

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it} \qquad (i = 1, ..., N;\ t = 2, ..., T)$$

The moment conditions proposed by Arellano and Bond (1991) are given by

$$E(y_{is}\Delta v_{it}) = 0 \qquad (s = 0, ..., t-2;\ t = 2, ..., T) \tag{4}$$

Note that the validity of these moment conditions does not depend on initial conditions. The FD-GMM estimator is defined as

$$\widehat{\alpha} = \frac{\left(\sum_{i=1}^N \Delta\mathbf{y}'_{i,-1}\mathbf{Z}_i^{L2}\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{L2\prime}\mathbf{H}\mathbf{Z}_i^{L2}\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{L2\prime}\Delta\mathbf{y}_i\right)}{\left(\sum_{i=1}^N \Delta\mathbf{y}'_{i,-1}\mathbf{Z}_i^{L2}\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{L2\prime}\mathbf{H}\mathbf{Z}_i^{L2}\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{L2\prime}\Delta\mathbf{y}_{i,-1}\right)} \tag{5}$$

where $\Delta\mathbf{y}_i = (\Delta y_{i2}, ..., \Delta y_{iT})'$, $\Delta\mathbf{y}_{i,-1} = (\Delta y_{i1}, ..., \Delta y_{i,T-1})'$

$$\mathbf{Z}_i^{L2} = \begin{bmatrix} y_{i0} & & & & & & \\ & y_{i0} & y_{i1} & & & & \\ & & & \ddots & & & \\ & & & & y_{i0} & \cdots & y_{i,T-2} \end{bmatrix} \tag{6}$$

and

$$\mathbf{H} = \begin{bmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ 0 & & & -1 & 2 \end{bmatrix}$$

[5]We use a terminology "mean-nonstationary" to distinguish from "nonstationary" which is conventionally used to indicate an integrated variable.

3 Asymptotic Properties

In this section, we derive the asymptotic properties of the FD-GMM estimator when persistency of data is strong. To this end, let us consider the following parameter sequence[6]:

$$\alpha_N = 1 - \frac{c}{N^p} \tag{7}$$

where $c > 0$ is a constant and $0 < p \leq 1$. We do not consider the case $p = 0$, since it is straightforward to show the consistency and asymptotic normality by applying the general results on GMM. We use a local to unity approach to approximate the distribution near unit root, which is the same spirit as Hahn, Hausman and Kuersteiner (2007)[7] .

We first provide an intuitive reason why initial conditions affect the strength of instruments, which is provided by Hayakawa (2009), and then derive the order of magnitude of the F statistic in the first-stage regression in a two-stage least squares (2SLS) regression form.

3.1 Strength of instruments and initial conditions

The Jacobian of the moment conditions (4), which is derived in Appendix, is given by

$$-\frac{dE(y_{is}\Delta v_{it})}{d\alpha} = E(y_{is}\Delta y_{i,t-1}) = h_s \alpha^{t-2}(1-\delta)\frac{\sigma_\eta^2}{1-\alpha} - \frac{\sigma_v^2 \alpha^{t-s-2}[1-(1-\lambda)\alpha^{2s}]}{1+\alpha} \tag{8}$$

By noting that $E(y_{is}\Delta y_{it})$ measures the strength of instruments, we find that an additional correlation between y_{is} and Δy_{it}, the first term in (8), appears when $\delta \neq 1$. Hence, depending on the relative size of the first and second terms in (8), instruments becomes weak in some cases and becomes strong in other cases.

Formally, let us consider the strength of instruments in terms of concentration parameter which is substituted by the first-stage F statistic[8]. To simplify the discussion, let us consider the case $T = 2$. When $T = 2$, the two-stage least squares regression form can be written as

$$\Delta y_{i2} = \alpha \Delta y_{i1} + \Delta v_{i2} \tag{9}$$

$$\Delta y_{i1} = \pi y_{i0} + u_i \tag{10}$$

Using (22) and (26) in Appendix, it is easy to show that the OLS estimator of π is given by

$$\widehat{\pi} = \frac{N^{-1}\sum_{i=1}^N y_{i0}\Delta y_{i1}}{N^{-1}\sum_{i=1}^N y_{i0}^2} = \frac{(1-\delta)O_p(N^p) + O_p(1)}{O_p(N^{2p})} = (1-\delta)O_p\left(\frac{1}{N^p}\right) + Op\left(\frac{1}{N^{2p}}\right)$$

[6]In a time series model, a similar approach is used by Phillips and Magdalinos (2007a,b), Magdalinos and Phillips (2009), Giraitis and Phillips (2006), Kurozumi and Hayakawa (2009).

[7]Hahn, Hausman and Kuersteiner (2007) assume $p = 1$.

[8]See Stock, Wright and Yogo (2002).

From this, it is observed that whether $\delta = 1$ or not affects the speed toward zero as N grows. The first-stage F statistic that tests $H_0 : \pi = 0$ is given by

$$F = \frac{\widehat{\pi}^2 \left(\sum_{i=1}^{N} y_{i0}^2 \right)}{N^{-1} \sum_{i=1}^{N} (\Delta y_{i1} - \widehat{\pi} y_{i0})^2} = (1-\delta)^2 O(N) + (1-\delta) O(N^{1-p}) + O(N^{1-2p}) \quad (11)$$

This suggests that initial conditions and the degree of persistence affect the strength of instruments. When $\delta = 1$, it follows that $F \to \infty$ when $0 < p < 1/2$, $F = O(1)$ when $p = 1/2$ and $F \to 0$ when $1/2 < p \leq 1$ as $N \to \infty$. However, when $\delta \neq 1$, $F \to \infty$ regardless of the value of p as $N \to \infty$. This result indicates that when $\delta \neq 1$, the instruments are strong even if the data are persistent. Depending on p and δ, we consider the following four cases:

Case A(a): $\delta = 1$ and $0 < p < \frac{1}{2}$

Case A(b): $\delta = 1$ and $p = \frac{1}{2}$

Case A(c): $\delta = 1$ and $\frac{1}{2} < p \leq 1$

Case B: $\delta \neq 1$

Note that the data are mean-stationary in Case A while they are mean-nonstationary in case B. Furthermore, note that among the Cases A, persistency of data in Case A(a) is weakest while strongest in Case A(c). We do not need to separate cases for the Case B.

The following theorem provides the asymptotic distributions of the GMM estimator in Cases A and B.

Theorem 1. *Let Assumptions 1 to 4 hold and $N \to \infty$ with T fixed. Then, we have*

$$\text{Case A(a):} \quad N^{1/2-p} (\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{8\sigma_\eta^2}{c^2 \sigma_v^2 \lambda^2 \boldsymbol{\iota}_m' (\mathbf{JHJ}')^+ \boldsymbol{\iota}_m}\right)$$

$$\text{Case A(b):} \quad \widehat{\alpha} - \alpha \xrightarrow{d} \frac{(\boldsymbol{\xi} + \boldsymbol{\mu}_1)' (\mathbf{JHJ}')^+ \boldsymbol{\zeta}}{(\boldsymbol{\xi} + \boldsymbol{\mu}_1)' (\mathbf{JHJ}')^+ (\boldsymbol{\xi} + \boldsymbol{\mu}_1)},$$

$$\text{Case A(c):} \quad \widehat{\alpha} - \alpha \xrightarrow{d} \frac{\boldsymbol{\xi}' (\mathbf{JHJ}')^+ \boldsymbol{\zeta}}{\boldsymbol{\xi}' (\mathbf{JHJ}')^+ \boldsymbol{\xi}},$$

$$\text{Case B:} \quad \sqrt{N} (\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\sigma_v^2}{\sigma_\eta^2 \delta^2 \boldsymbol{\iota}_m' (\mathbf{JHJ}')^+ \boldsymbol{\iota}_m}\right)$$

where $\boldsymbol{\iota}_m$ is an $m = (T-1)T/2$ dimensional vector of ones, $\mathbf{A}^+$ denotes the Moore-Penrose inverse of $\mathbf{A}$, $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are zero mean random vectors defined by (34) in Appendix, $\boldsymbol{\mu}_1$ is defined in Lemma A3(c.1) in Appendix, and

$$\mathbf{J} = \begin{bmatrix} \iota_1 & & & \\ & \boldsymbol{\iota}_2' & & \\ & & \ddots & \\ & & & \boldsymbol{\iota}_{T-1}' \end{bmatrix}. \quad (12)$$

Remark 1. In case A, where data are mean-stationary, we find that the rate of convergence of the GMM estimator depends on p. In Case A(a), the GMM estimator is $N^{1/2-p}$-consistent, while in Cases A(b) and (c), it is inconsistent. This indicates that the GMM estimator behaves poorly when data are mean-stationary and persistency is strong. Also note that a similar discontinuity is obtained by Hahn and Kuersteiner (2002) and Caner (2010) in weak instruments literature.

Remark 2. In Case B, where data are mean-nonstationary, we find that the GMM estimator is $\sqrt{N}$-consistent for $0 < p \leq 1$. In other words, when data are mean-nonstationary, the GMM estimator performs well even when persistency of data is strong.

Remark 3. Comparing the asymptotic variance of Case A(a) and Case B, we find that the variance ratio σ_η^2/σ_v^2 has an opposite effect. When σ_η^2/σ_v^2 is large, the variance in Case A(a) becomes large while that in Case B becomes small.

3.2 Asymptotic results with "near mean-stationary" initial conditions

Theorem 1 shows that the degree of persistence and initial conditions affect the rate of convergence. However, when δ is not equal to one, but very close to one, say, $\delta = 1.0001$, the behavior of the GMM estimator might be close to the asymptotic distribution of Case A not that of Case B. One possible way to explain this is that the closeness between δ and 1 is not taken into consideration when δ is fixed. In other words, when δ is fixed, $\delta = 1.0001$ and $\delta = 10$ are not distinguished. To take the closeness of δ to 1 into account, we introduce the following sequence:

$$\delta_N = 1 - \frac{d}{N^q} \tag{13}$$

where $0 < q \leq 1$ and $d \neq 0$ is a constant[9]. In this sequence, q controls the closeness of data to mean-stationarity. When q is close to 0, δ_N is deviated from 1, which implies that data tend to be mean-nonstationary. As q gets larger, δ_N gets close to 1, implying that the data are near mean-stationary. Thus, with (13), the closeness between δ and 1 are taken into consideration through q.

With the sequence (13), the first-stage F statistic that tests $H_0 : \pi = 0$ given in (11) becomes

$$F = O_p(N^{1-2q}) + O_p(N^{1-p-q}) + O_p(N^{1-2p}).$$

Depending on the values of p and q, the order of magnitudes of F statistic differs. Totally eleven cases are summarized in Table 1. From the Table 1, we find that for

[9] We do not consider the case $q = 0$ since this case is equivalent to fixed δ.

Cases C1, C2, C3, C4 and C7, F statistic diverges while for Cases C5, C6, C8 and C9, F statistic is bounded as $N \to \infty$. This implies that when persistency of data is not so strong or when δ is sufficiently deviated from 1, the instruments becomes strong. However, as the persistency of data gets stronger or as the data get close to mean-stationary, the instruments become weak. The following theorem shows the asymptotic distribution of eleven cases in Table 1.

Table 1: Order of magnitudes of F statistic in first-stage regression

	$0 < p < \frac{1}{2}$	$p = \frac{1}{2}$	$\frac{1}{2} < p \leq 1$
$0 < q < \frac{1}{2}$	Case C1(a): $O_p(N^{1-2q})$ if $p > q$ Case C1(b): $O_p(N^{1-2r})$ if $p = q = r$ Case C1(c): $O_p(N^{1-2p})$ if $p < q$	Case C2: $O_p(N^{1-2q})$	Case 3: $O_p(N^{1-2q})$
$q = \frac{1}{2}$	Case C4: $O_p(N^{1-2p})$	Case C5: $O_p(1)$	Case C6: $O_p(1)$
$\frac{1}{2} < q \leq 1$	Case C7: $O_p(N^{1-2p})$	Case C8: $O_p(1)$	Case C9: $o_p(1)$

Theorem 2. *Let Assumptions hold and $N \to \infty$ with T fixed. Then, we have*

$$\text{Case C1(a), C2, C3:} \quad N^{1/2-q}\left(\widehat{\alpha}_{GMM} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\sigma_v^2}{d^2\sigma_\eta^2 \boldsymbol{\iota}_m'(\mathbf{JHJ}')^{+}\boldsymbol{\iota}_m}\right)$$

$$\text{Case C1(b):} \quad N^{1/2-r}\left(\widehat{\alpha}_{GMM} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{8\sigma_v^2\sigma_\eta^2}{\left[2d\sigma_\eta^2 - c\lambda^2\sigma_v^2\right]^2 \boldsymbol{\iota}_m'(\mathbf{JHJ}')^{+}\boldsymbol{\iota}_m}\right)$$

$$\text{Case C1(c), C4, C7:} \quad N^{1/2-p}\left(\widehat{\alpha}_{GMM} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{8\sigma_\eta^2}{c^2\lambda^2\sigma_v^2 \boldsymbol{\iota}_m'(\mathbf{JHJ}')^{+}\boldsymbol{\iota}_m}\right)$$

$$\text{Case C5:} \quad \widehat{\alpha}_{GMM} - \alpha \xrightarrow{d} \frac{(\boldsymbol{\xi} + \boldsymbol{\mu}_3)'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}\boldsymbol{\zeta}}{(\boldsymbol{\xi} + \boldsymbol{\mu}_3)'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}(\boldsymbol{\xi} + \boldsymbol{\mu}_3)},$$

$$\text{Case C6:} \quad \widehat{\alpha}_{GMM} - \alpha \xrightarrow{d} \frac{(\boldsymbol{\xi} + \boldsymbol{\mu}_2)'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}\boldsymbol{\zeta}}{(\boldsymbol{\xi} + \boldsymbol{\mu}_2)'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}(\boldsymbol{\xi} + \boldsymbol{\mu}_2)},$$

$$\text{Case C8:} \quad \widehat{\alpha}_{GMM} - \alpha \xrightarrow{d} \frac{(\boldsymbol{\xi} + \boldsymbol{\mu}_1)'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}\boldsymbol{\zeta}}{(\boldsymbol{\xi} + \boldsymbol{\mu}_1)'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}(\boldsymbol{\xi} + \boldsymbol{\mu}_1)},$$

$$\text{Case C9:} \quad \widehat{\alpha}_{GMM} - \alpha \xrightarrow{d} \frac{\boldsymbol{\xi}'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}\boldsymbol{\zeta}}{\boldsymbol{\xi}'\left(\mathbf{J}'\mathbf{HJ}\right)^{+}\boldsymbol{\xi}}$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are zero mean random vectors defined by (34) in Appendix, $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ are defined in Lemma A3(c.1), (c.5) and (c.6) in Appendix, respectively.

Remark 4. We find that when $0 < p < 1/2$ or $0 < q < 1/2$, the GMM estimator is consistent. However, its rate of convergence depends on p or q. When $p > q$, the rate of convergence is $N^{1/2-q}$, while when $p < q$ it is $N^{1/2-p}$. Also note that these convergence rates are slower than $\sqrt{N}$ for Case B where δ is fixed. This means that closeness of data to mean-stationarity affects the rate of convergence, which is the main purpose to introduce the parameter sequence (13).

Remark 5. When $1/2 \leq p \leq 1$ and $1/2 \leq q \leq 1$, the GMM estimator is inconsistent. This implies that as data become persistent and mean-stationary, the GMM estimator deteriorates.

Remark 6. The asymptotic distribution of Cases C1(c), C4 and C7, and Case A(a) and identical. An implication for this result is that in the presence of strong mean-nonstationarity the behavior of the GMM estimator is not affected by the degree of persistence so much.

Remark 7. When the variance ratio σ_η^2/σ_v^2 is large, the asymptotic variance of Cases C1(a),C1(b), C2 and C3 becomes small, while that of Cases C1(c), C4 and C7 becomes large. This implies that when $p > q$, large σ_η^2/σ_v^2 makes the asymptotic variance small while when $p < q$ it makes the variance large. Hence, depending on the relative size of p and q, the effect of σ_η^2/σ_v^2 changes dramatically.

Remark 8.

4 Extension of the model

In this section, we extend an AR(1) model to a model with an exogenous variable and show that similar results to the previous section could occur. The key point in the previous section is that an additional correlation between an endogenous variable and instruments appears as a result of mean-nonstationarity(see (8)). This section shows that the same is true for a model with an exogenous variable and for simplicity, we do not derive asymptotic distributions. Let us consider the following model:

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \eta_i + v_{it} \\ x_{it} &= \rho x_{i,t-1} + \kappa_i + e_{it} \end{aligned}$$

where e_{it} is *i.i.d.* and κ_i and η_i can be correlated. For initial conditions, we assume

$$y_{i0} = \delta_\mu \mu_i + \delta_\pi \pi_i + \varepsilon_{i0}, \qquad x_{i0} = \delta_\theta \theta_i + \nu_{i0}$$

where $\mu_i = \frac{\eta_i}{1-\alpha}$, $\theta_i = \frac{\kappa_i}{1-\rho}$, $\pi_i = \frac{\beta\theta_i}{1-\alpha}$. For simplicity, we assume that

$$\varepsilon_{i0} = \sum_{j=0}^{\infty} \alpha^j (v_{i,-j} + \beta r_{i,-j}), \qquad \nu_{i0} = \sum_{j=0}^{\infty} \rho^j e_{i,-j}$$

Then, we have[10]

$$\begin{aligned}
y_{it} &= \alpha^t y_{i0} + \beta \sum_{j=0}^{t-1} \alpha^j x_{i,t-j} + (1-\alpha^t)\mu_i + \sum_{j=0}^{t-1} \alpha^j v_{i,t-j} \\
&= \left[1-(1-\delta_\mu)\alpha^t\right]\mu_i + \left[\frac{1-(1-\delta_\pi)\alpha^t}{1-\alpha} - (1-\delta_\theta)\left(\sum_{j=0}^{t-1} \rho^{t-j}\alpha^j\right)\right]\beta\theta_i \\
&\quad + \sum_{j=0}^{\infty} \alpha^j (v_{i,t-j} + \beta r_{i,t-j}) \qquad (14) \\
&= h_t\mu_i + m_t\beta\theta_i + w_{it}
\end{aligned}$$

where $w_{it} = \sum_{j=0}^{\infty} \alpha^j (v_{i,t-j} + \beta r_{i,t-j})$ and $r_{it} = \sum_{j=0}^{\infty} \rho^j e_{i,t-j}$. We also have

$$\Delta y_{it} = (1-\delta_\mu)\alpha^{t-1}\eta_i + \left[(1-\delta_\pi)\alpha^{t-1} - (1-\delta_\theta)\left(\rho^t - (1-\alpha)\sum_{j=0}^{t-2}\alpha^j\rho^{t-1-j}\right)\right]\beta\theta_i + \Delta w_{it}$$

Let us consider the following moment conditions[11]:

$$E(y_{is}\Delta v_{it}) = 0 \qquad (s = 0, ..., t-2;\; t = 2, ..., T)$$

Then, the Jacobian of the moment condition is

$$\begin{aligned}
-\frac{\partial E(y_{is}\Delta v_{it})}{\partial \alpha} &= E(y_{is}\Delta y_{i,t-1}) \\
&= h_s\alpha^{t-2}(1-\delta_\mu)\frac{\sigma_\eta^2}{1-\alpha} - m_s\left[(\delta_\pi - 1) + (1-\delta_\theta)\rho\right]\beta\alpha^{t-1}\sigma_\theta^2 + E(w_{is}\Delta w_{i,t-1}) \qquad (15) \\
-\frac{\partial E(y_{is}\Delta v_{it})}{\partial \beta} &= E(y_{is}\Delta x_{it}) \\
&= m_s\rho^{t-1}(1-\delta_\theta)\frac{\sigma_\kappa^2}{1-\rho} + E(w_{is}\Delta r_{it}) \qquad (16)
\end{aligned}$$

[10]Using $x_{it} = \left[1-(1-\delta_\theta)\rho^t\right]\theta_i + \sum_{j=0}^{\infty}\rho^j e_{i,t-j} = k_t\theta_i + r_{it}$ where $k_t = 1-(1-\delta_\theta)\rho^t$ and $r_{it} = \sum_{j=0}^{\infty}\rho^j e_{i,t-j}$, we get

$$\begin{aligned}
\sum_{j=0}^{t-1}\alpha^j x_{i,t-j} &= (k_t + \alpha k_{t-1} + \cdots + \alpha^{t-1}k_1)\theta_i + \sum_{j=0}^{t-1}\alpha^j r_{i,t-j} \\
&= \left[\frac{1-\alpha^t}{1-\alpha} - (1-\delta_\theta)\sum_{j=0}^{t-1}\rho^{t-j}\alpha^j\right]\theta_i + \sum_{j=0}^{t-1}\alpha^j r_{i,t-j}
\end{aligned}$$

[11]The moment conditions $E(\Delta y_{is}u_{it}) = 0,\ (s = 1, ..., t-2;\ t = 3, ..., T)$ and $E(\Delta x_{is}u_{it}) = 0,\ (s = 1, ..., t-1;\ t = 3, ..., T)$ with $u_{it} = \eta_i + v_{it}$ which are used in constructing a system GMM estimator are valid only when both y_{it} and x_{it} are mean-stationary. Hence, in general, if one of the regressors are mean-nonstationary, Δy_{is} is not valid instruments. In such a case, only first-differenced variables which are considered to be mean-stationary can be used as instruments. If all the regressors are considered to be mean-nonstationary, all the moment conditions are invalid and a system GMM estimator is inconsistent.

From (15) and (16), it is easy to see that both initial conditions and the degree of persistence of y_{it} and x_{it} affect the strength of instruments. Hence, depending on underlying parameters, instruments may be strong in some cases and weak in other cases although we do not know in advance. The same result applies to the moment conditions $E(x_{is}\Delta v_{it}) = 0,\ (s = 0, ..., t-1;\ t = 2, ..., T)$.

5 Simulation studies

In this section, we conduct a simulation. There are mainly three purposes. The first is to assess the theoretical results obtained in the previous section. The second is to assess the size and power property of mean-stationarity test. The third is to compare the size and power property of weak instruments robust tests.

5.1 Simulation design

Although we considered a model with an exogenous variable in the previous section, to reduce the computational burden, we consider a panel AR(1) model which is mainly discussed in Section 3. The data are generated as

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}, \qquad (i = 1, ..., N;\ t = 1, ..., T) \tag{17}$$

For initial conditions, we use

$$y_{i0} = \frac{\eta_i}{1-\bar{\alpha}} + \varepsilon_{i0} \tag{18}$$

where $\bar{\alpha}$ and δ in (2) are related with $\delta = (1-\alpha)/(1-\bar{\alpha})$.

We set $\alpha = 0.95$, $T = 8$ and $N = 300$. We generate $v_{it} \sim iid\mathcal{N}(0,1)$, $\eta_i \sim iid\mathcal{N}(0,\sigma_\eta^2)$ with $\sigma_\eta^2 = 0.2, 1, 4, 10$ and $\varepsilon_{i0} \sim \mathcal{N}(0, \lambda/(1-\alpha^2))$ with $\lambda = 0.2, 1, 5$. We move $\bar{\alpha}$ from 0.9000 to 0.9975 with steps 0.0025. Note that data are mean-stationary when $\bar{\alpha} = 0.95$ and mean-nonstationary when $\bar{\alpha} \neq 0.95$. Number of replication for each design is 1000[12].

5.2 Mean, standard deviation and empirical size

We first see the behavior of mean. In addition to the FD-GMM estimator given by (5), we also compute the FD-GMM estimator using a smaller number of instruments $\mathbf{Z}_i^{L1} = diag(y_{i0}, y_{i1}, ..., y_{i,T-2})$ and one- and two-step system GMM estimators using instruments $\mathbf{Z}_i^{S21} = diag(\mathbf{Z}_i^{L2}, \mathbf{Z}_i^{D1})$ and $\mathbf{Z}_i^{S11} = diag(\mathbf{Z}_i^{L1}, \mathbf{Z}_i^{D1})$ with $\mathbf{Z}_i^{D1} = diag(\Delta y_{i1}, ..., \Delta y_{iT-1})$. Note that the system GMM estimators are consistent only when $\bar{\alpha} = 0.95$ under conventional asymptotic theory where α is fixed.

[12]Number of total designs is $3(\lambda) \times 4(\sigma_\eta^2) \times 40(\bar{\alpha}) = 480$.

Simulation results of the mean of these estimators are summarized in Figures 1 to 3[13]. For FD-GMM estimators in Figure 1, it is observed that the FD-GMM estimator performs poorly when $\lambda = 1$. However, the bias of the FD-GMM estimator becomes small when $\lambda = 5$ compared with the case of $\lambda = 0.2, 1$. As to the effect of the number of instruments, we find that reducing the number of instruments has a substantial effect on the behavior; for example, compare Figure 1(e) and 1(f). For system GMM estimators, we find that one-step and two-step estimators behave very similarly. Although the system GMM estimators are consistent only for $\bar{\alpha} = 0.95$, the magnitude of bias tend to be small when $\lambda = 5$ and σ_η^2/σ_v^2=0.2.

Next, we consider the standard deviation. Since the system GMM estimators are inconsistent except when $\bar{\alpha} = 0.95$, we only consider the FD-GMM estimators. Looking at Figures 4, we find that standard deviation becomes large around $\bar{\alpha} = 0.95$. This is consistent with the theoretical result that FD-GMM estimators become inconsistent when data are close to mean-stationary and persistency of data is strong. As in the mean, reducing the number of instruments has a substantial effects on the behavior of standard deviation.

For empirical sizes of hypothesis test $H_0 : \alpha = 0.95$, which are shown in Figures 5, size distortion around $\bar{\alpha} = 0.95$ is substantial when $\mathbf{Z}_i^{L2}$ is used as instruments. However, as $\bar{\alpha}$ moves away from 0.95, size distortion becomes small except for the case of $\sigma_\eta^2/\sigma_v^2 = 0.2$. This result seems to be natural since the bias of the FD-GMM estimator is very small when $\bar{\alpha}$ is sufficiently deviated from 0.95. For the case of the FD-GMM estimator using $\mathbf{Z}_i^{L1}$ as instruments, we find that size distortion is small. This result seems to be odd since the FD-GMM estimator is heavily biased. One possible reason for this is that the confidence interval is very wide.

5.3 Testing for mean-stationary

In empirical studies, the system GMM estimator is widely applied. However, its consistency is obtained under a very restrictive assumption that data are mean-stationary. Hence, it is important to check whether data satisfies mean-stationarity or not. A common approach to test for mean-stationarity is to use a difference Sargan test proposed by Blundell, Bond and Windmeijer (2000, p.70). Under the assumption of serially uncorrelated errors, the moment conditions of the FD-GMM estimators are valid regardless of whether data are mean-stationary or not. However, a subset of moment conditions used in constructing the system GMM estimator, i.e., $E(\Delta y_{is}(\eta_i + v_{it})), (s = 1, ..., t-1; t = 2, ..., T)$ is valid only when data are mean-stationary. Hence by computing the difference of J statistic based on the FD-GMM

[13] In figures 1 to 9, the horizontal axis is $\bar{\alpha}$ ranging from 0.9000 to 0.9975 and four lines corresponds to $\sigma_\eta^2 = 0.2, 1, 4, 10$. "IV=XX" means that $\mathbf{Z}_i^{XX}$ is used as instruments. For example, "IV=L2" means that $\mathbf{Z}_i^{L2}$ is used as instruments.

and system GMM estimators, we can check whether data are mean-stationary or not. However although testing for mean-stationarity is simple to implement, to the best of author's knowledge, there are few studies that assess the performance of mean-stationarity test[14]. Hence, we conduct a simulation to assess the size and power of mean-stationarity test. Since we consider the difference of J statistics of the FD- and system GMM estimators, several combinations can be considered. We consider two cases. The first case is taking a difference of J statistics based on FD- and system GMM estimators using $\mathbf{Z}_i^{L2}$ and $\mathbf{Z}_i^{S21}$, respectively. In this case, we write "IV=S21-L2". The second case is using instruments $\mathbf{Z}_i^{L1}$ for the FD-GMM estimator and $\mathbf{Z}_i^{S11} = diag(\mathbf{Z}_i^{L1}, \mathbf{Z}_i^{D1})$, denoted by "IV=S11-L1". Simulation results are summarized in Figure 6. Note that $\bar{\alpha} = 0.95$ corresponds to the size and other values of $\bar{\alpha}$ correspond to power. Figure shows that size distortion is substantial for all cases. For the (size unadjusted) power, we find that the power gets higher in the region where $\bar{\alpha}$ is moderately deviated from 0.95 as σ_η^2/σ_v^2 grows. It should be noted that when σ_η^2/σ_v^2 is small, power is very low and almost equal to empirical size in some cases. These results suggest that we should use a mean-stationarity test very carefully.

5.4 Weak instruments robust test

In section 3, we showed that initial conditions affect the strength of instruments. However, in practice, it is difficult to judge whether instruments are strong or weak. Hence, in some cases the estimates are very precise and in other cases, the estimates might be largely biased and inference might be unreliable. Fortunately, several methods have been proposed to conduct an inference which does not depend on the strength of instruments. The first is K statistic by Kleibergen (2005, 2007), the second is S statistic by Stock and Wright (2000), and the third is conditional likelihood ratio (CLR) test by Moreira (2003), which is extended to GMM by Kleibergen (2005, 2007). We call the test statistic of the conditional likelihood CLR) test as M statistic. These three methods have a correct size regardless of whether instruments are weak or strong. Figures 7 to 9 show that size distortion is very small for all cases[15]. We now consider the power properties of K, S and M statistics to test for $H_0 : \alpha = a$ where a moves from 0.850 to 0.995 with 0.005 steps. We set $\bar{\alpha} = 0.94, 0.95, 0.96$. Simulation results are summarized in Figures 10 to 18. Note that the horizontal axis of these figures are a not $\bar{\alpha}$ as in the previous figures. From these figures, we find that the power is affected by σ_η^2/σ_v^2 and $\bar{\alpha}$. For $\bar{\alpha} = 0.94, 0.96$, the power gets higher as σ_η^2/σ_v^2 grows. However, when $\bar{\alpha} = 0.95$ or when $\bar{\alpha} = 0.94$ with $\sigma_\eta^2/\sigma_v^2 = 0.2$, all three tests have very low power; in some

[14]Blundell, Bond and Windmeijer (2000) discuss only size property and do not consider the power.
[15]To implement the CLR test, we use a rank test by Kleibergen and Paap (2006).

cases, the powers are almost equal to empirical sizes. Hence, these weak instruments robust tests are not always a useful method. Among the three methods, the performance of K and M statistics is very similar and better than S statistic.

6 Conclusion

In this paper, we investigated the behavior of the FD-GMM estimator for persistent dynamic panel data models where the autoregressive parameter is modeled as $\alpha_N = 1 - c/N^p,\ (0 < p \leq 1)$. We showed that assumption on initial conditions affects the rate of convergence. If initial conditions are mean-stationary, the FD-GMM estimator is $N^{1/2-p}$-consistent when $0 < p < 1/2$, while inconsistent when $1/2 \leq p \leq 1$. When data are mean-nonstationary, the FD-GMM estimator is shown to be $\sqrt{N}$-consistent for $0 < p \leq 1$. Furthermore, to take the closeness of data to mean-stationarity into consideration, we introduced a notion of "near mean-stationarity" where data are deviated from mean-stationarity by $O(1/N^q), (0 < q \leq 1)$. Under a local to unity and near mean-stationary systems, we showed that both p and q affect the rate of convergence. When both p and q are in the region $1/2 \leq p,\ q \leq 1$, the FD-GMM estimator was shown to be inconsistent while for other smaller values of p and q, the FD-GMM estimator was shown to be consistent with a convergence rate $N^{1/2-p}$ or $N^{1/2-q}$. Simulation results showed that the derived theoretical results capture the finite sample behavior reasonably well. Additional simulation results showed that the performance of a mean-stationarity test is heavily affected by the variance ratio of individual effects and errors; in some cases, the power is almost equal to the empirical size. We also assessed the performance of three weak instruments robust tests. Simulation results showed that although size is close to the nominal one, their powers are substantially affected by the initial conditions; when data are mean-stationary, these tests have almost no power and when data are mean-nonstationary their power depend on the variance ratio of individual effects and errors. These simulation results imply that initial conditions substantially affect the behavior of the GMM estimator and related tests. Hence it is an important topic to develop an estimator which is not affected by the initial conditions so much. Also, constructing a statistic that measures the strength of instruments such as Shea (1997) or Stock and Yogo (2005) in dynamic panel context might be useful for applied research.

Appendix

Lemma A1. *For $0 < p \leq 0$ and $t \geq 1$,*

$$\alpha_N^t = 1 - \frac{ct}{N^p} + O\left(\frac{1}{N^{2p}}\right) \tag{19}$$

$$
\begin{aligned}
h_t &= 1-(1-\delta)\alpha_N^t = \begin{cases} 1 & \textit{for case A} \\ \delta + \dfrac{ct(1-\delta)}{N^p} + O\left(\dfrac{1}{N^{2p}}\right) & \textit{for case B} \end{cases} && (20) \\
h_t &= 1-(1-\delta_N)\alpha_N^t = 1 - \frac{d}{N^q} + o\left(\frac{1}{N^q}\right) \qquad \textit{for case C} && (21)
\end{aligned}
$$

Proof of Lemma A1 Using asymptotic equivalence $\log(1+x) = x + O(x^2)$ as $x \to 0$ and $e^x = 1 + x + O(x^2)$, we have

$$
\alpha_N^t = \exp\left[t\log\left(1-\frac{c}{N^p}\right)\right] = \exp\left[\frac{-ct}{N^p} + O\left(\frac{1}{N^{2p}}\right)\right] = 1 - \frac{ct}{N^p} + O\left(\frac{1}{N^{2p}}\right)
$$

The results for h_t are easily obtained from the first result. □

Below, we denote α_N and δ_N as α and δ, respectively, for brevity.

Lemma A2. *Let us define* $w_{i,t-1} = v_{i,t-1} + (\alpha-1)\sum_{j=0}^{t-3}\alpha^j v_{i,t-j-2}$ *and* $\psi_j = (1-\alpha^{2j})/(1-\alpha^2)$. *Then, for* $t \geq s$, *we have*

$$
(a) \quad E(y_{is}\Delta y_{i,t-1}) = h_s\alpha^{t-2}(1-\delta)\frac{\sigma_\eta^2}{1-\alpha} - \frac{\sigma_v^2\alpha^{t-s-2}[1-(1-\lambda)\alpha^{2s}]}{1+\alpha}, \qquad (22)
$$

$$
(s = 0, ..., t-2, \ t \geq 2)
$$

$$
= \begin{cases} \dfrac{-\lambda\sigma_v^2}{2-c/N^p} + o(1) & \textit{for case A} \\ \dfrac{\delta(1-\delta)\sigma_\eta^2 N^p}{c} + O(1) & \textit{for case B} \\ \dfrac{d\sigma_\eta^2 N^{p-q}}{c} - \dfrac{\lambda\sigma_v^2}{2-c/N^p} + o(N^{p-q}) + o(1) & \textit{for case C} \end{cases} \qquad (23)
$$

$$
(b) \quad var(y_{is}\Delta y_{i,t-1})
$$

$$
= \begin{cases} (1-\delta)^2 h_0^2 \dfrac{var(\eta_i^2)}{(1-\alpha)^2} + (1-\alpha)^2 var(\varepsilon_{i0}^2) & \\ +\sigma_v^2(h_0^2\sigma_\mu^2 + \sigma_0^2) + \sigma_\eta^2\sigma_0^2(1-\delta-h_0)^2 & (s=0, t=2) \\ & \\ (1-\delta)^2 h_s^2 \alpha^{2(t-2)} \dfrac{var(\eta_i^2)}{(1-\alpha)^2} + var(\xi_{is}w_{i,t-1}) & \\ +(1-\alpha)^2\alpha^{2(t+s-2)}var(\varepsilon_{i0}^2) & \\ +[\sigma_v^2 + (1-\alpha)^2\sigma_v^2\psi_{t-2}][h_s^2\sigma_\mu^2 + \alpha^{2s}\sigma_0^2] & \\ +\alpha^{2(t-2)}[(1-\delta)\alpha^s - h_s]^2\sigma_\eta^2\sigma_0^2 & \\ +\sigma_v^2\psi_s\alpha^{2(t-2)}[(1-\delta)^2\sigma_\eta^2 + (1-\alpha)^2\sigma_0^2] & \\ -2\psi_s\sigma_v^2\alpha^{2t-s-4}[(1-\delta)h_s\sigma_\eta^2 - \alpha^s(1-\alpha)^2\sigma_0^2] & (s = 0, .., t-2, \ t \geq 3) \end{cases} \qquad (24)
$$

$$
= \begin{cases} \dfrac{\sigma_v^2\sigma_\eta^2 N^{2p}}{c^2} + o(N^{2p}) & \textit{for case A, C} \\ \dfrac{[\delta^2(1-\delta)^2 var(\eta_i^2) + \sigma_v^2\sigma_\eta^2\delta^2]N^{2p}}{c^2} + o(N^{2p}) & \textit{for case B} \end{cases} \qquad (25)
$$

$$
(c) \quad E(y_{it}y_{is}) = h_t h_s \sigma_\mu^2 + \frac{\alpha^{t-s}[1-(1-\lambda)\alpha^{2s}]\sigma_v^2}{1-\alpha^2} \qquad (26)
$$

$$= \begin{cases} \dfrac{\sigma_\eta^2 N^{2p}}{c^2} + o(N^{2p}) & \textit{for case A, C} \\ \dfrac{\sigma_\eta^2(1-\delta)^2 N^{2p}}{c^2} + O(N^p) & \textit{for case B} \end{cases} \tag{27}$$

$$(d) \quad var(y_{it}y_{is}) = h_t^2 h_s^2 var(\mu_i^2) + var(\xi_{it}\xi_{is}) + \alpha^{2(t+s)} var(\varepsilon_{i0}^2)$$
$$+ \left(h_t\alpha^s + h_s\alpha^t\right)^2 \sigma_\mu^2\sigma_0^2 + \left(h_t^2\psi_s + h_s^2\psi_t + 2h_sh_t\alpha^{t-s}\psi_s\right)\sigma_\mu^2\sigma_v^2$$
$$+ \left(\alpha^{2s}\psi_t + 3\alpha^{2t}\psi_s\right)\sigma_v^2\sigma_0^2 \tag{28}$$
$$= O(N^{4p}) \qquad \textit{for case A, B, C} \tag{29}$$

$$(e) \quad var(y_{is}\Delta v_{it}) = 2h_th_s\sigma_v^2\sigma_\mu^2 + \frac{2\alpha^{t-s}[1-(1-\lambda)\alpha^{2s}]\sigma^4}{1-\alpha^2} \qquad (s = 0, .., t-2) \tag{30}$$

$$= \begin{cases} \dfrac{2\sigma_v^2\sigma_\eta^2 N^{2p}}{c^2} + o\left(N^{2p}\right) & \textit{for Case A, C} \\ \dfrac{2(1-\delta)^2\sigma_v^2\sigma_\eta^2 N^{2p}}{c^2} + O\left(N^p\right) & \textit{for Case B} \end{cases} \tag{31}$$

Proof of Lemma A2 Note that $\Delta y_{i,t-1}$ can be written as

$$\Delta y_{i,t-1} = \begin{cases} (1-\delta)\eta_i + v_{i1} - (1-\alpha)\varepsilon_{i0}, & t = 2 \\ (1-\delta)\alpha^{t-2}\eta_i + w_{i,t-1} - (1-\alpha)\alpha^{t-2}\varepsilon_{i0} & t \geq 3 \end{cases}. \tag{32}$$

Also, note that

$$\frac{1}{1-\alpha} = \frac{N^p}{c}, \quad \frac{1}{1+\alpha} = \frac{1}{2-c/N^p}, \quad \frac{1}{1-\alpha^2} = \frac{N^p}{2c-c^2/N^p}. \tag{33}$$

(a): Using $E(w_{i,t-1}^2) = \sigma_v^2 + (1-\alpha)^2\sigma_v^2\psi_{t-2}$, $E(\xi_{is}\xi_{it}) = \sigma_v^2\alpha^{t-s}\psi_s$ for $s \leq t$, and $E(\xi_{is}w_{i,t-1}) = -(1-\alpha)\alpha^{t-s-2}\psi_s\sigma_v^2$ for $s = 0, .., t-2$, we get (22). The results for cases A and C can be obtained by using (33) and Lemma A1. The result for case B can be obtained by noting that the first term in (22) is dominating the second one.
(b): We prove the case of $t \geq 3$ since the result of $t = 2$ is straightforward. Using (32), (24) for the case of $t \geq 3$ is obtained as follows:

$$\begin{aligned} var(y_{is}\Delta y_{i,t-1}) &= var\Bigg[\left(\frac{1-\delta}{1-\alpha}\right)\alpha^{t-2}h_s\eta_i^2 + \xi_{is}w_{i,t-1} - (1-\alpha)\alpha^{t+s-2}\varepsilon_{i0}^2 \\ &\quad +(1-\delta)\alpha^{t-2}\eta_i\xi_{is} + \left((1-\delta)\alpha^{t+s-2} - h_s\alpha^{t-2}\right)\eta_i\varepsilon_{i0} \\ &\quad +h_s\mu_iw_{i,t-1} + \xi_{is}w_{i,t-1} + \alpha^s w_{i,t-1}\varepsilon_{i0} - (1-\alpha)\alpha^{t-2}\xi_{is}\varepsilon_{i0}\Bigg] \\ &= \frac{(1-\delta)^2}{(1-\alpha)^2}\alpha^{2(t-2)}h_s^2 var(\eta_i^2) + var(\xi_{is}w_{i,t-1}) + (1-\alpha)^2\alpha^{2(t+s-2)}var(\varepsilon_{i0}^2) \\ &\quad +(1-\delta)^2\alpha^{2(t-2)}var(\eta_i\xi_{is}) + \left((1-\delta)\alpha^{t+s-2} - h_s\alpha^{t-2}\right)^2 var(\eta_i\varepsilon_{i0}) \\ &\quad +h_s^2 var(\mu_iw_{i,t-1}) + \alpha^{2s}var(w_{i,t-1}\varepsilon_{i0}) + (1-\alpha)^2\alpha^{2(t-2)}var(\xi_{is}\varepsilon_{i0}) \\ &\quad +2(1-\delta)h_s\alpha^{t-2}cov(\eta_i\xi_{is}, \mu_iw_{i,t-1}) - 2(1-\alpha)\alpha^{t+s-2}cov(w_{i,t-1}\varepsilon_{i0}, \xi_{is}\varepsilon_{i0}) \\ &= \frac{(1-\delta)^2}{(1-\alpha)^2}\alpha^{2(t-2)}h_s^2 var(\eta_i^2) + var(\xi_{is}w_{i,t-1}) + (1-\alpha)^2\alpha^{2(t+s-2)}var(\varepsilon_{i0}^2) \\ &\quad +(1-\delta)^2\alpha^{2(t-2)}\sigma_v^2\sigma_\eta^2\psi_s + \left((1-\delta)\alpha^{t+s-2} - h_s\alpha^{t-2}\right)^2\sigma_\eta^2\sigma_0^2 \end{aligned}$$

$$
\begin{aligned}
&+h_s^2\sigma_\mu^2(\sigma_v^2+(1-\alpha)^2\sigma_v^2\psi_{t-2})+\alpha^{2s}\sigma_0^2(\sigma_v^2+(1-\alpha)^2\sigma_v^2\psi_{t-2})\\
&+(1-\alpha)^2\alpha^{2(t-2)}\psi_s\sigma_v^2\sigma_0^2\\
&-2(1-\delta)h_s\psi_s\alpha^{2t-s-4}\sigma_v^2\sigma_\eta^2+2(1-\alpha)^2\alpha^{2t-4}\psi_s\sigma_v^2\sigma_0^2.
\end{aligned}
$$

To prove (25), we need to derive $var(\xi_{is}w_{i,t-1})$. Using $\xi_{is}=\sum_{j=1}^{s}\alpha^{s-j}v_{ij}$ and $w_{i,t-1}=v_{i,t-1}-(1-\alpha)\sum_{\ell=1}^{t-2}\alpha^{t-\ell-2}v_{i\ell}=\sum_{\ell=1}^{t-1}b_\ell v_{i\ell}$ where $b_{t-1}=1$, $b_\ell=-(1-\alpha)\alpha^{t-\ell-2}$, $(\ell=1,...,t-2)$, we have

$$
\begin{aligned}
E(\xi_{is}^2w_{i,t-1}^2) &= \sum_{j_1,j_2=1}^{s}\sum_{\ell_1,\ell_2=1}^{t-1}\alpha^{2s-j_1-j_2}b_{\ell_1}b_{\ell_2}E(v_{ij_1}v_{ij_2}v_{i\ell_1}v_{i\ell_2})\\
&= E(v_{ij}^4)\sum_{j=1}^{s}\alpha^{2s-2j}b_j^2+\sigma^4\sum_{j=1}^{s}\sum_{\ell=1}^{t-1}\alpha^{2s-2j}b_\ell^2+2\sigma^4\sum_{j=1}^{s}\sum_{\ell=1}^{s}\alpha^{2s-j-\ell}b_jb_\ell\\
&= E(v_{ij}^4)(1-\alpha)^2\alpha^{t-s-2}\frac{1-\alpha^{2s}}{1-\alpha^2}+\sigma^4\left(\frac{1-\alpha^{2s}}{1-\alpha^2}\right)\left(1+(1-\alpha)^2\frac{1-\alpha^{2(t-2)}}{1-\alpha^2}\right)\\
&\quad+2\sigma^4(1-\alpha)^2\alpha^{2(t-s-2)}\frac{(1-\alpha^{2s})^2}{(1-\alpha^2)^2}\\
&= \frac{\sigma^4N^p}{2c-c^2/N^p}+o(N^p)
\end{aligned}
$$

Hence, by noting that $E(\xi_{is}w_{i,t-1})=O(1)$, we have

$$
var(\xi_{is}w_{i,t-1})=\frac{\sigma^4N^p}{2c-c^2/N^p}+o(N^p)
$$

For cases A and C, the result follows by noting that the fourth term in (24) is $O(N^{2p})$ and dominating other terms. The result for case B is obtained by noting that the first and fourth terms are $O(N^{2p})$ and dominating other terms.

(c): (26) is straightforward to show using (3). (27) is obtained by noting that the first term in (26) is dominating.

(d): Using (3), (28) is obtained as follows:

$$
\begin{aligned}
var(y_{it}y_{is}) &= var\left[h_th_s\mu_i^2+\xi_{it}\xi_{is}+\alpha^{t+s}\varepsilon_{i0}^2+h_t\mu_i\xi_{is}\right.\\
&\qquad\left.+(h_t\alpha^s+h_s\alpha^t)\mu_i\varepsilon_{i0}+h_s\mu_i\xi_{it}+\alpha^s\xi_{it}\varepsilon_{i0}+\alpha^t\xi_{is}\varepsilon_{i0}\right]\\
&= h_t^2h_s^2var(\mu_i^2)+var(\xi_{is}\xi_{it})+\alpha^{2(t+s)}var(\varepsilon_{i0}^2)+h_t^2var(\mu_i\xi_{is})\\
&\quad+\left(h_t\alpha^s+h_s\alpha^t\right)^2var(\mu_i\varepsilon_{i0})+h_s^2var(\mu_i\xi_{it})+\alpha^{2s}var(\xi_{it}\varepsilon_{i0})+\alpha^{2t}var(\xi_{is}\varepsilon_{i0})\\
&\quad+2h_th_scov(\mu_i\xi_{is},\mu_i\xi_{it})+2\alpha^{t+s}cov(\xi_{is}\varepsilon_{i0},\xi_{it}\varepsilon_{i0})\\
&= h_t^2h_s^2var(\mu_i^2)+var(\xi_{is}\xi_{it})+\alpha^{2(t+s)}var(\varepsilon_{i0}^2)+h_t^2\sigma_\mu^2\sigma_v^2\psi_s\\
&\quad+\left(h_t\alpha^s+h_s\alpha^t\right)^2\sigma_\mu^2\sigma_0^2+h_s^2\sigma_\mu^2\sigma_v^2\psi_t+\alpha^{2s}\sigma_v^2\sigma_0^2\psi_t+\alpha^{2t}\sigma_v^2\sigma_0^2\psi_s\\
&\quad+2h_th_s\sigma_v^2\sigma_\mu^2\alpha^{t-s}\psi_s+2\alpha^{2t}\sigma_v^2\sigma_0^2\psi_s\\
&= a_1+a_2+\cdots+a_{10},
\end{aligned}
$$

To show (29), we need to assess the magnitude of orders of each term. From (20) and (21), it is easy to show that a_1 is $O(N^{4p})$ and $a_4, \cdots, a_{10}$ are smaller than $O(N^{3p})$. To complete the proof, we show that $a_2 = O(N^{2p})$ and $a_3 = O(N^{2p})$.

$a_3 = O(N^{2p})$ is obtained by noting that $E(\varepsilon_{i0}^4) = \kappa\sigma_0^4 = O(N^{2p})$ from Assumption 3. To show $a_2 = O(N^{2p})$, we use the following relationship, which is obtained by applying the Cauchy-Schwartz inequality:

$$var(\xi_{is}\xi_{it}) < E(\xi_{is}^2\xi_{it}^2) \leq \sqrt{E(\xi_{is}^4)E(\xi_{it}^4)}.$$

Using $\xi_{is} = \sum_{j=1}^{s} \alpha^{s-j} v_{ij}$, we have

$$\begin{aligned}
E(\xi_{is}^4) &= \sum_{j_1,j_2,j_3,j_4=1}^{s} \alpha^{4s-j_1-j_2-j_3-j_4} E(v_{ij_1}v_{ij_2}v_{ij_3}v_{ij_4}) \\
&= \sum_{j=1}^{s} \alpha^{4(s-j)} E(v_{ij}^4) + 3\sigma^4 \sum_{\substack{j_1,j_2=1 \\ j_1\neq j_2}}^{s} \alpha^{4s-2j_1-2j_2} \\
&= \frac{E(v_{ij}^4)(1-\alpha^{4s})}{1-\alpha^4} + 3\sigma^4 \left(\frac{1-\alpha^{2s}}{1-\alpha^2}\right)^2 = O(N^{2p}).
\end{aligned}$$

Hence, $var(y_{it}y_{is}) = O(N^{4p})$ follows.
(e): (30) is obtained by noting that $var(y_{is}\Delta v_{it}) = 2\sigma_v^2 E(y_{is}^2)$. (31) is obtained from (27). □

Next, we show the convergence results which are used in deriving the asymptotic behavior of the GMM estimator. For simplicity, we denote $\mathbf{Z}_i^{L2}$ as $\mathbf{Z}_i$.

Lemma A3. *Let Assumptions 1 to 4 hold. Then as $N \to \infty$ with T fixed, we have*

$$(a.1) \quad \frac{1}{N^{2p+1}} \sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i \xrightarrow{p} \frac{\sigma_\eta^2}{c^2}\mathbf{J}'\mathbf{H}\mathbf{J}, \qquad \textit{for cases A and C}$$

$$(a.2) \quad \frac{1}{N^{2p+1}} \sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i \xrightarrow{p} \frac{\sigma_\eta^2(1-\delta)^2}{c^2}\mathbf{J}'\mathbf{H}\mathbf{J}, \qquad \textit{for case B}$$

$$(b.1) \quad \frac{1}{N^{(2p+1)/2}} \sum_{i=1}^{N} \mathbf{Z}_i'\Delta\mathbf{v}_i \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{2\sigma_v^2\sigma_\eta^2}{c^2}\mathbf{J}'\mathbf{H}\mathbf{J}\right), \qquad \textit{for cases A and C}$$

$$(b.2) \quad \frac{1}{N^{(2p+1)/2}} \sum_{i=1}^{N} \mathbf{Z}_i'\Delta\mathbf{v}_i \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{2\sigma_v^2\sigma_\eta^2(1-d)^2}{c^2}\mathbf{J}'\mathbf{H}\mathbf{J}\right), \qquad \textit{for case B}$$

$$(c.1) \quad \frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \frac{-\lambda\sigma_v^2}{2}\boldsymbol{\iota}_m \equiv \boldsymbol{\mu}_1, \qquad \textit{for case A(a)}$$

$$(c.2)\quad \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \boldsymbol{\xi}+\boldsymbol{\mu}_1 \\ \boldsymbol{\zeta} \end{bmatrix} \qquad \textit{for case A(b)}$$

$$(c.3)\quad \begin{bmatrix} \frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \\ \frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \qquad \textit{for case A(c)}$$

$$(c.4)\quad \frac{1}{N^{p+1}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \frac{\delta(1-\delta)\sigma_\eta^2}{c}\boldsymbol{\iota}_m, \qquad \textit{for case B}$$

$$(c.5)\quad \frac{1}{N^{p-q+1}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \frac{d\sigma_\eta^2}{c}\boldsymbol{\iota}_m \equiv \boldsymbol{\mu}_2, \qquad \textit{for case C1(a)}$$

$$(c.6)\quad \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \left(\frac{d\sigma_\eta^2}{c} - \frac{\lambda\sigma_v^2}{2}\right)\boldsymbol{\iota}_m \equiv \boldsymbol{\mu}_3, \qquad \textit{for case C1(b)}$$

$$(c.7)\quad \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \boldsymbol{\mu}_1, \qquad \textit{for case C1(c)}$$

$$(c.8)\quad \frac{1}{N^{3/2-q}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \boldsymbol{\mu}_2, \qquad \textit{for case C2}$$

$$(c.9)\quad \frac{1}{N^{p-q+1}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \boldsymbol{\mu}_2, \qquad \textit{for case C3}$$

$$(c.10)\quad \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \boldsymbol{\mu}_1, \qquad \textit{for case C4}$$

$$(c.11)\quad \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \boldsymbol{\xi}+\boldsymbol{\mu}_3 \\ \boldsymbol{\zeta} \end{bmatrix} \qquad \textit{for case C5}$$

$$(c.12)\quad \begin{bmatrix} \frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \\ \frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \boldsymbol{\xi}+\boldsymbol{\mu}_2 \\ \boldsymbol{\zeta} \end{bmatrix} \qquad \textit{for case C6}$$

$$(c.13) \quad \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \xrightarrow{p} \boldsymbol{\mu}_1, \qquad \textit{for case C7}$$

$$(c.14) \quad \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \boldsymbol{\xi}+\boldsymbol{\mu}_1 \\ \boldsymbol{\zeta} \end{bmatrix} \qquad \textit{for case C8}$$

$$(c.15) \quad \begin{bmatrix} \frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1} \\ \frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \qquad \textit{for case C9(a),(b),(c)}$$

where

$$\begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{\sigma_\eta^2\sigma_v^2}{c^2}\boldsymbol{\iota}_m\boldsymbol{\iota}_m' & \mathbf{C} \\ \mathbf{C}' & \frac{2\sigma_v^2\sigma_\eta^2}{c^2}\mathbf{J}'\mathbf{H}\mathbf{J} \end{pmatrix}\right] \tag{34}$$

$$\mathbf{C} = \lim_{N\to\infty} cov\left(\frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1}, \frac{1}{N^{(2p+1)/2}}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i\right).$$

Proof of Lemma A3 (a.1)-(a.2): The results are obtained by noting that $E(y_{it}y_{is})$ does not depend on s and t when $\alpha = 1 - c/N^p$(see Lemma A2(c)).
(b.1)-(b.2): Since $E(\mathbf{Z}_i'\Delta\mathbf{v}_i) = \mathbf{0}$ and $var\left(\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{v}_i\right) = 2\sigma_v^2 N E(\mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i)$, by applying the central limit theorem and from (a.1)-(a.2), the results follow.
(c.1): Since $E\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1}\right) = O(1)$ and $var\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1}\right) = O(1/N) \to \mathbf{0}$, the result follows. (c.4), (c.5), (c.6), (c.7), (c.8), (c.9), (c.10), and (c.13) can be proved in the same way.
(c.2): Since $E\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1}\right) = O(1)$ and $var\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1}\right) = O(1)$, by applying the central limit theorem, the result follows. (c.11), (c.12), and (c.14) can be proved in the same way.
(c.3): Since $E\left(N^{-(2p+1)/2}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1}\right) \to \mathbf{0}$ and $var\left(N^{-(2p+1)/2}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\mathbf{y}_{i,-1}\right) = O(1)$, by applying the central limit theorem, the result follows. (c.15) can be proved in the same way. □

Proof of Theorem 1 and 2 Using Lemma A3, the results are obtained. □

References

[1] Alvarez, J. and M. Arellano (2003) "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71 (4), 1121–1159.

[2] Anderson, T. W. and C. Hsiao (1981) "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76 (375), 598–606.

[3] Anderson, T. W. and C. Hsiao (1982) "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18 (1), 47–82.

[4] Arellano, M. (2003a) "Modelling Optimal Instrumental Variables For Dynamic Panel Data Models," working paper.

[5] Arellano, M. (2003b) *Panel Data Econometrics*, Oxford: Oxford University Press.

[6] 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 (2), 277–97.

[7] Arellano, M. and O. Bover (1995) "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68 (1), 29–51.

[8] Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87 (1), 115–143.

[9] Blundell, R., S. Bond, and F. Windmeijer (2000) "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator," in B. H. Baltagi ed. *Nonstationary Panels, Panel Cointegration and Dynamic Panels*, 15 of Advances in Econometrics, Amsterdam: JAI Press, 53–91.

[10] Bun, M. J. G. and F. Windmeijer (2010) "The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models," *Econometrics Journal*, 13 (1), 95–126.

[11] Caner, M. (2010) "Testing, Estimation in GMM and CUE with Nearly-Weak Identification," *Econometric Reviews*, 29 (3), 330–363.

[12] Elliott, G. and U. K. Müller (2006) "Minimizing the Impact of the Initial Condition on Testing for Unit Roots," *Journal of Econometrics*, 135 (1-2), 285–310.

[13] Giraitis, L. and P. C. B. Phillips (2006) "Uniform Limit Theory for Stationary Autoregression," *Journal of Time Series Analysis*, 27 (1), 51–60.

[14] Hahn, J. and G. Kuersteiner (2002) "Discontinuities of Weak Instrument Limiting Distributions," *Economics Letters*, 75 (3), 325–331.

[15] Hahn, J., J. Hausman, and G. Kuersteiner (2007) "Long Difference Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects," *Journal of Econometrics*, 127 (2), 574–617.

[16] Harris, D., D. I. Harvey, S. J. Leybourne, and N. D. Sakkas (2010) "Local Asymptotic Power of the Im-Pesaran-Shin Panel Unit Root Test and the Impact of Initial Observations," *Econometric Theory*, 26 (1), 311–324.

[17] Harvey, D. I. and S. J. Leybourne (2005) "On Testing for Unit Roots and the Initial Observation," *Econometrics Journal*, 8 (1), 97–111.

[18] Harvey, D. I. and S. J. Leybourne (2006) "Power of a Unit-Root Test and the Initial Condition," *Journal of Time Series Analysis*, 27 (5), 739–752.

[19] Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2009) "Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition," *Econometric Theory*, 25 (3), 587–636.

[20] Hayakawa, K. (2009) "On the Effect of Mean-Nonstationarity in Dynamic Panel Data Models," *Journal of Econometrics*, 153 (2), 133–135.

[21] Holtz-Eakin, D., W. K. Newey, and H. S. Rosen (1988) "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56 (6), 1371–95.

[22] Hsiao, C. (2003) *Analysis of Panel Data*, Cambridge and New York: Cambridge University Press.

[23] Im, K. S., M. H. Pesaran, and Y. Shin (2003) "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115 (1), 53–74.

[24] Kiviet, J. F. (2007) "Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models," in G. D. A. Phillips and E. Tzavalis eds. *The Refinement of Econometric Estimation and Test Procedures*: Cambridge University Press.

[25] Kleibergen, F. (2005) "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73 (4), 1103–1123.

[26] Kleibergen, F. (2007) "Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, 139 (1), 181–216.

[27] Kleibergen, F. and R. Paap (2006) "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133 (1), 97–126.

[28] Kruiniger, H. (2009) "GMM Estimation and Inference in Dynamic Panel Data Models with Persistent Data," *Econometric Theory*, 25 (5), 1348–1391.

[29] Kurozumi, E. and K. Hayakawa (2009) "Asymptotic Properties of the Efficient Estimators for Cointegrating Regression Models with Serially Dependent Errors," *Journal of Econometrics*, 149 (2), 118–135.

[30] Madsen, E. (2010) "Unit Root Inference in Panel Data Models Where the Time-series Dimension is Fixed: A comparison of Different Test," *Econometrics Journal*, 13 (1), 63–94. forthcoming.

[31] Magdalinos, T. and P. C. Phillips (2009) "Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors," *Econometric Theory*, 25 (2), 482–526.

[32] Moon, H. R. and P. C. B. Phillips (2000) "Estimation of Autoregressive Roots Near Unity Using Panel Data," *Econometric Theory*, 16 (6), 927–997.

[33] Moon, H. R. and P. C. B. Phillips (2004) "GMM Estimation of Autoregressive Roots Near Unity with Panel Data," *Econometrica*, 72 (2), 467–522.

[34] Moreira, M. J. (2003) "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71 (4), 1027–1048.

[35] Müller, U. K. and G. Elliott (2003) "Tests for Unit Roots and the Initial Condition," *Econometrica*, 71 (4), 1269–1286.

[36] Phillips, P. C. B. and T. Magdalinos (2007a) "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics*, 127 (1), 115–130.

[37] Phillips, P. C. B. and T. Magdalinos (2007b) "Limit Theory for Moderate Deviations from a Unit Root Under Weak Dependence," in G. D. A. Phillips and E. Tzavalis eds. *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*: Cambridge University Press., 123–162.

[38] Shea, J. (1997) “Instrument Relevance in Multivariate Linear Models: A Simple Measure,” *Review of Economics and Statistics*, 79 (2), 348–352.

[39] Stock, J. H. and J. Wright (2000) “GMM with Weak Identification,” *Econometrica*, 68 (5), 1055–1096.

[40] Stock, J. H. and M. Yogo (2005) “Testing for Weak Instruments in Linear IV Regression,” in D. W. K. Andrews and J. H. Stock eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*: Cambridge University Press, Chap. 5, 80–108.

[41] 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 (4), 518–529.

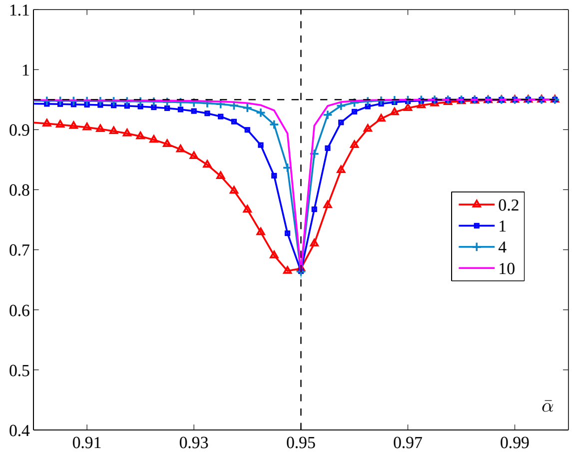


Figure 1(a): Mean of the first-difference GMM estimator
(IV=L2, $\lambda = 0.2$)

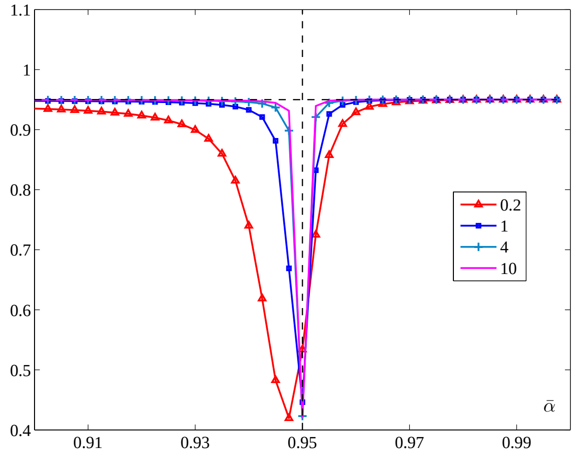


Figure 1(b): Mean of the first-difference GMM estimator
(IV=L1, $\lambda = 0.2$)

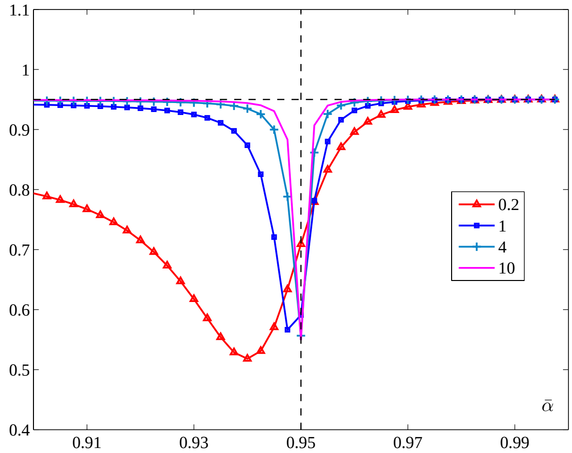


Figure 1(c): Mean of the first-difference GMM estimator
(IV=L2, $\lambda = 1$)

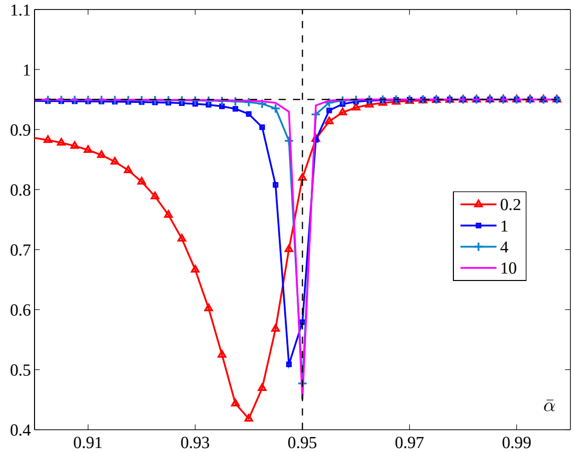


Figure 1(d): Mean of the first-difference GMM estimator
(IV=L1, $\lambda = 1$)

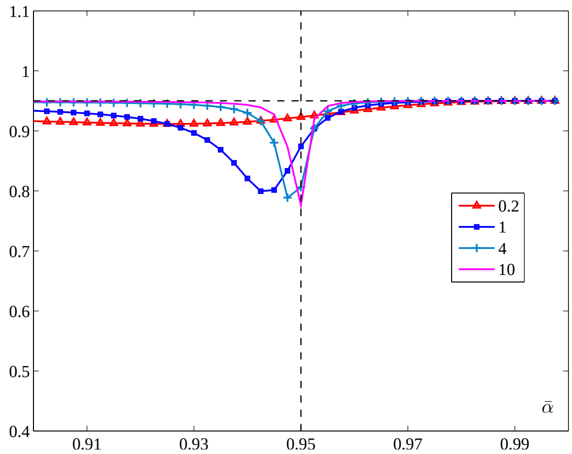


Figure 1(e): Mean of the first-difference GMM estimator
(IV=L2, $\lambda = 5$)

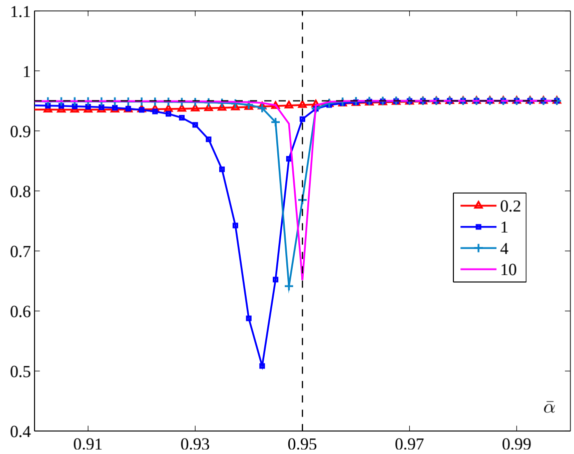


Figure 1(f): Mean of the first-difference GMM estimator
(IV=L1, $\lambda = 5$)

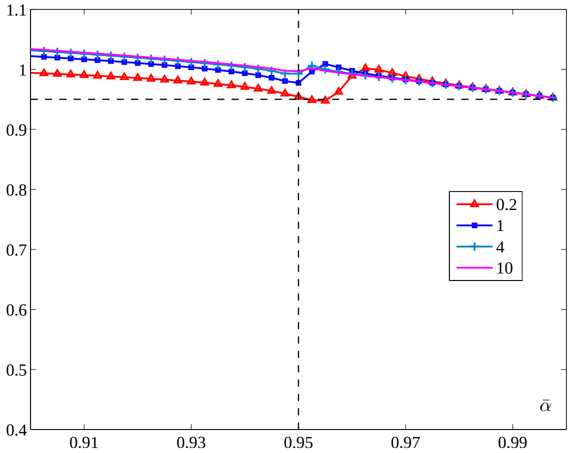


Figure 2(a): Mean of 1 step system GMM estimator (IV=S21, $\lambda = 0.2$)

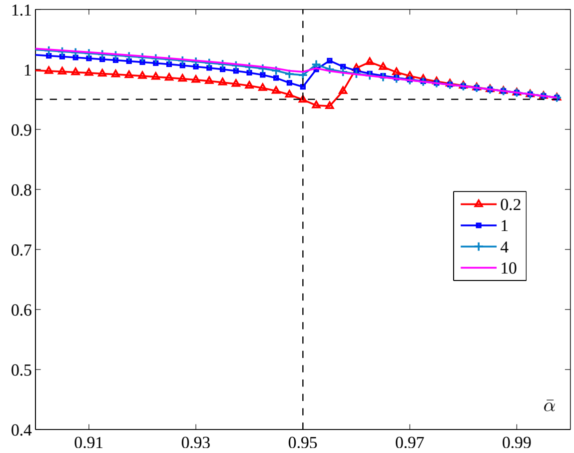


Figure 2(b): Mean of 2 step system GMM estimator (IV=S21, $\lambda = 0.2$)

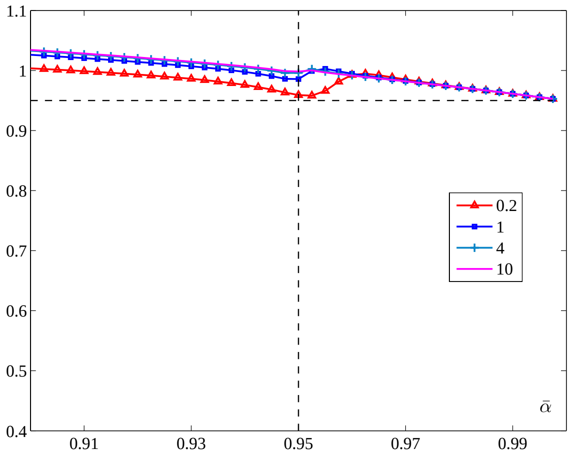


Figure 2(c): Mean of 1 step system GMM estimator (IV=S21, $\lambda = 1$)

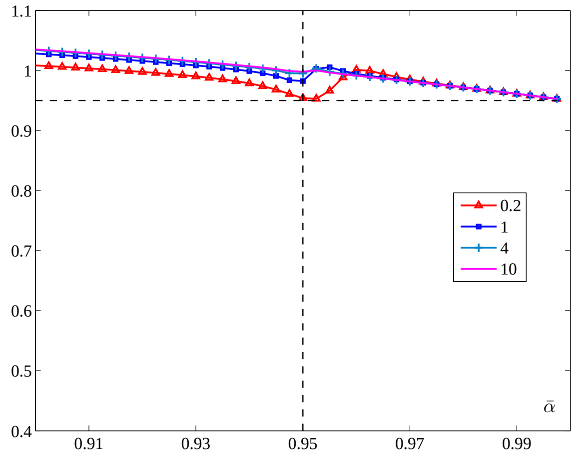


Figure 2(d): Mean of 2 step system GMM estimator (IV=S21, $\lambda = 1$)

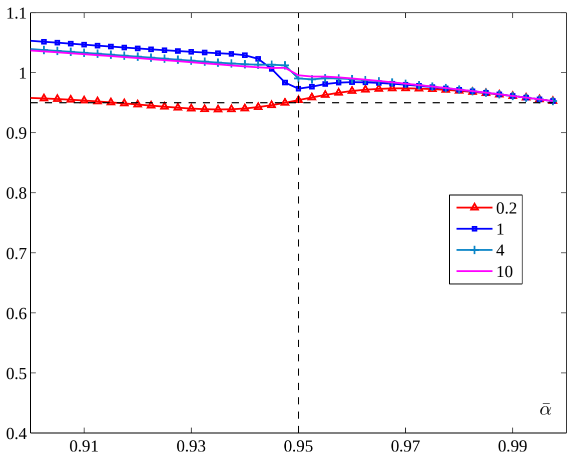


Figure 2(e): Mean of 1 step system GMM estimator (IV=S21, $\lambda = 5$)

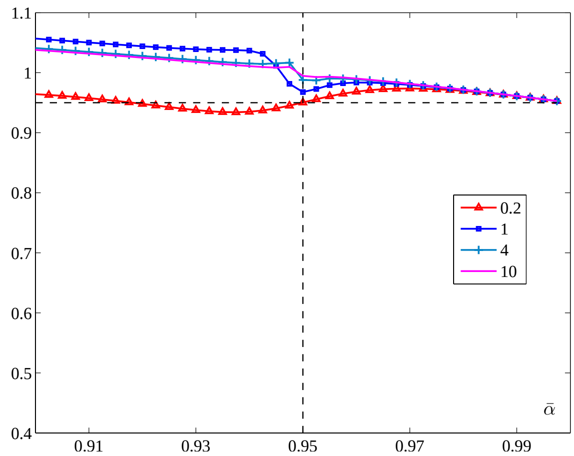


Figure 2(f): Mean of 2 step system GMM estimator (IV=S21, $\lambda = 5$)

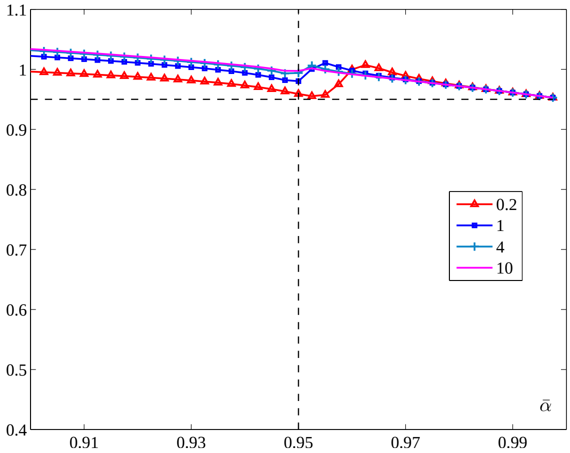


Figure 3(a): Mean of 1 step system GMM estimator (IV=S11, $\lambda = 0.2$)

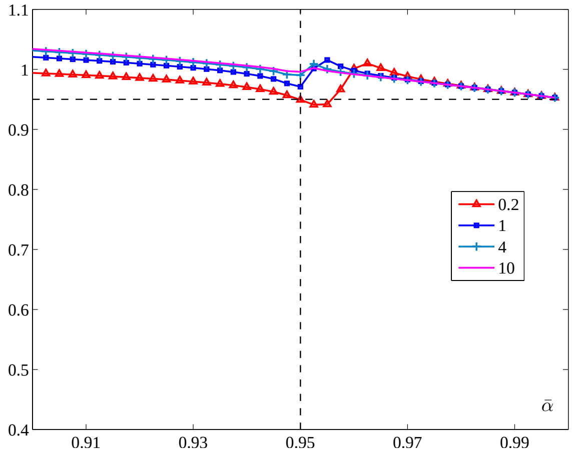


Figure 3(b): Mean of 2 step system GMM estimator (IV=S11, $\lambda = 0.2$)

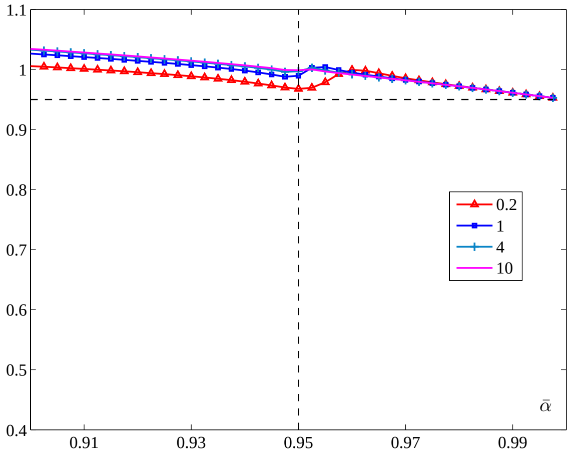


Figure 3(c): Mean of 1 step system GMM estimator (IV=S11, $\lambda = 1$)

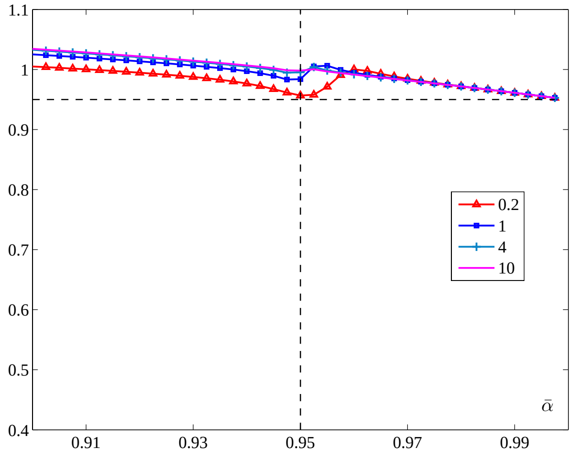


Figure 3(d): Mean of 2 step system GMM estimator (IV=S11, $\lambda = 1$)

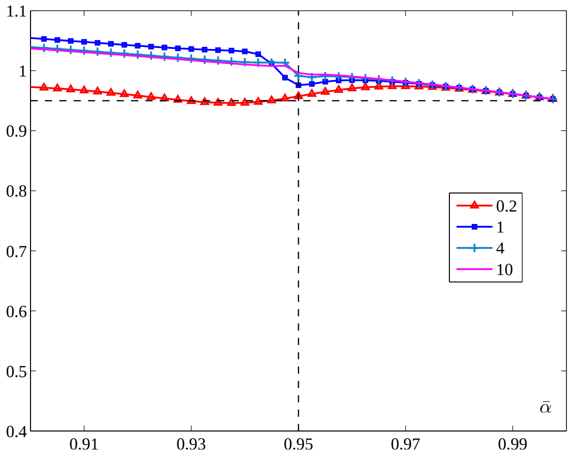


Figure 3(e): Mean of 1 step system GMM estimator (IV=S11, $\lambda = 5$)

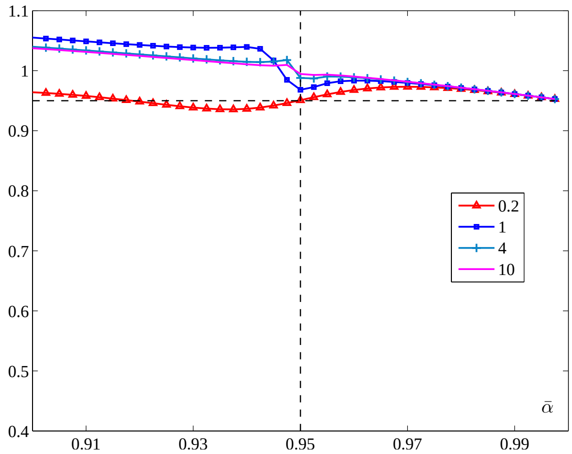


Figure 3(f): Mean of 2 step system GMM estimator (IV=S11, $\lambda = 5$)

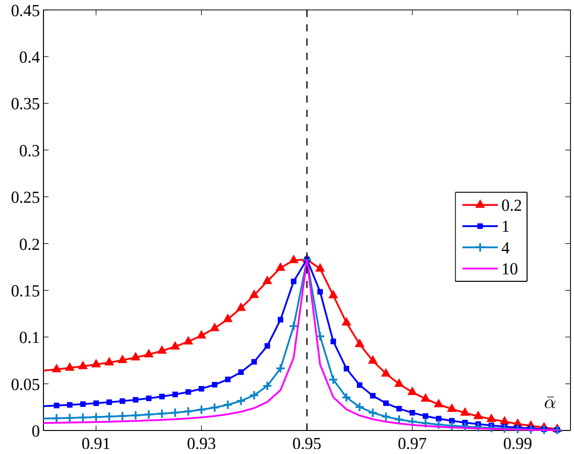


Figure 4(a): Standard deviation of the GMM estimator (IV=L2, $\lambda = 0.2$)

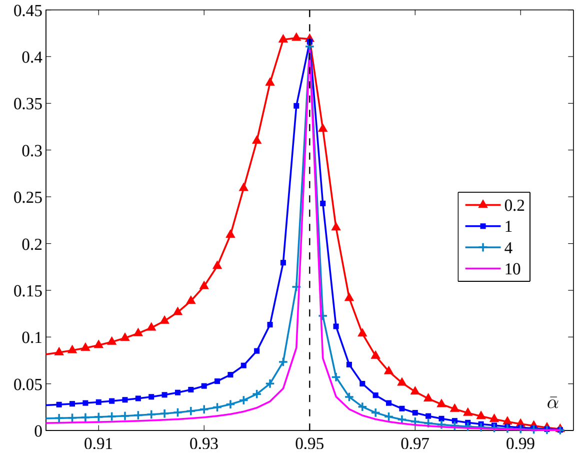


Figure 4(b): Standard deviation of the GMM estimator (IV=L1, $\lambda = 0.2$)

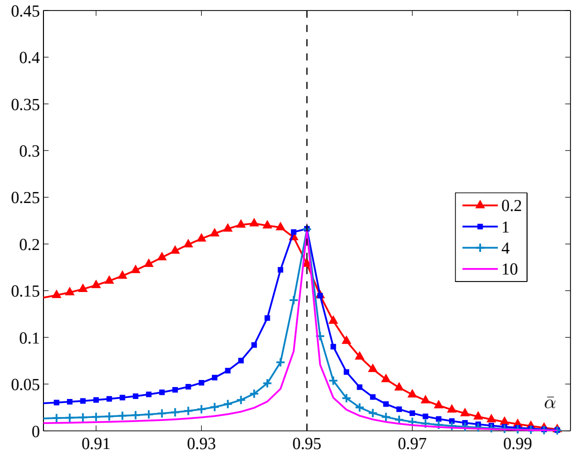


Figure 4(c): Standard deviation of the GMM estimator (IV=L2, $\lambda = 1$)

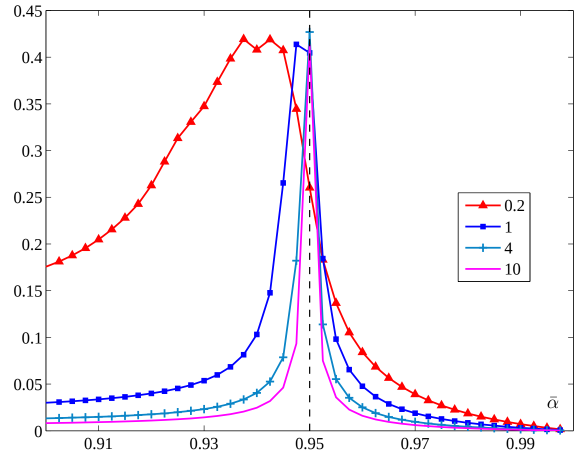


Figure 4(d): Standard deviation of the GMM estimator (IV=L1, $\lambda = 1$)

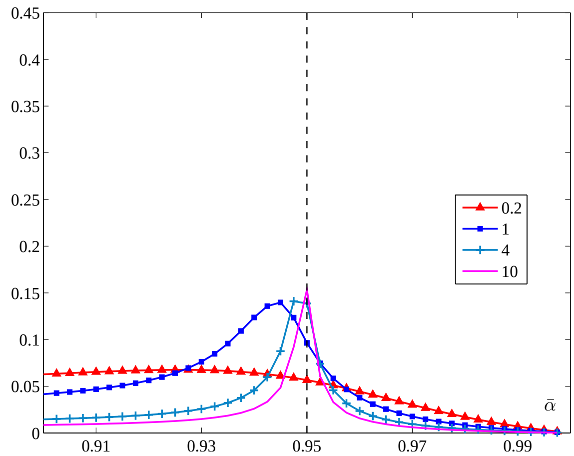


Figure 4(e): Standard deviation of the GMM estimator (IV=L2, $\lambda = 5$)

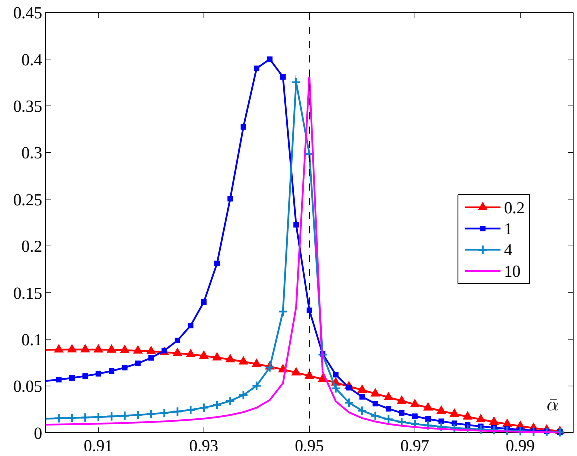


Figure 4(f): Standard deviation of the GMM estimator (IV=L1, $\lambda = 5$)

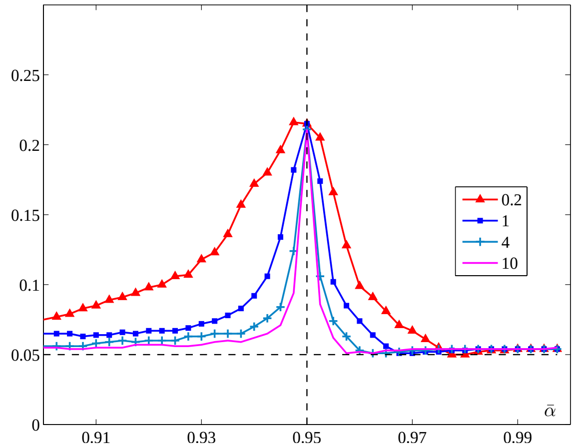


Figure 5(a): Empirical size of the GMM estimator (IV=L2, $\lambda = 0.2$)

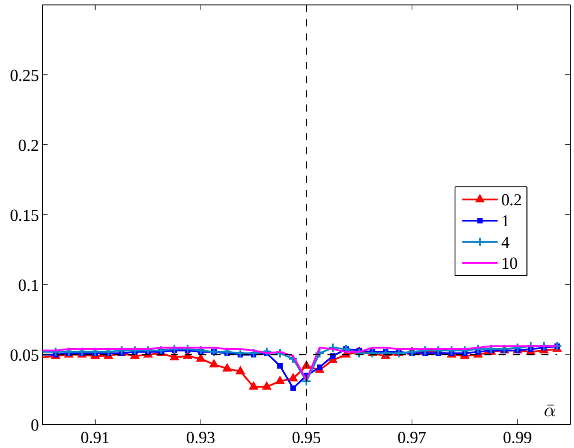


Figure 5(b): Empirical size of the GMM estimator (IV=L1, $\lambda = 0.2$)

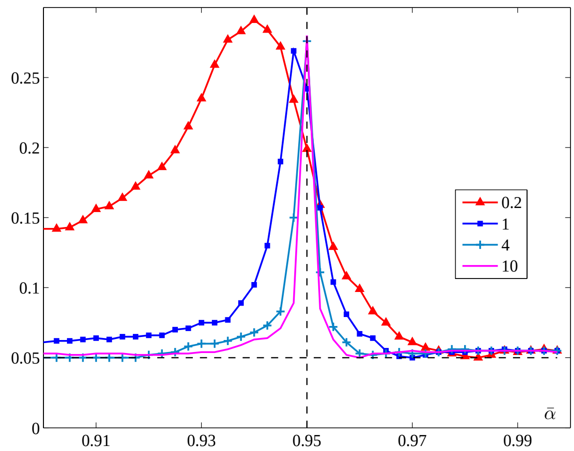


Figure 5(c): Empirical size of the GMM estimator (IV=L2, $\lambda = 1$)

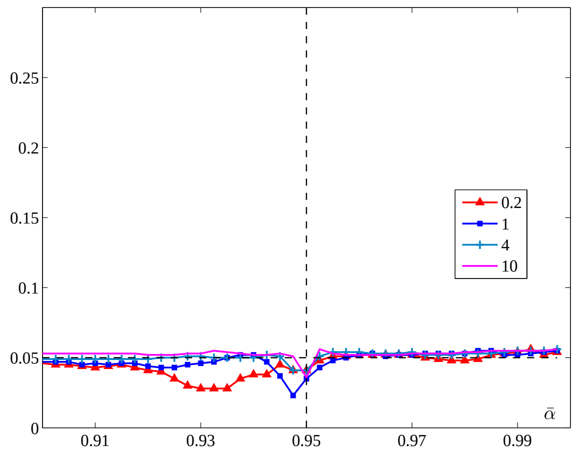


Figure 5(d): Empirical size of the GMM estimator (IV=L1, $\lambda = 1$)

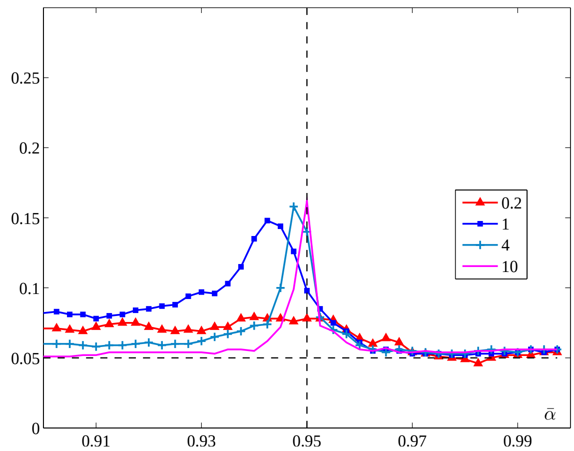


Figure 5(e): Empirical size of the GMM estimator (IV=L2, $\lambda = 5$)

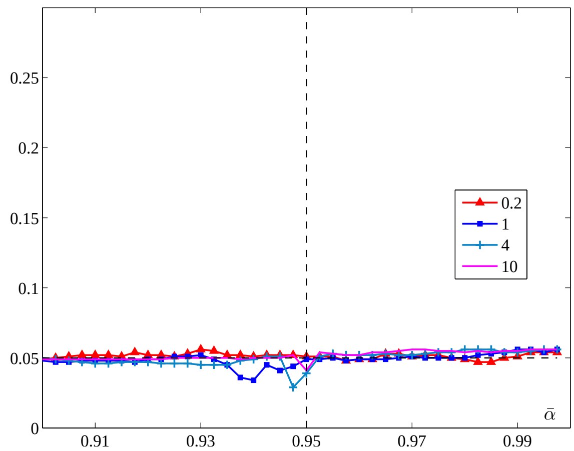


Figure 5(f): Empirical size of the GMM estimator (IV=L1, $\lambda = 5$)

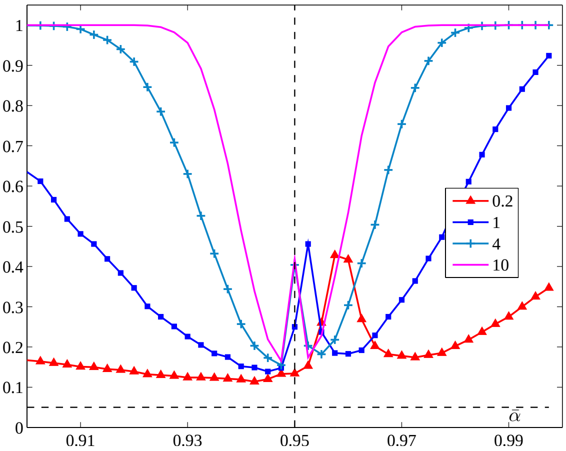


Figure 6(a):Power plot of mean-stationarity test (IV=$S21-L2$, $\lambda=0.2$)

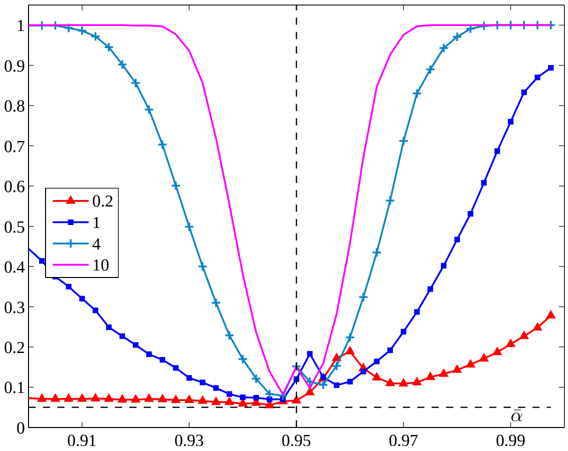


Figure 6(b):Power plot of mean-stationarity test (IV=$S11-L1$, $\lambda=0.2$)

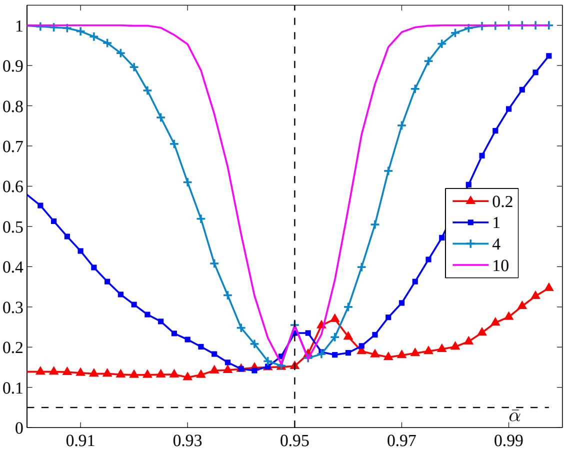


Figure 6(c):Power plot of mean-stationarity test (IV=$S21-L2$, $\lambda=1$)

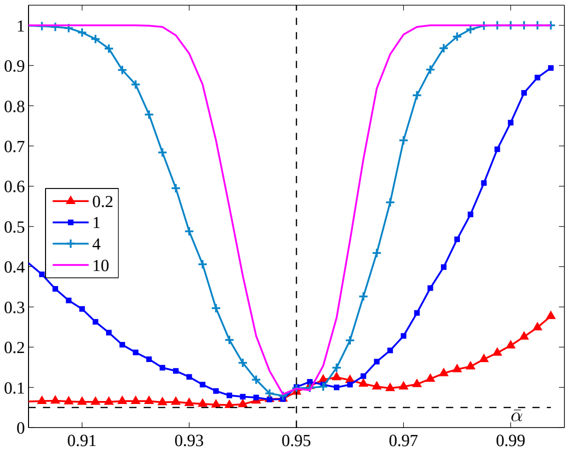


Figure 6(d):Power plot of mean-stationarity test (IV=$S11-L1$, $\lambda=1$)

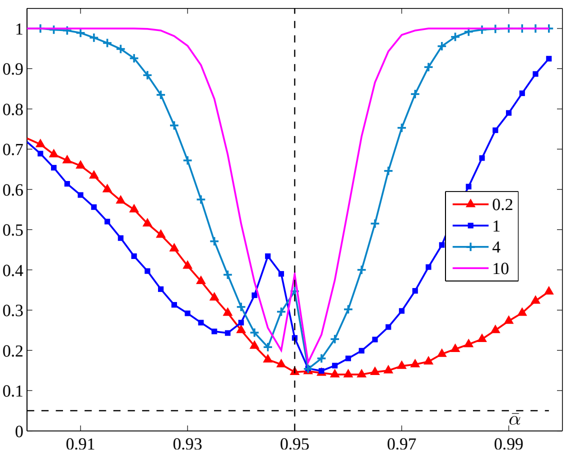


Figure 6(e):Power plot of mean-stationarity test (IV=$S21-L2$, $\lambda=5$)

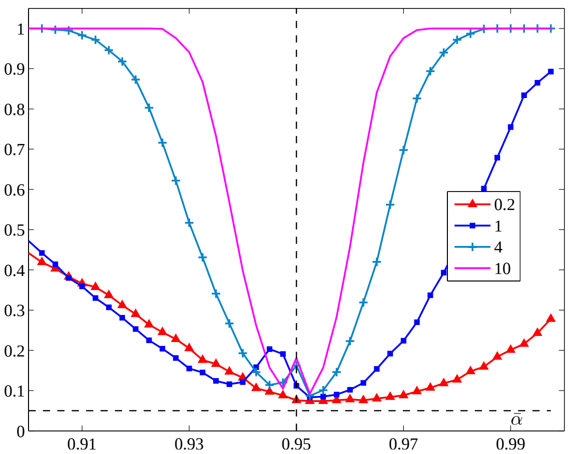


Figure 6(f):Power plot of mean-stationarity test (IV=$S11-L1$, $\lambda=5$)

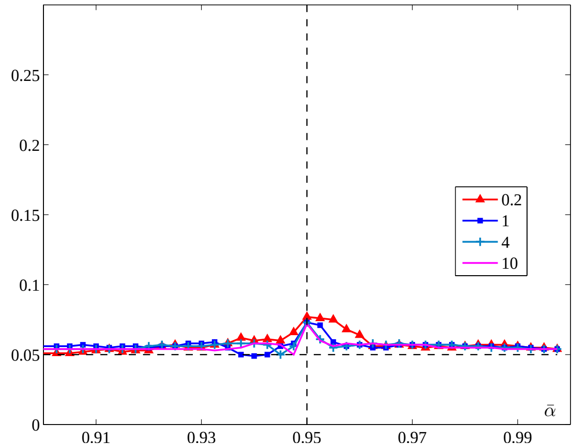


Figure 7(a): Empirical size of K statistic (IV=L2, $\lambda = 0.2$)

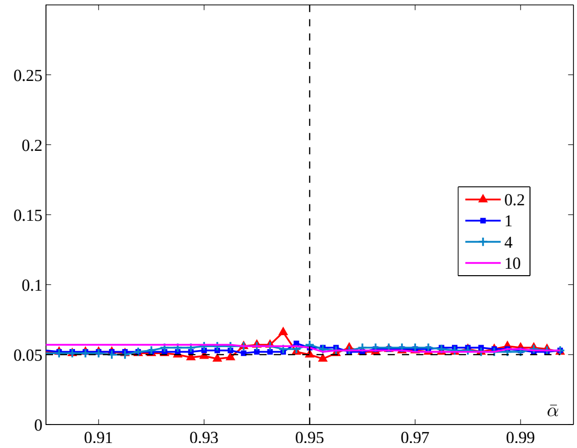


Figure 7(b): Empirical size of K statistic (IV=L1, $\lambda = 0.2$)

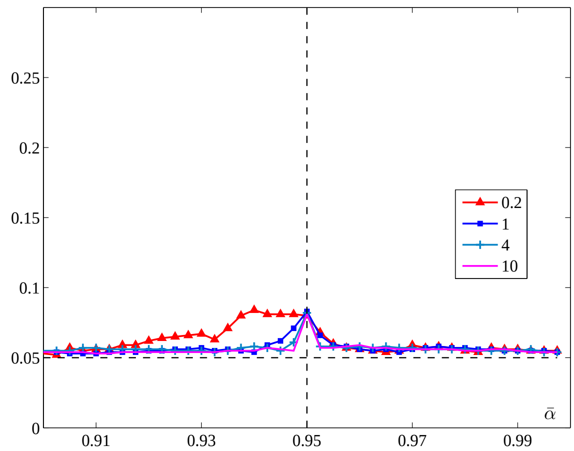


Figure 7(c): Empirical size of K statistic (IV=L2, $\lambda = 1$)

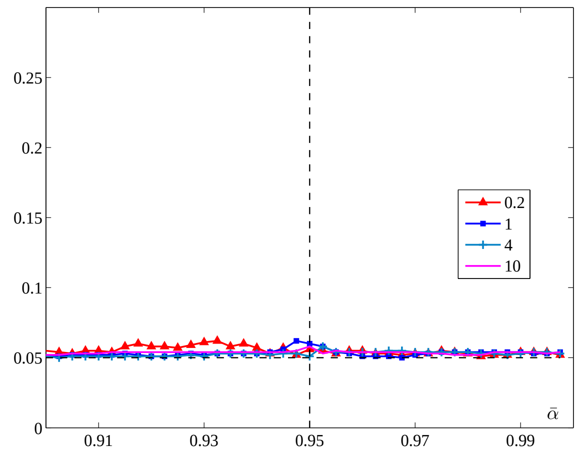


Figure 7(d): Empirical size of K statistic (IV=L1, $\lambda = 1$)

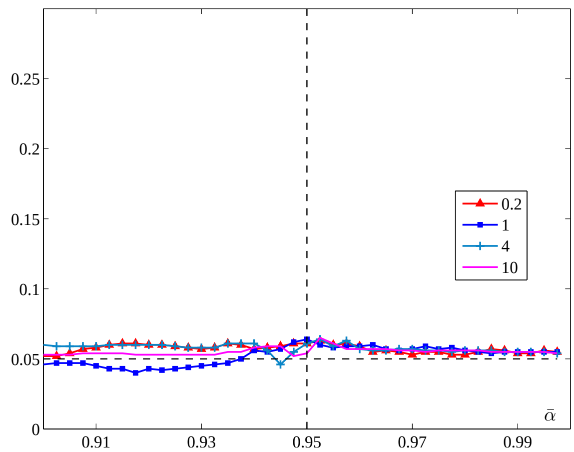


Figure 7(e): Empirical size of K statistic (IV=L2, $\lambda = 5$)

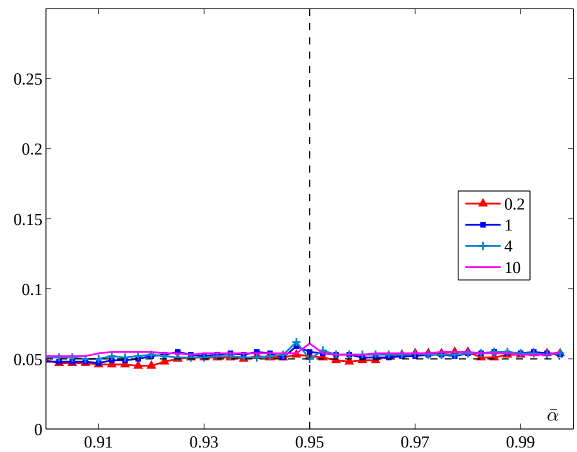


Figure 7(f): Empirical size of K statistic (IV=L1, $\lambda = 5$)

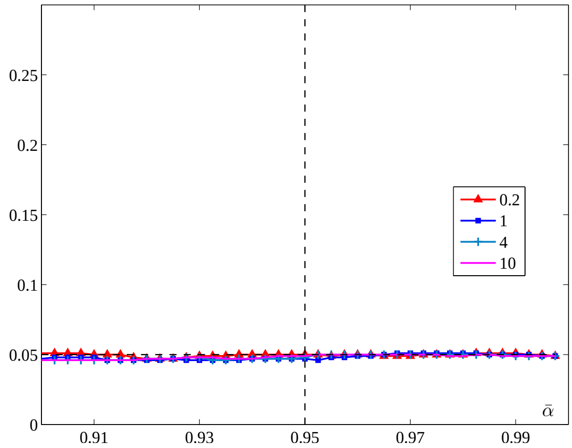


Figure 8(a): Empirical size of S statistic (IV=L2, $\lambda = 0.2$)

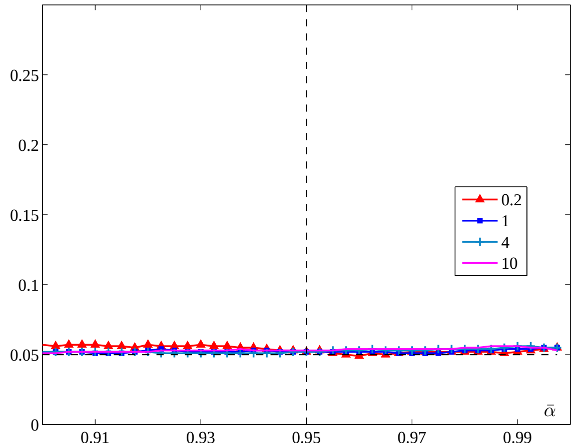


Figure 8(b): Empirical size of S statistic (IV=L1, $\lambda = 0.2$)

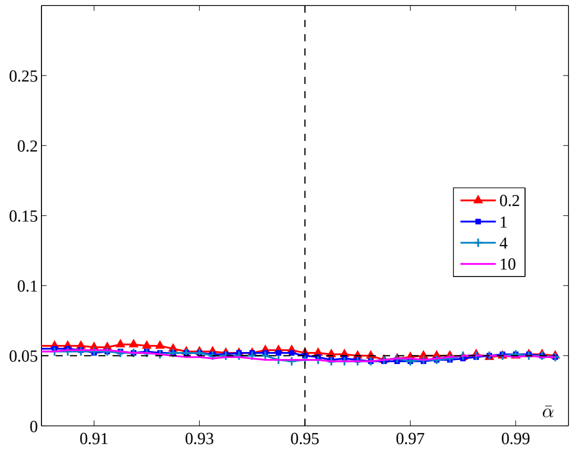


Figure 8(c): Empirical size of S statistic (IV=L2, $\lambda = 1$)

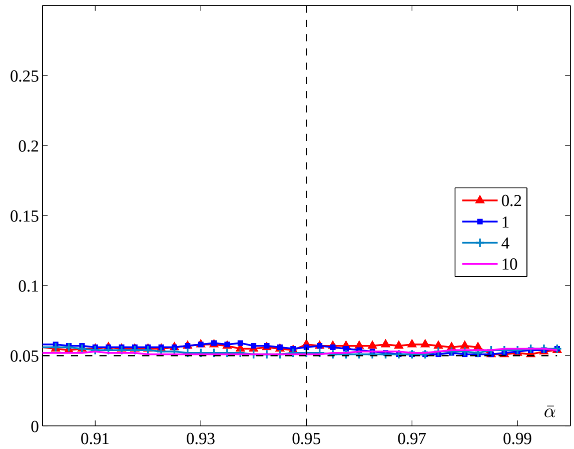


Figure 8(d): Empirical size of S statistic (IV=L1, $\lambda = 1$)

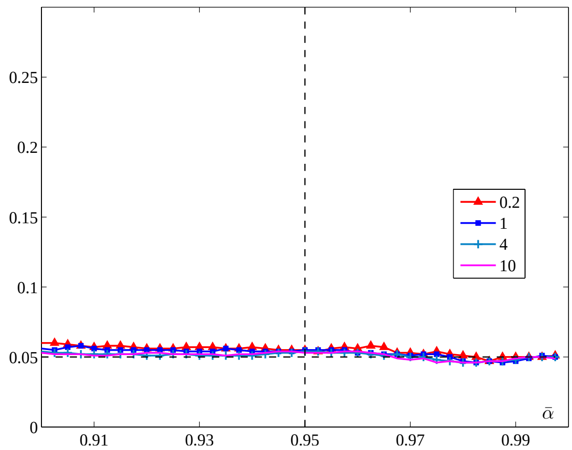


Figure 8(e): Empirical size of S statistic (IV=L2, $\lambda = 5$)

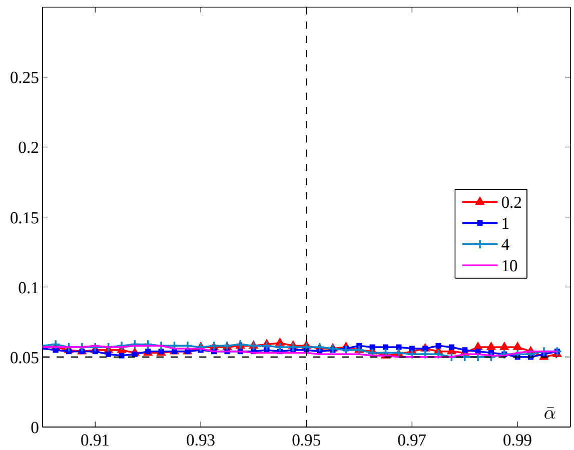


Figure 8(f): Empirical size of S statistic (IV=L1, $\lambda = 5$)

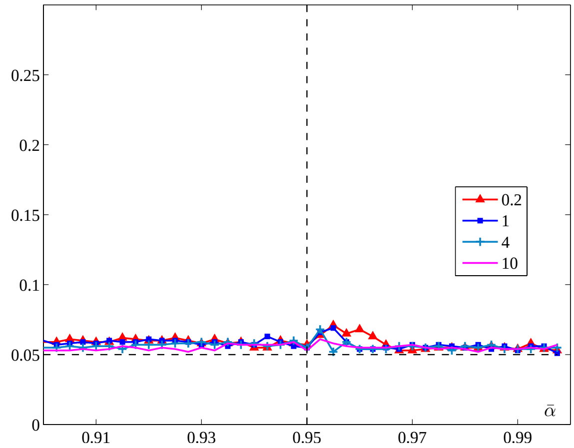


Figure 9(a): Empirical size of M statistic (IV=L2, $\lambda = 0.2$)

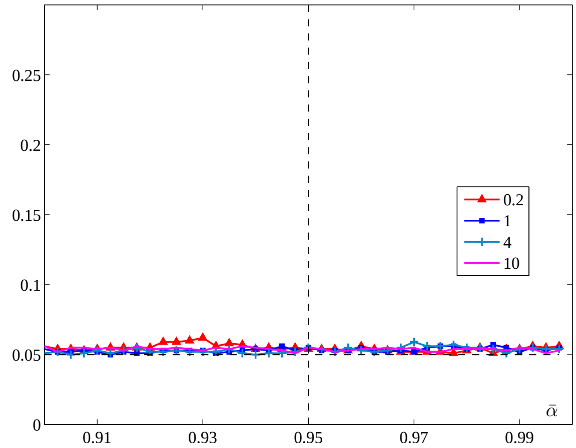


Figure 9(b): Empirical size of M statistic (IV=L1, $\lambda = 0.2$)

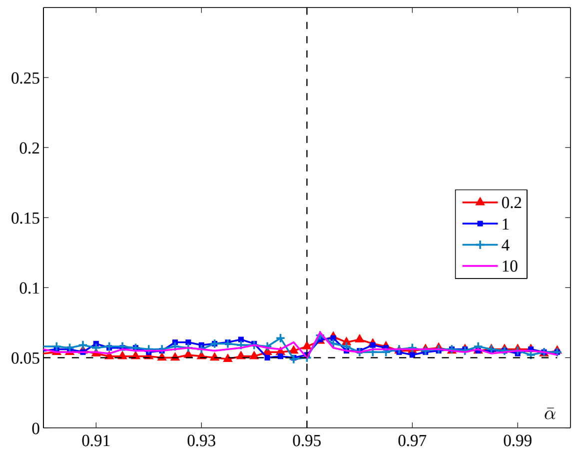


Figure 9(c): Empirical size of M statistic (IV=L2, $\lambda = 1$)

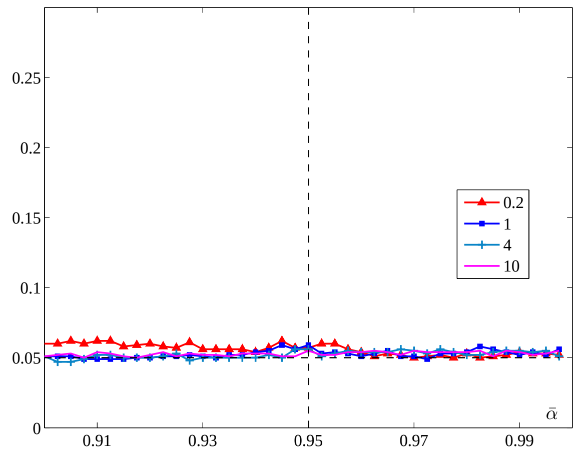


Figure 9(d): Empirical size of M statistic (IV=L1, $\lambda = 1$)

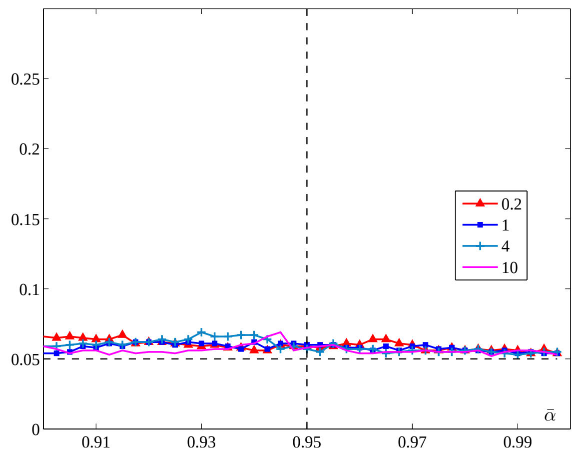


Figure 9(e): Empirical size of M statistic (IV=L2, $\lambda = 5$)

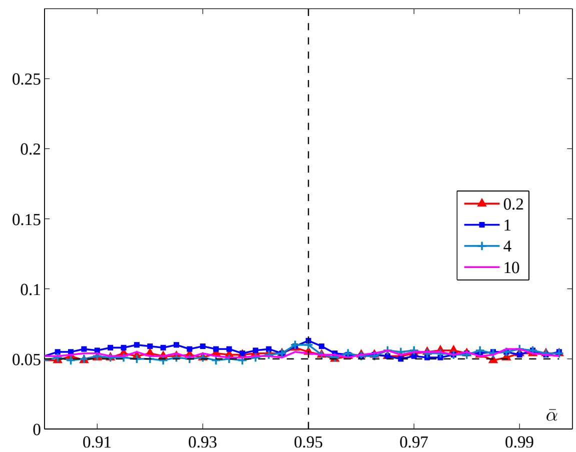


Figure 9(f): Empirical size of M statistic (IV=L1, $\lambda = 5$)

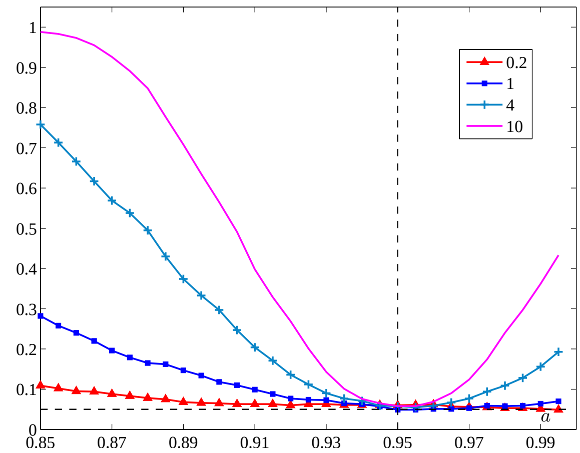


Figure 10(a): Power plot of K statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.94$)

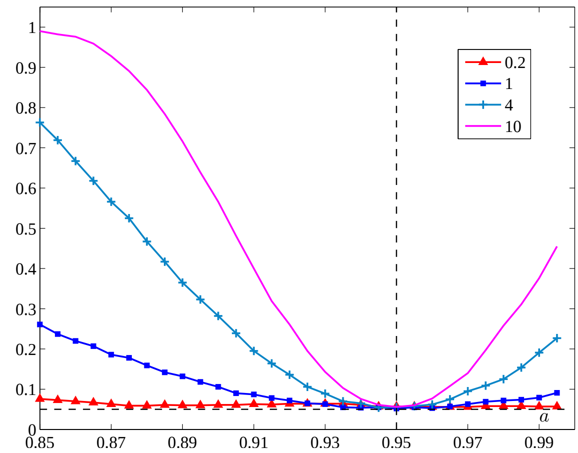


Figure 10(b): Power plot of K statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.94$)

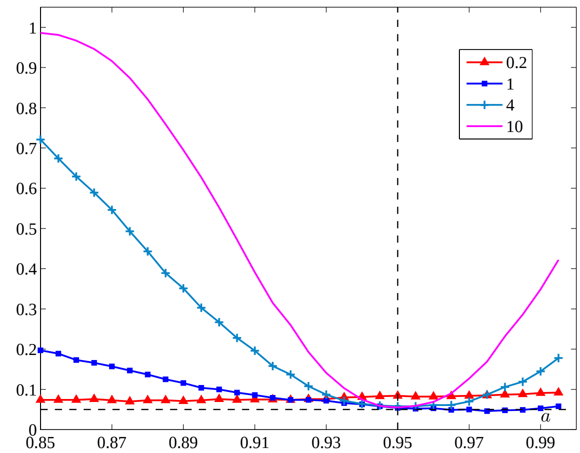


Figure 10(c): Power plot of K statistic ($\lambda = 1$, $\bar{\alpha} = 0.94$)

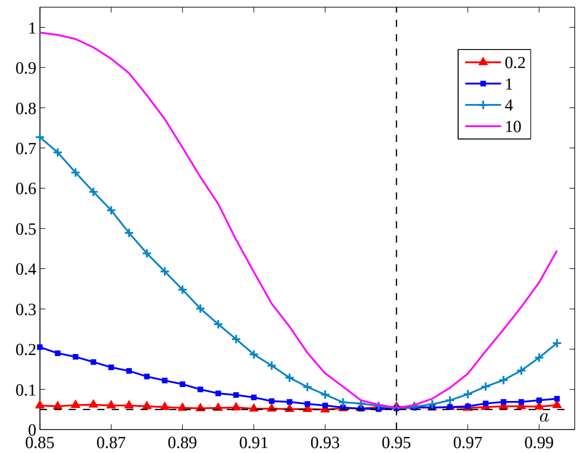


Figure 10(d): Power plot of K statistic ($\lambda = 1$, $\bar{\alpha} = 0.94$)

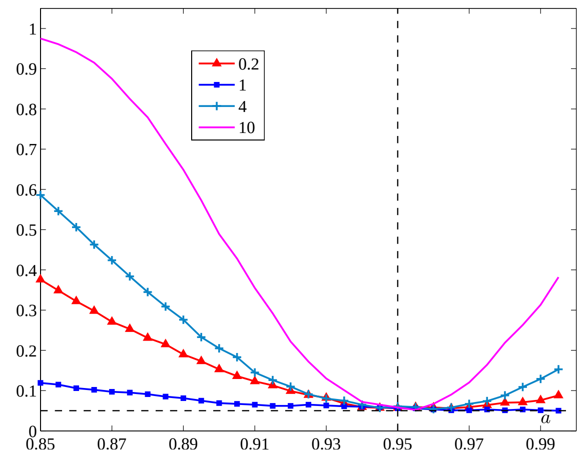


Figure 10(e): Power plot of K statistic ($\lambda = 5$, $\bar{\alpha} = 0.94$)

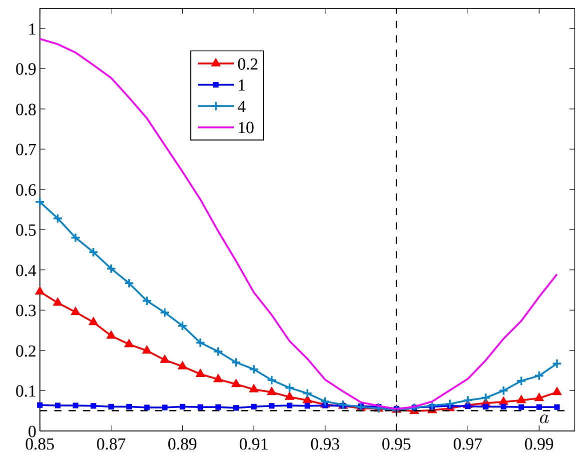


Figure 10(f): Power plot of K statistic ($\lambda = 5$, $\bar{\alpha} = 0.94$)

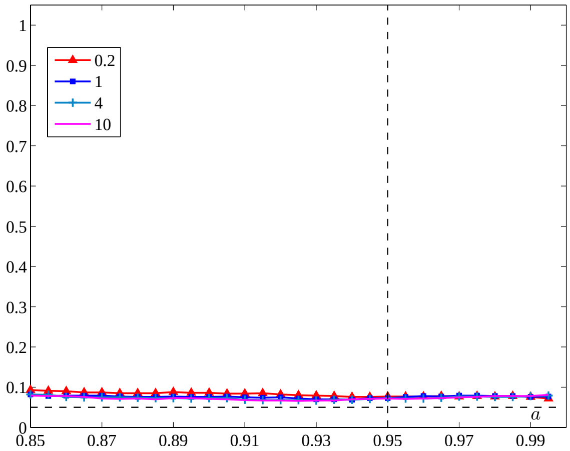


Figure 11(a): Power plot of K statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.95$)

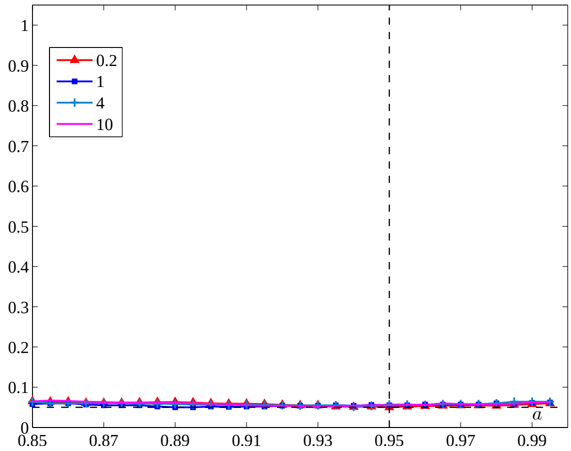


Figure 11(b): Power plot of K statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.95$)

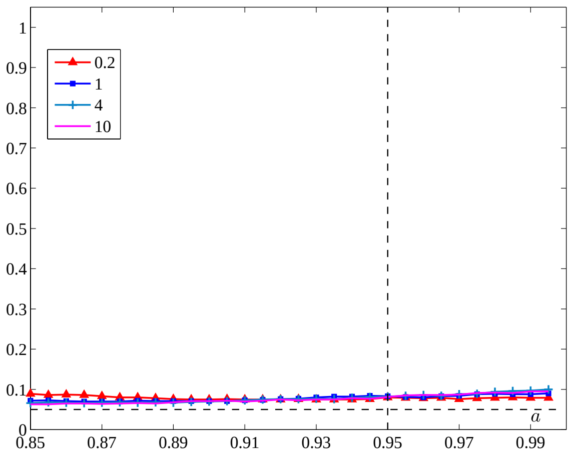


Figure 11(c): Power plot of K statistic
($\lambda = 1$, $\bar{\alpha} = 0.95$)

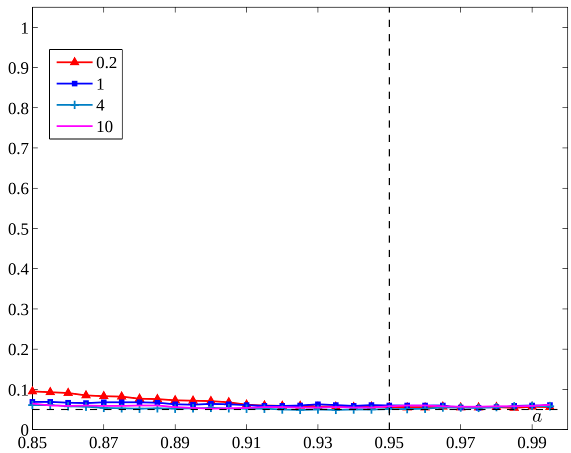


Figure 11(d): Power plot of K statistic
($\lambda = 1$, $\bar{\alpha} = 0.95$)

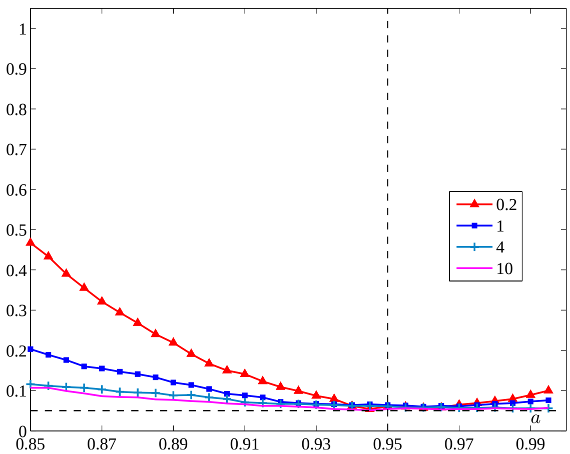


Figure 11(e): Power plot of K statistic
($\lambda = 5$, $\bar{\alpha} = 0.95$)

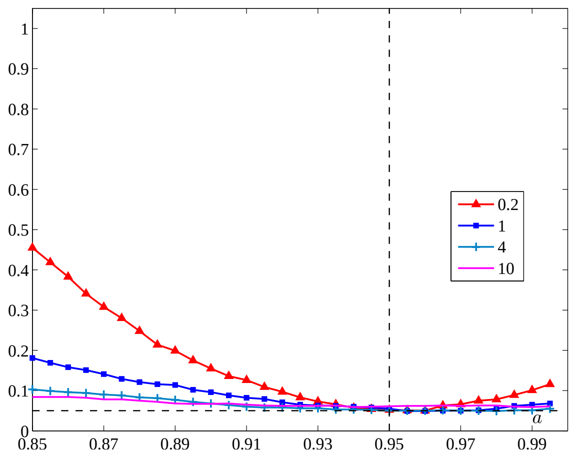


Figure 11(f): Power plot of K statistic
($\lambda = 5$, $\bar{\alpha} = 0.95$)

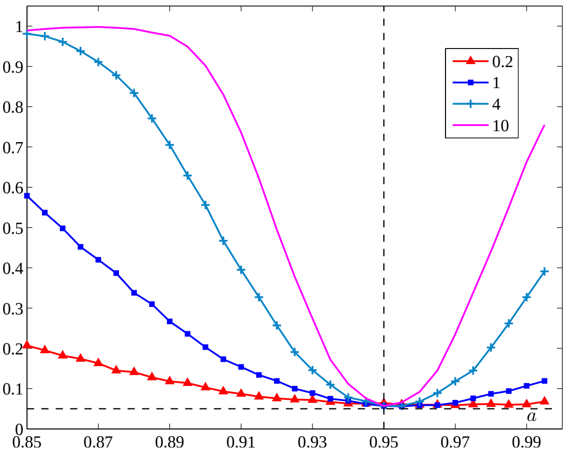


Figure 12(a): Power plot of K statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.96$)

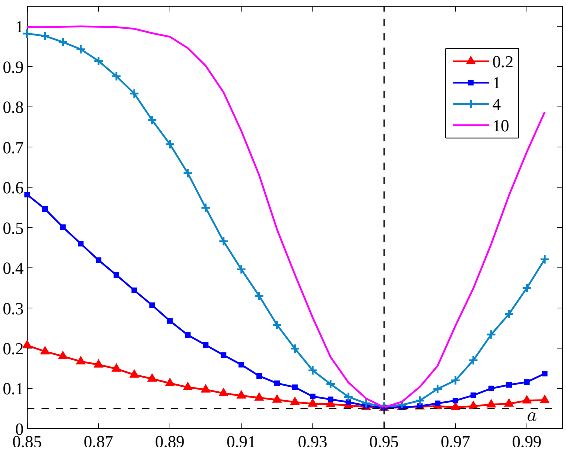


Figure 12(b): Power plot of K statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.96$)

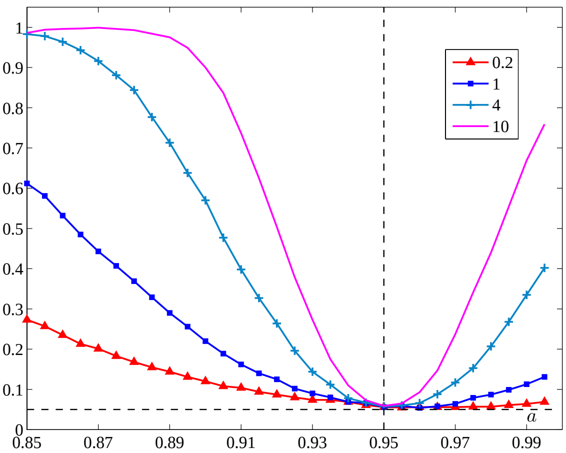


Figure 12(c): Power plot of K statistic
($\lambda = 1$, $\bar{\alpha} = 0.96$)

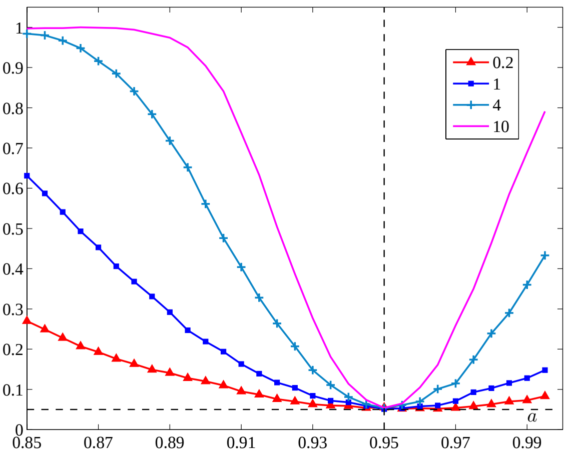


Figure 12(d): Power plot of K statistic
($\lambda = 1$, $\bar{\alpha} = 0.96$)

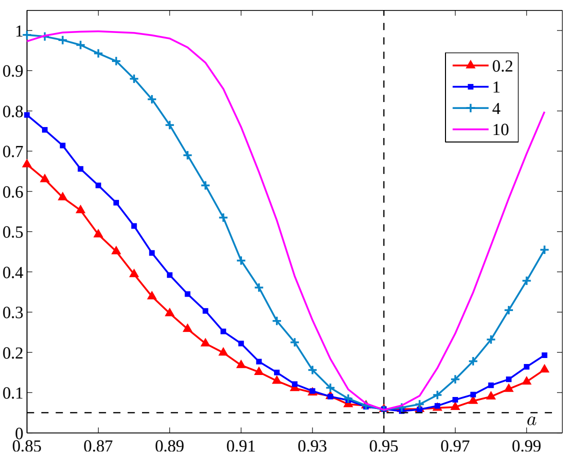


Figure 12(e): Power plot of K statistic
($\lambda = 5$, $\bar{\alpha} = 0.96$)

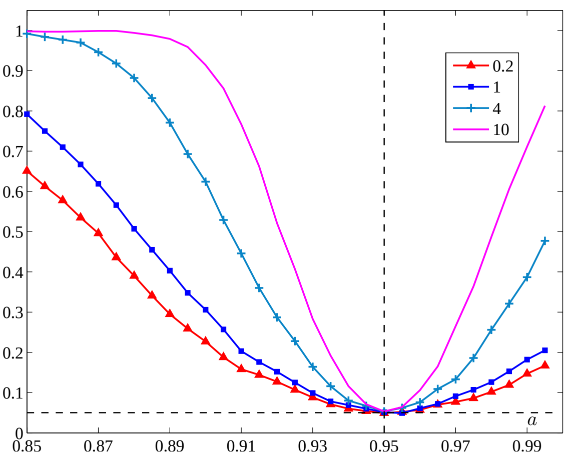
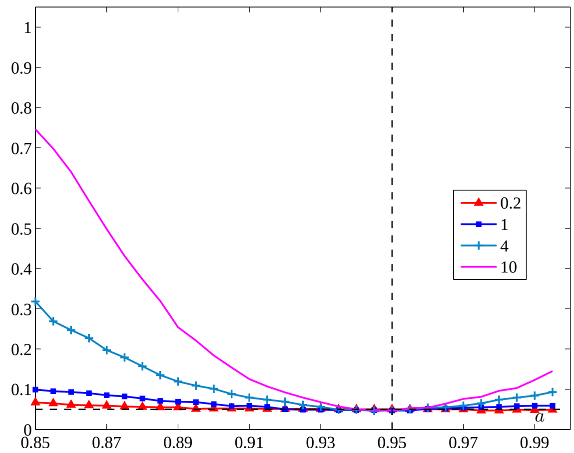


Figure 12(f): Power plot of K statistic
($\lambda = 5$, $\bar{\alpha} = 0.96$)



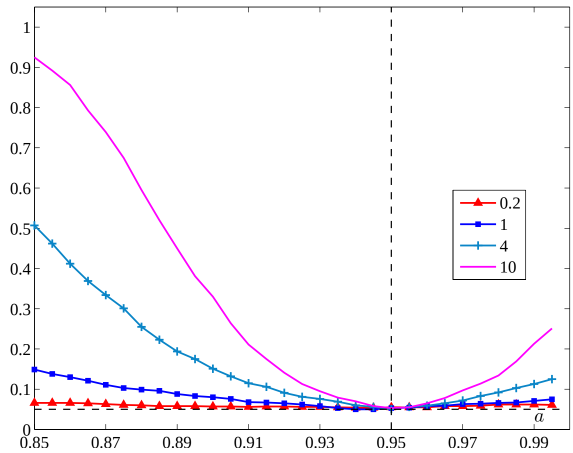
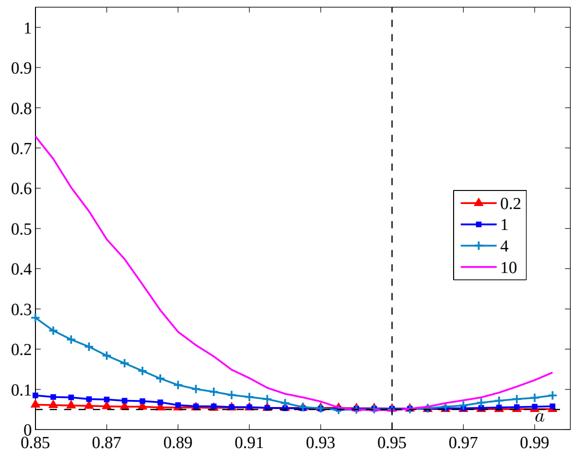


Figure 13(a): Power plot of S statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.94$)

Figure 13(b): Power plot of S statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.94$)



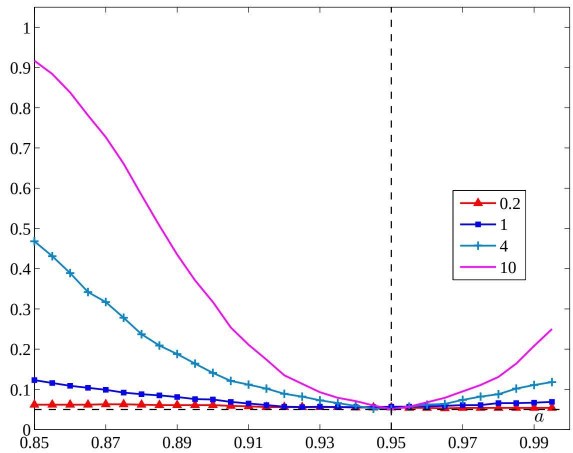
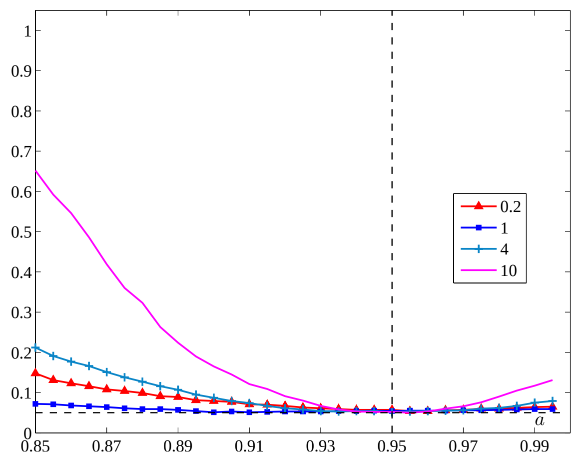


Figure 13(c): Power plot of S statistic ($\lambda = 1$, $\bar{\alpha} = 0.94$)

Figure 13(d): Power plot of S statistic ($\lambda = 1$, $\bar{\alpha} = 0.94$)



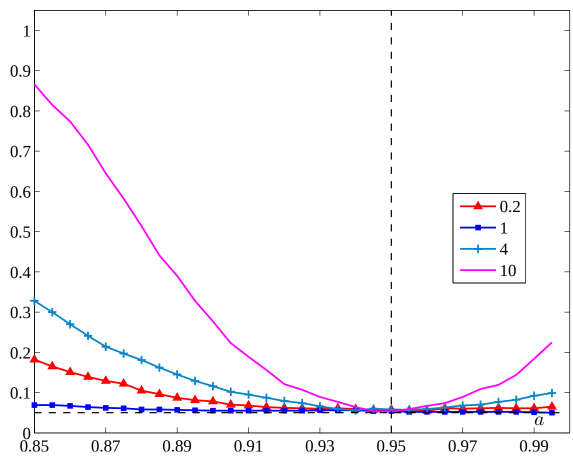


Figure 13(e): Power plot of S statistic ($\lambda = 5$, $\bar{\alpha} = 0.94$)

Figure 13(f): Power plot of S statistic ($\lambda = 5$, $\bar{\alpha} = 0.94$)

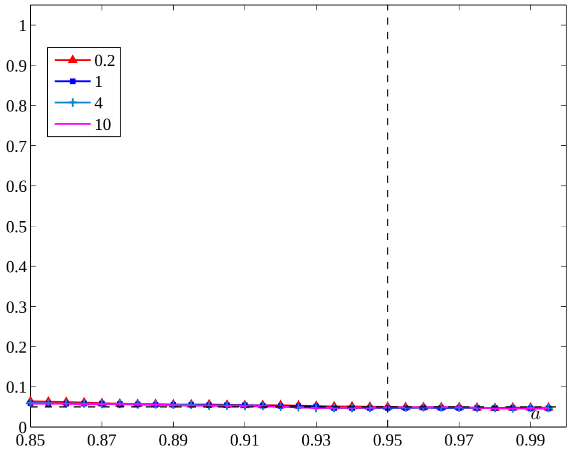


Figure 14(a): Power plot of S statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.95$)

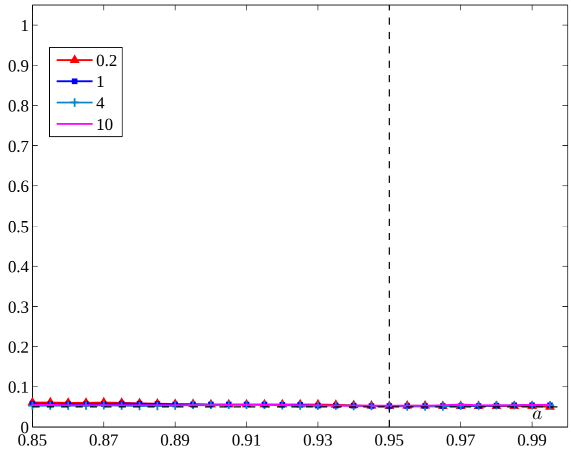


Figure 14(b): Power plot of S statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.95$)

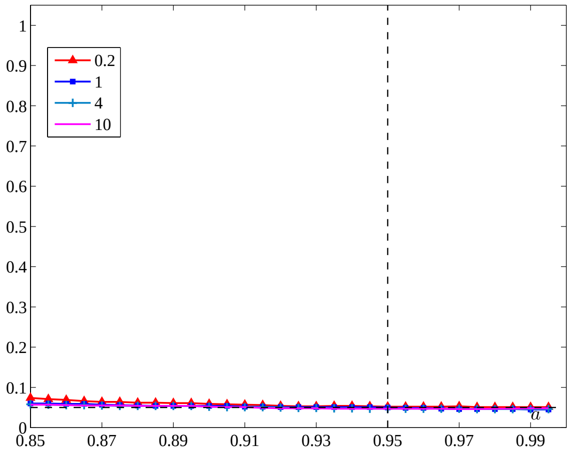


Figure 14(c): Power plot of S statistic
($\lambda = 1$, $\bar{\alpha} = 0.95$)

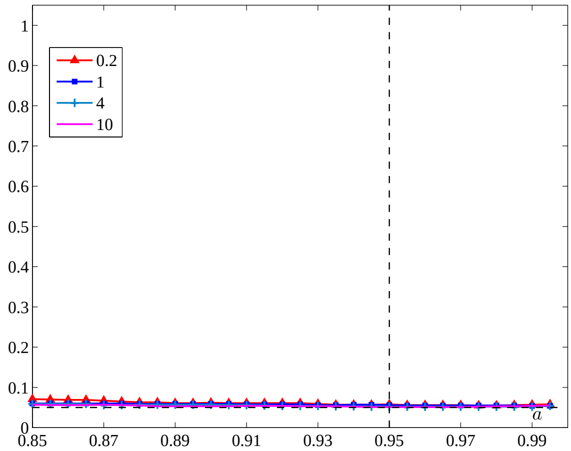


Figure 14(d): Power plot of S statistic
($\lambda = 1$, $\bar{\alpha} = 0.95$)

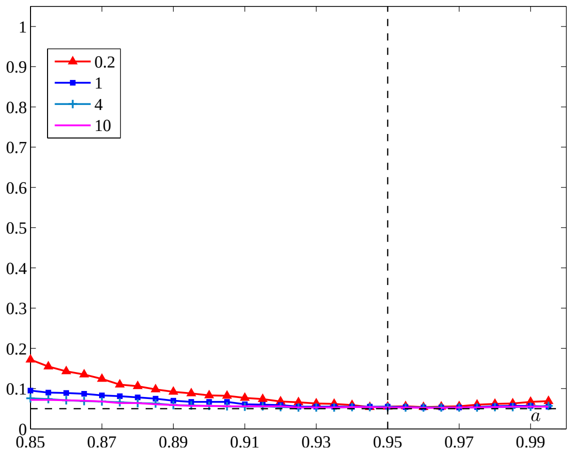


Figure 14(e): Power plot of S statistic
($\lambda = 5$, $\bar{\alpha} = 0.95$)

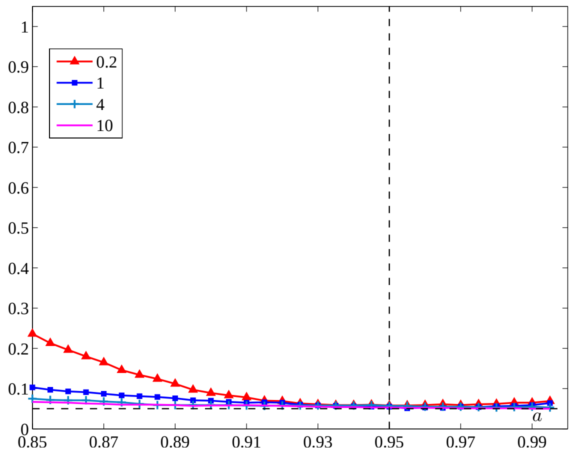


Figure 14(f): Power plot of S statistic
($\lambda = 5$, $\bar{\alpha} = 0.95$)

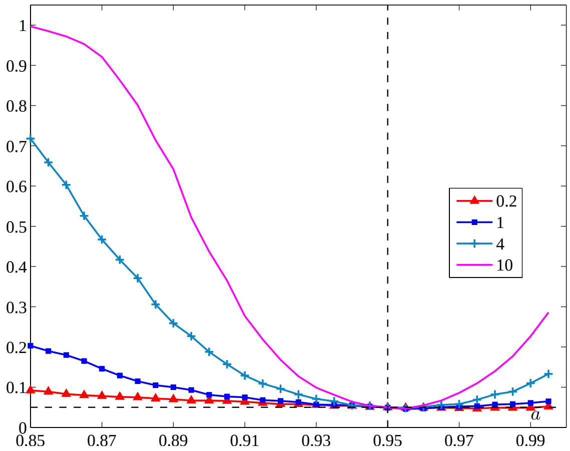


Figure 15(a): Power plot of S statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.96$)

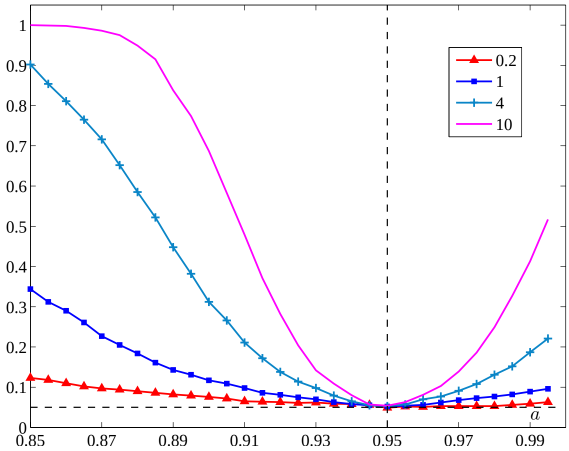


Figure 15(b): Power plot of S statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.96$)

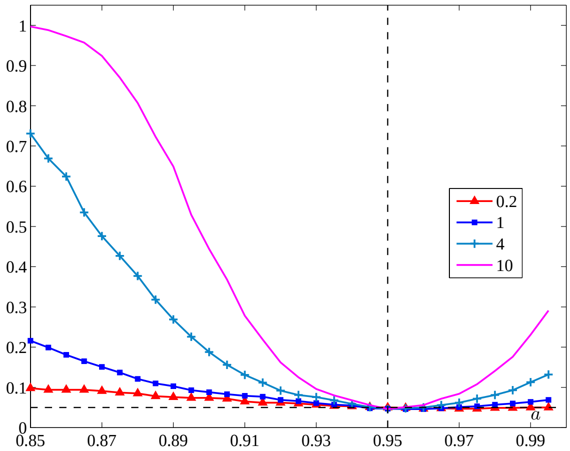


Figure 15(c): Power plot of S statistic ($\lambda = 1$, $\bar{\alpha} = 0.96$)

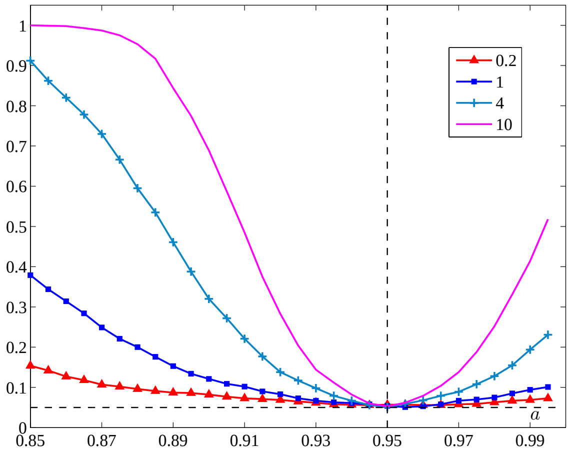


Figure 15(d): Power plot of S statistic ($\lambda = 1$, $\bar{\alpha} = 0.96$)

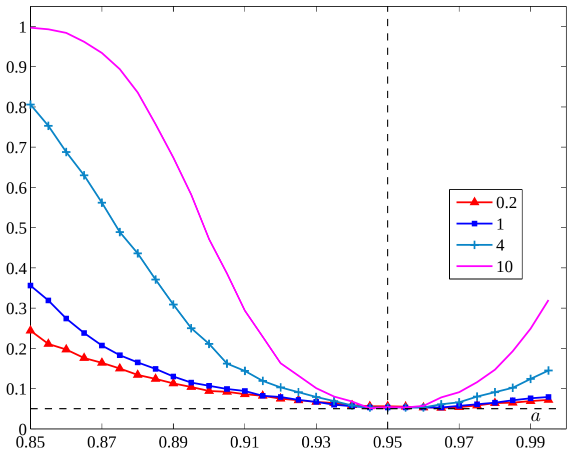


Figure 15(e): Power plot of S statistic ($\lambda = 5$, $\bar{\alpha} = 0.96$)

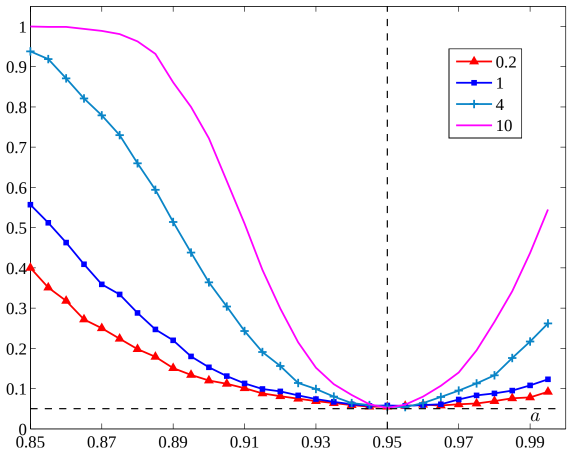


Figure 15(f): Power plot of S statistic ($\lambda = 5$, $\bar{\alpha} = 0.96$)

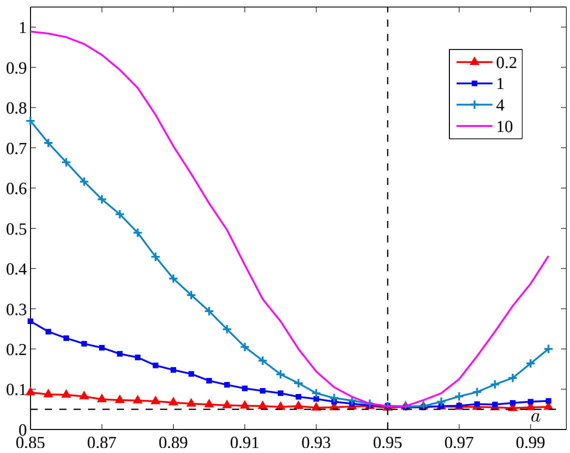


Figure 16(a): Power plot of M statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.94$)

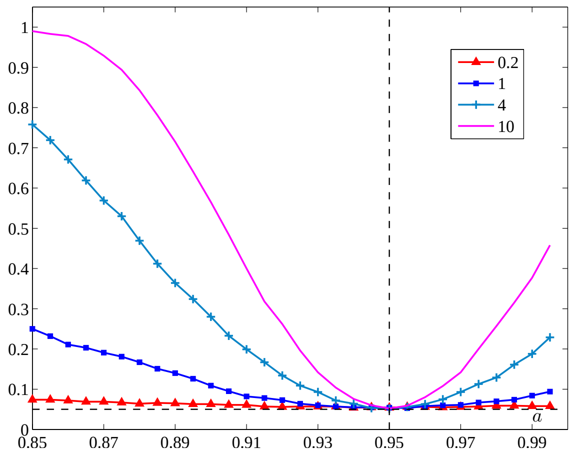


Figure 16(b): Power plot of M statistic ($\lambda = 0.2$, $\bar{\alpha} = 0.94$)

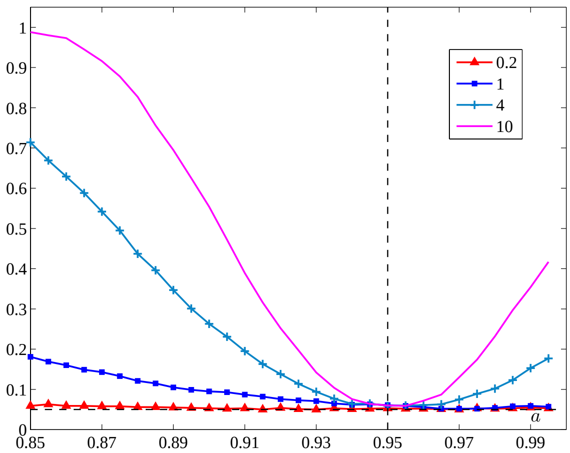


Figure 16(c): Power plot of M statistic ($\lambda = 1$, $\bar{\alpha} = 0.94$)

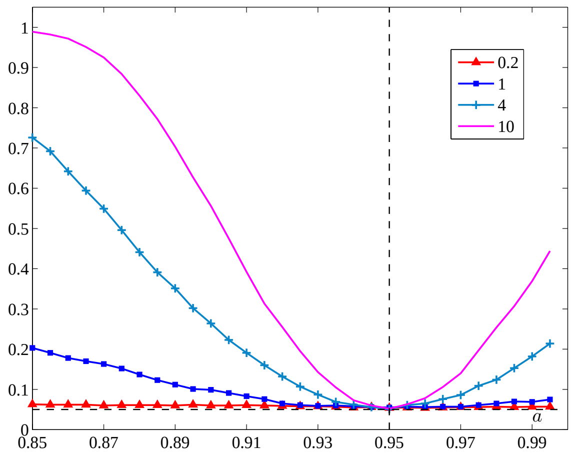


Figure 16(d): Power plot of M statistic ($\lambda = 1$, $\bar{\alpha} = 0.94$)

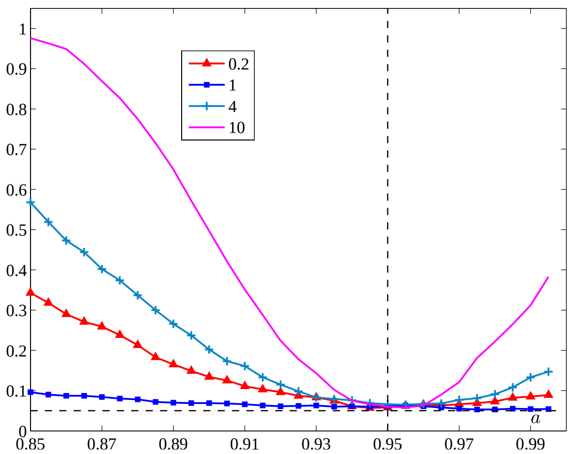


Figure 16(e): Power plot of M statistic ($\lambda = 5$, $\bar{\alpha} = 0.94$)

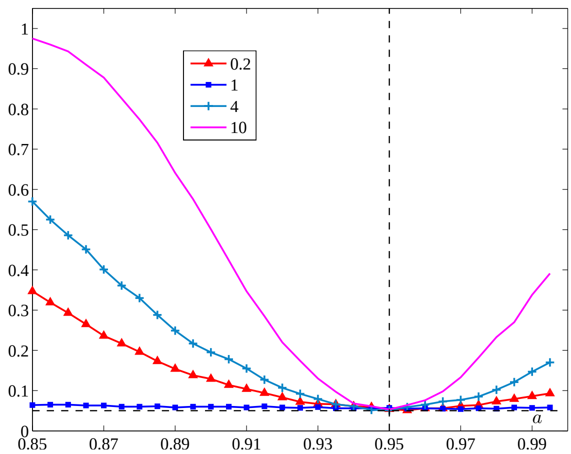


Figure 16(f): Power plot of M statistic ($\lambda = 5$, $\bar{\alpha} = 0.94$)

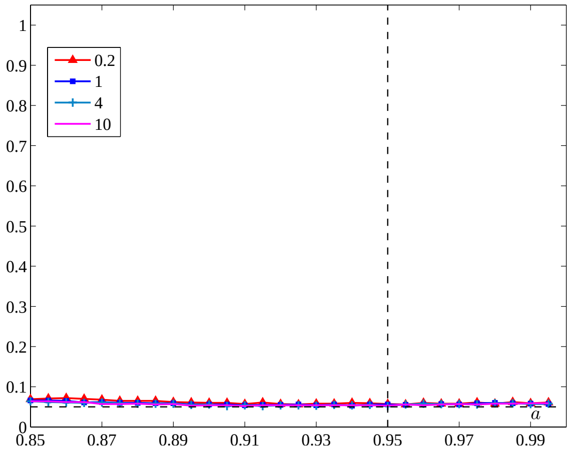


Figure 17(a): Power plot of M statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.95$)

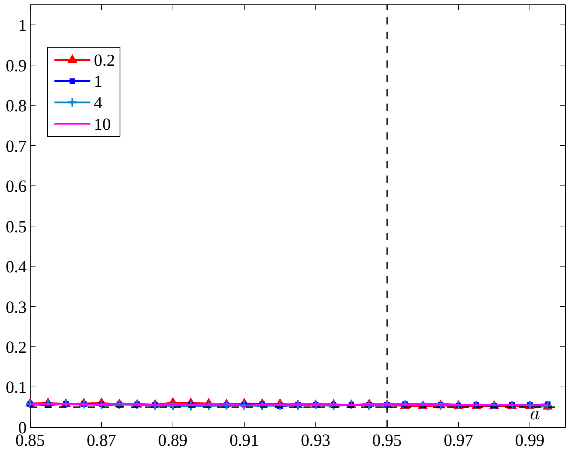


Figure 17(b): Power plot of M statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.95$)

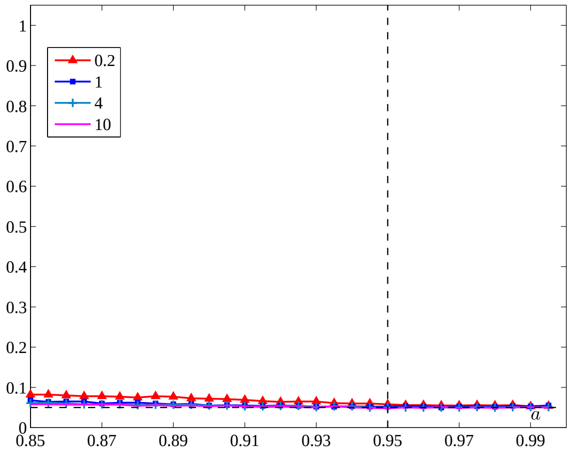


Figure 17(c): Power plot of M statistic
($\lambda = 1$, $\bar{\alpha} = 0.95$)

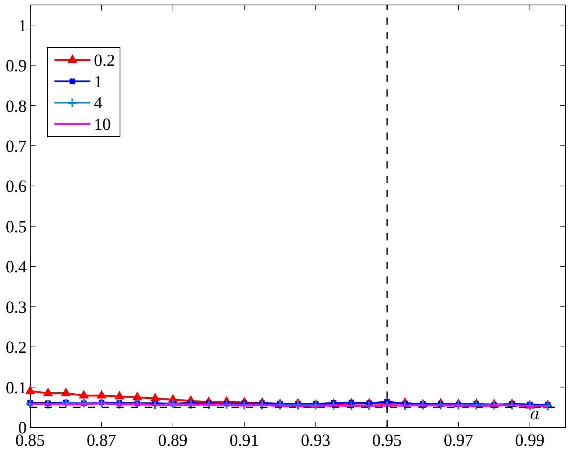


Figure 17(d): Power plot of M statistic
($\lambda = 1$, $\bar{\alpha} = 0.95$)

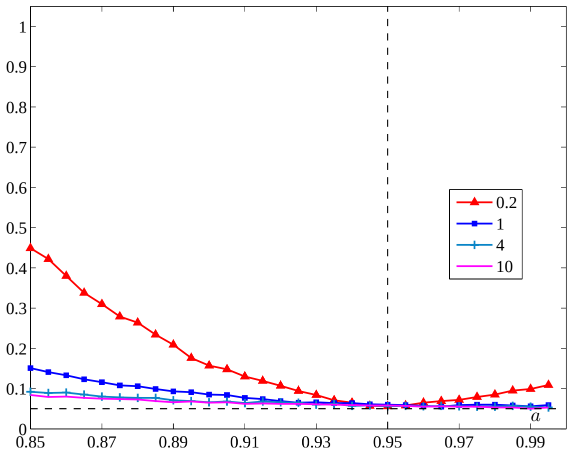


Figure 17(e): Power plot of M statistic
($\lambda = 5$, $\bar{\alpha} = 0.95$)

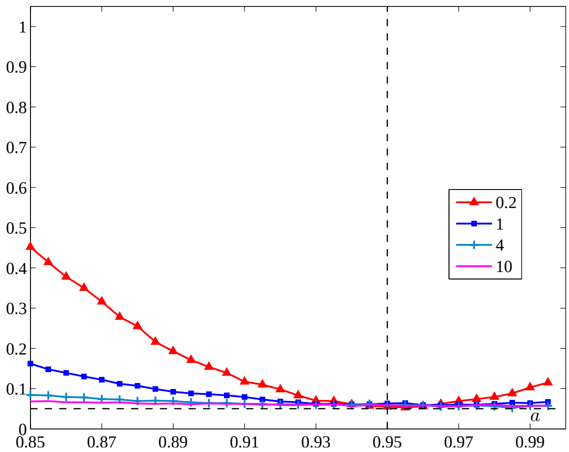


Figure 17(f): Power plot of M statistic
($\lambda = 5$, $\bar{\alpha} = 0.95$)

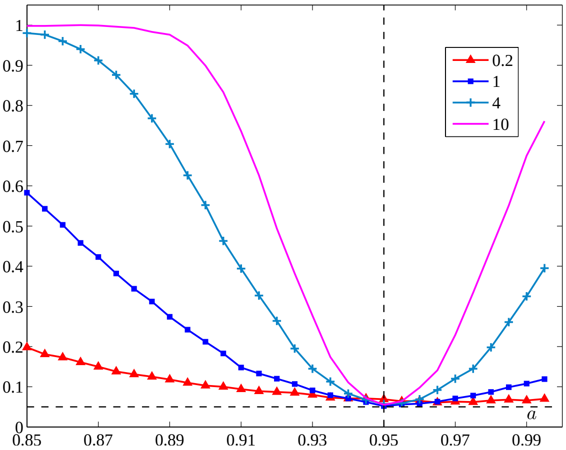


Figure 18(a): Power plot of M statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.96$)

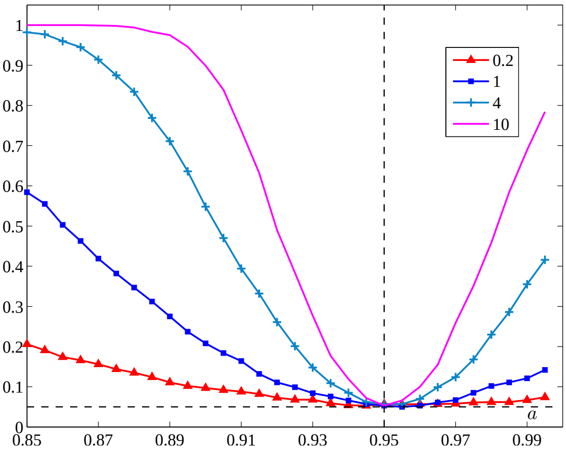


Figure 18(b): Power plot of M statistic
($\lambda = 0.2$, $\bar{\alpha} = 0.96$)

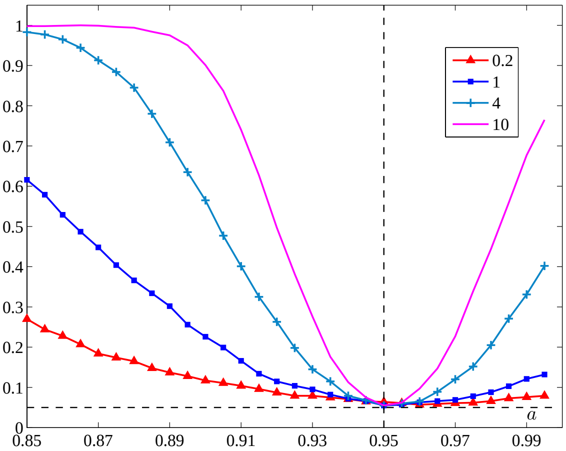


Figure 18(c): Power plot of M statistic
($\lambda = 1$, $\bar{\alpha} = 0.96$)

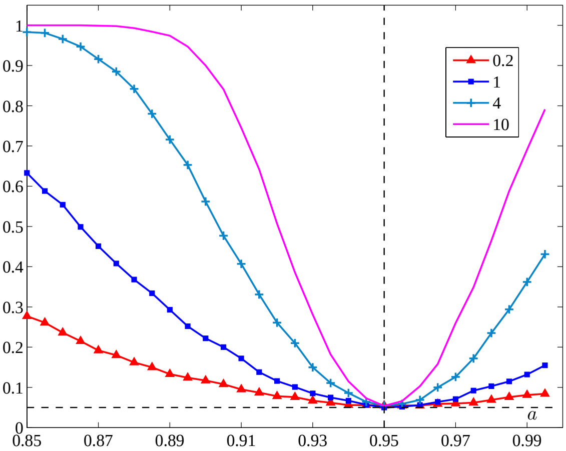


Figure 18(d): Power plot of M statistic
($\lambda = 1$, $\bar{\alpha} = 0.96$)

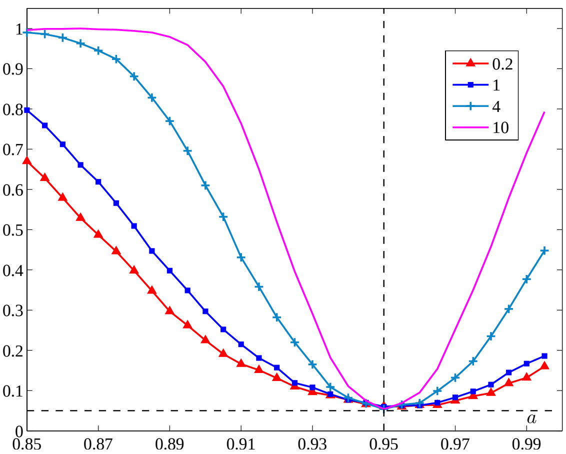


Figure 18(e): Power plot of M statistic
($\lambda = 5$, $\bar{\alpha} = 0.96$)

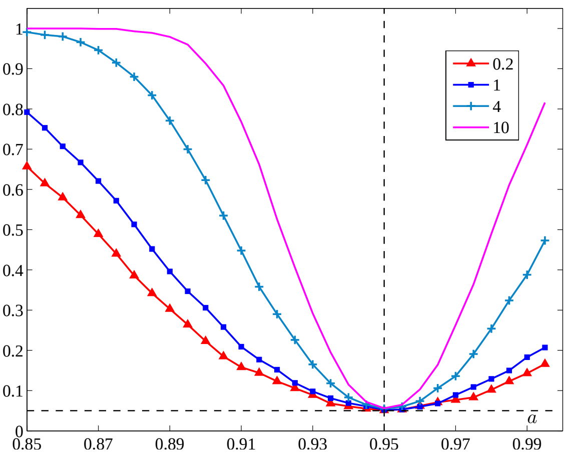


Figure 18(f): Power plot of M statistic
($\lambda = 5$, $\bar{\alpha} = 0.96$)