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

Kazuhiko Hayakawa

Department of Economics, Hiroshima University

July 4, 2010

1 Purpose of the paper

- In the literature, it is well known that the first-difference (FD-) GMM estimator in a dynamic panel data model suffers from the weak instruments problem when persistency is strong(Blundell and Bond, 1998).

- However, recently, in a panel AR(1) model, Hayakawa (2009) provides a simulation evidence that the performance of the FD-GMM estimator crucially depends on initial conditions.

- In some cases FD-GMM performs *poorly*, but in other cases it performs *very well* even if persistency is strong.

- The purpose of this paper is to theoretically demonstrate this result.

Outline of the Presentation

1. Purpose of the paper
2. Setup
3. Simulation evidence
4. Theoretical results
5. Conclusion
6. Future topic

2 Setup

2.1 Model

- We consider the following panel AR(1) model:

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

where $|\alpha| < 1$.

- N is large and T is small.

2.2 Assumption

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

Assumption 2 $\eta_i \sim iid(0, \sigma_\eta^2)$ *over* $i = 1, \cdots, N$.

Assumption 3 *For initial conditions, we assume*

$$y_{i0} = \delta\mu_i + \varepsilon_{i0} \tag{1}$$

where $\mu_i = \eta_i/(1-\alpha)$ *and* $\delta \neq 0$. *For* ε_{i0}, *we assume that*

$$\varepsilon_{i0} \sim iid\left(0, \frac{\lambda\sigma_v^2}{1-\alpha^2}\right) \tag{2}$$

over $i = 1, \cdots, N$ *where with* $\lambda > 0$.

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

2.3 Mean and covariance

- With initial conditions (1), we have the following expression:

$$y_{it} = \underbrace{\left[1-(1-\delta)\alpha^t\right]}_{h_t}\mu_i + \sum_{j=0}^{t-1}\alpha^j v_{i,t-j} + \alpha^t \varepsilon_{i0} \qquad (3)$$

- The mean and autocovariance of y_{it} given η_i are

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

- Both δ and λ determine the characteristics of data.

$\boxed{\delta = 1 \text{ and } \lambda = 1}$ $\Rightarrow$ y_{it} is **covariance stationary**.

$$
\begin{aligned}
E(y_{it}|\eta_i) &= \mu_i \\
cov(y_{is}, y_{it}|\eta_i) &= \mu_i^2 + \left(\frac{\sigma_v^2 \alpha^{t-s}}{1-\alpha^2}\right), \quad (s \leq t)
\end{aligned}
$$

$\boxed{\delta = 1 \text{ and } \lambda \neq 1}$ $\Rightarrow$ y_{it} is **mean-stationary**.

$$
\begin{aligned}
E(y_{it}|\eta_i) &= \mu_i \\
cov(y_{is}, y_{it}|\eta_i) &= \mu_i^2 + \sigma_v^2 \alpha^{t-s} \left(\frac{1-(1-\lambda)\alpha^{2s}}{1-\alpha^2}\right), \quad (s \leq t)
\end{aligned}
$$

$\boxed{\delta \neq 1 \text{ and } \lambda = 1 \text{ or } \neq 1}$ $\Rightarrow$ y_{it} is **mean-nonstationary**

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

- Most of the previous studies assume $\delta = \lambda = 1$(covariance stationarity) or $\delta = 1$(mean-stationarity).

- Blundell and Bond (1998) examined the behavior of several GMM estimators by simulation when $\delta \neq 1$(mean-nonstationarity). But they set $\alpha = 0.5$.

- Arellano (2003a,b) provide empirical evidence and examples of mean-nonstationarity.

2.4 The first difference GMM Estimator

- Consider the first-difference model:

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it} \quad (4)$$

where $\mathbf{z}_{it} = (y_{i0}, y_{i1}, ..., y_{i,t-2})'$ is used as instruments.

- The FD-GMM estimator is defined as

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

where $\Delta\mathbf{y}_i = (\Delta y_{i2}, \Delta y_{i3}, \cdots \Delta y_{iT})'$, $\Delta\mathbf{y}_{i,-1} = (\Delta y_{i1}, \Delta y_{i2}, \cdots \Delta y_{i,T-1})'$,

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

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

3 Simulation evidence

3.1 Simulation design

$$
\begin{aligned}
y_{it} &= 0.95\ y_{i,t-1} + \eta_i + v_{it}, \qquad\qquad (i = 1, ..., N;\ t = 1, ..., T) \\
&\quad \eta_i \sim iid\mathcal{N}(0, \sigma_\eta^2), \qquad v_{it} \sim iid\mathcal{N}(0, \sigma_v^2), \\
&\quad \sigma_\eta^2 = 1, 4, \qquad\qquad \sigma_v^2 = 1, \\
y_{i0} &= \frac{\eta_i}{1-\bar{\alpha}} + \varepsilon_{i0}, \quad \text{where } \delta \text{ and } \bar{\alpha} \text{ is related as } \delta = \frac{(1-\alpha)}{(1-\bar{\alpha})} \\
&\quad \varepsilon_{i0} \sim iid\mathcal{N}\left(0, \frac{1}{1-\alpha^2}\right) \text{ i.e. } \lambda = 1
\end{aligned}
$$

- $\bar{\alpha} = 0.93, 0.94, 0.95, 0.96, 0.97.$
- $T = 8$ and $N = 300$.
- Number of replication for each design is 1000.

3.2 Simulation results

Table 1: Median of the GMM estimator ($\alpha = 0.95$, $\lambda = 1$)

data	$\bar{\alpha}$	δ	$\sigma_\eta^2/\sigma_v^2 = 1$	$\sigma_\eta^2/\sigma_v^2 = 4$
mean-nonstationary	0.93	0.714	0.925	0.945
	0.94	0.833	0.877	0.936
covariance stationary	0.95	1.000	0.611	0.577
mean-nonstationary	0.96	1.250	0.932	0.945
	0.97	1.667	0.948	0.949

- When $\delta = 1$,
  - ◇ the GMM performs poorly.
  - ◇ $\sigma_\eta^2/\sigma_v^2 \uparrow \Rightarrow$ bias $\uparrow$.
- As $\delta(\bar{\alpha})$ moves away from $1(0.95)$,
  - ◇ the GMM tends to perform well.
  - ◇ $\sigma_\eta^2/\sigma_v^2 \uparrow \Rightarrow$ bias $\downarrow$.

Purpose of this paper

Theoretically demonstrate the following behaviors:

- Poor performance when δ is near 1, i.e, data are near mean-stationary.
- Good performance when δ is away from 1, i.e., data are mean-nonstationary.

4 Theoretical results

4.1 Possible approaches

(Conventional) asymptotic theory

⇒ Not useful to see the finite sample behavior since the consistency of the FD-GMM estimator does not depend on initial conditions.

Higher order theory

⇒ May be useful. But, it is generally complicated and more importantly, Hahn, Hausman and Kuersteiner (2007)(HHK) show that higher order theory does not appoximate the finite sample behavior for persistent panel data.

Local to unity approach

⇒ Hahn, Hausman and Kuersteiner (2007) (hereafter HHK) employ a local to unity system such that α approaches unity as N gets larger with T being fixed to approximate the finite sample behavior for persistent panel data.

⇒ Using a local to unity approach, HHK show that the FD-GMM estimator is inconsistent when data are covariance stationary.

⇒ This paper employs a local to unity approach.

- The difference between HHK and this paper is the assumption on data characteristics as follows:

$$\text{HHK} \quad \Rightarrow \quad \text{covariance stationary data}$$

$$\text{This paper} \quad \Rightarrow \quad \begin{cases} \text{covariance stationary data} \\ \text{mean-stationary data} \\ \text{mean-nonstationary data} \end{cases}$$

4.2 Local to unity system for α

- Assume the following parameter sequence.

$$\alpha_N = 1 - \frac{c}{N}, \qquad (c > 0) \tag{6}$$

- We use a local to unity system to approximate the distribution near unit root, which is the same spirit as HHK.

- The fact that $\alpha_N \to 1$ with $N \to \infty$ does not reflect any realistic data. Employing this local to unity system is an analytical device to obtain a better finite sample properties than that obtained under conventional asymptotics where parameter is fixed.

4.3 Strength of instruments

- HHK use a local to unity system $\alpha_N = 1 - \frac{c}{N}$ to discuss the weak instruments problem for covariance stationary data.

- Reconsider the strength of instruments when data may not be covariance stationary.

- Consider the F statistic of the first-stage regression in 2SLS regression form, which is a measure of instruments.

4.3.1 F statistic in the first-stage regression

- Consider the case $T = 2$. The 2SLS regression form is given by

$$\Delta y_{i2} = \alpha \Delta y_{i1} + \Delta v_{i2} \tag{7}$$

$$\Delta y_{i1} = \pi y_{i0} + u_i \tag{8}$$

- The first-stage F statistic to test $H_0 : \pi = 0$ is given by

$$F = (1-\delta)^2 O_p(N) + (1-\delta) O_p(1) + O_p\left(\frac{1}{N}\right) \tag{9}$$

δ	data	F statisitc	instruments
$\delta = 1$	mean-stationary	$O_p(1/N)$	weak
$\delta \neq 1$	mean-nonstationary	$O_p(N)$	strong

4.3.2 Intuitive discussion

- Unremoved individual effects in Δy_{it} makes an additional correlation.
- Correlation between endogenous variable and instruments can be measured by

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

where

$$\Delta y_{it} = (1-\delta)\alpha^{t-1}\eta_i + v_{it} + (\alpha-1)\sum_{j=0}^{t-2}\alpha^j v_{i,t-j-2} - (1-\alpha)\alpha^{t-1}\varepsilon_{i0}$$

- Whether $\delta = 1$ or not affects the strength of instruments.
  - $\Rightarrow$ When $\delta = 1$ (y_{it} is mean-stationary), the first term vanishes.
  - $\Rightarrow$ When $\delta \neq 1$ (y_{it} is mean-nonstationary), $1-\delta$ and σ_η^2/σ_v^2 affect the strength of instruments.

4.4 Asymptotic distributions I

Theorem 1 *Let Assumptions 1 to 4 hold and $N \to \infty$ with T fixed. Then, under a local to unity system (6), we have*

$$\textbf{When } \delta = 1: \quad \widehat{\alpha} - \alpha \xrightarrow{d} \frac{\boldsymbol{\xi}' \left(\mathbf{JHJ}'\right)^{+} \boldsymbol{\zeta}}{\boldsymbol{\xi}' \left(\mathbf{JHJ}'\right)^{+} \boldsymbol{\xi}},$$

$$\textbf{When } \delta \neq 1: \quad \sqrt{N}\left(\widehat{\alpha} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\sigma_v^2}{\sigma_\eta^2 \delta^2 \boldsymbol{\iota}_m' (\mathbf{JHJ}')^{+} \boldsymbol{\iota}_m}\right)$$

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

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

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

and $\mathbf{J}$ *are defined as*

$$\mathbf{J} = diag(\iota_1, \boldsymbol{\iota}_2', \cdots \boldsymbol{\iota}_{T-1}')$$

4.5 Local to unity approach for δ

- This result may not be useful to explain the behavior of the GMM estimator in finite sample.
- Consider a GMM estimator with $\delta = 1.00001$. Asymptotic theory indicates that this GMM estimator is consistent since $\delta \neq 1$ although very close to 1.
- It is expected that the GMM estimator with $\delta = 1.00001$ behaves similarly to the GMM estimator with $\delta = 1$.
- The drawback of Theorem 1 is that when δ is fixed, we cannot distinguish $\delta = 1.00001$ and $\delta = 1$.

- To take the closeness of δ to 1 into account, we introduce the following sequence:

$$\delta_N = 1 - \frac{d}{N^q}, \qquad d \neq 0, \quad 0 < q \leq 1 \tag{10}$$

- In this sequence, q controls the closeness of data to mean-stationarity.

  $\Rightarrow$ When q is close to 1, δ_N is close to 1.

  $\Rightarrow$ When q is close to 0, δ_N tends to be deviated from 1.

4.6 F statistic under double local to unity systems

- With the sequence δ_N, the first-stage F statistic given in (9) becomes

$$\begin{aligned} F &= (1-\delta_N)^2 O_p(N) + (1-\delta_N) O_p(1) + O_p(1/N) \\ &= O_p(N^{1-2q}) + O_p(N^{-q}) + O_p(N^{-1}). \end{aligned}$$

- q determines the order of magnitudes.

q	data	F statisitc	instruments
$1/2 < q \leq 1$	mean-stationary	$o_p(1)$	weak
$q = 1/2$	$\updownarrow$	$O_p(1)$	weak
$0 < q < 1/2$	mean-nonstationary	$O_p(N^{1-2q})$	strong

4.7 Asymptotic distributions II

Theorem 2 *Let Assumptions hold and* $N \to \infty$ *with* T *fixed. Then, under local to unity systems (6) and (10), we have*

When $\frac{1}{2} < q \leq 1$**:** $\quad \widehat{\alpha} - \alpha \xrightarrow{d} \dfrac{\boldsymbol{\xi}' \left(\mathbf{J}'\mathbf{H}\mathbf{J}\right)^{+} \boldsymbol{\zeta}}{\boldsymbol{\xi}' \left(\mathbf{J}'\mathbf{H}\mathbf{J}\right)^{+} \boldsymbol{\xi}}$

When $q = \frac{1}{2}$**:** $\quad \widehat{\alpha} - \alpha \xrightarrow{d} \dfrac{(\boldsymbol{\xi} + \boldsymbol{\mu}_2)' \left(\mathbf{J}'\mathbf{H}\mathbf{J}\right)^{+} \boldsymbol{\zeta}}{(\boldsymbol{\xi} + \boldsymbol{\mu}_2)' \left(\mathbf{J}'\mathbf{H}\mathbf{J}\right)^{+} (\boldsymbol{\xi} + \boldsymbol{\mu}_2)},$

When $0 < q < \frac{1}{2}$**:** $\quad N^{1/2-q} \left(\widehat{\alpha} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \dfrac{2\sigma_v^2}{d^2 \sigma_\eta^2 \boldsymbol{\iota}_m' (\mathbf{J}\mathbf{H}\mathbf{J}')^{+} \boldsymbol{\iota}_m}\right)$

where $\boldsymbol{\mu}_2 = \frac{d\sigma_\eta^2}{c} \boldsymbol{\iota}_m$.

- When data are near mean-statioanry in the sense that $1/2 \leq q \leq 1$, the FD-GMM is inconsistent.

- When there are strong mean-nonstationarity in data, the FD-GMM is consistent.

- These asymptotic distributions explain the finite sample behavior well.

5 Conclusion

- This paper derived the asymptotic distribusion of the FD-GMM estimator under local to unity systems when data may not be covariance stationary.

- Confirmed that derived asymptotic distributions explained the finite sample behavior of the FD-GMM well.

6 Future topics

- Although not reported here, we confirmed that the long difference IV estimator by HHK is less affected by the data characteristics and the bias-corrected WG estimator by Bun and Carree (2005) performs well irrespective of data characterisitics.

- Extend the model to have regressors and investigate the behavior of FD-GMM, long difference IV and bias-corrected WG estimators by Monte Carlo simulation.

7 Additional simulation results

$$
\begin{aligned}
y_{it} &= 0.95\, y_{i,t-1} + \eta_i + v_{it}, \qquad\qquad (i = 1, ..., N;\ t = 1, ..., T) \\
& \quad \eta_i \sim iid\mathcal{N}(0, \sigma_\eta^2), \qquad v_{it} \sim iid\mathcal{N}(0, 1), \\
y_{i0} &= \frac{\eta_i}{1 - \bar{\alpha}} + \varepsilon_{i0}, \quad \text{where } \delta \text{ and } \bar{\alpha} \text{ is related as } \delta = \frac{(1 - \alpha)}{(1 - \bar{\alpha})} \\
& \quad \varepsilon_{i0} \sim iid\mathcal{N}\left(0, \frac{1}{1 - \alpha^2}\right) \text{ i.e. } \lambda = 1
\end{aligned}
$$

- $\bar{\alpha} = 0.9000, 0.9025, \cdots, 0.9500, \cdots, 0.9975.$
- $T = 8$ and $N = 300$.
- $\sigma_\eta^2 = 0.2, 1, 4, 10$

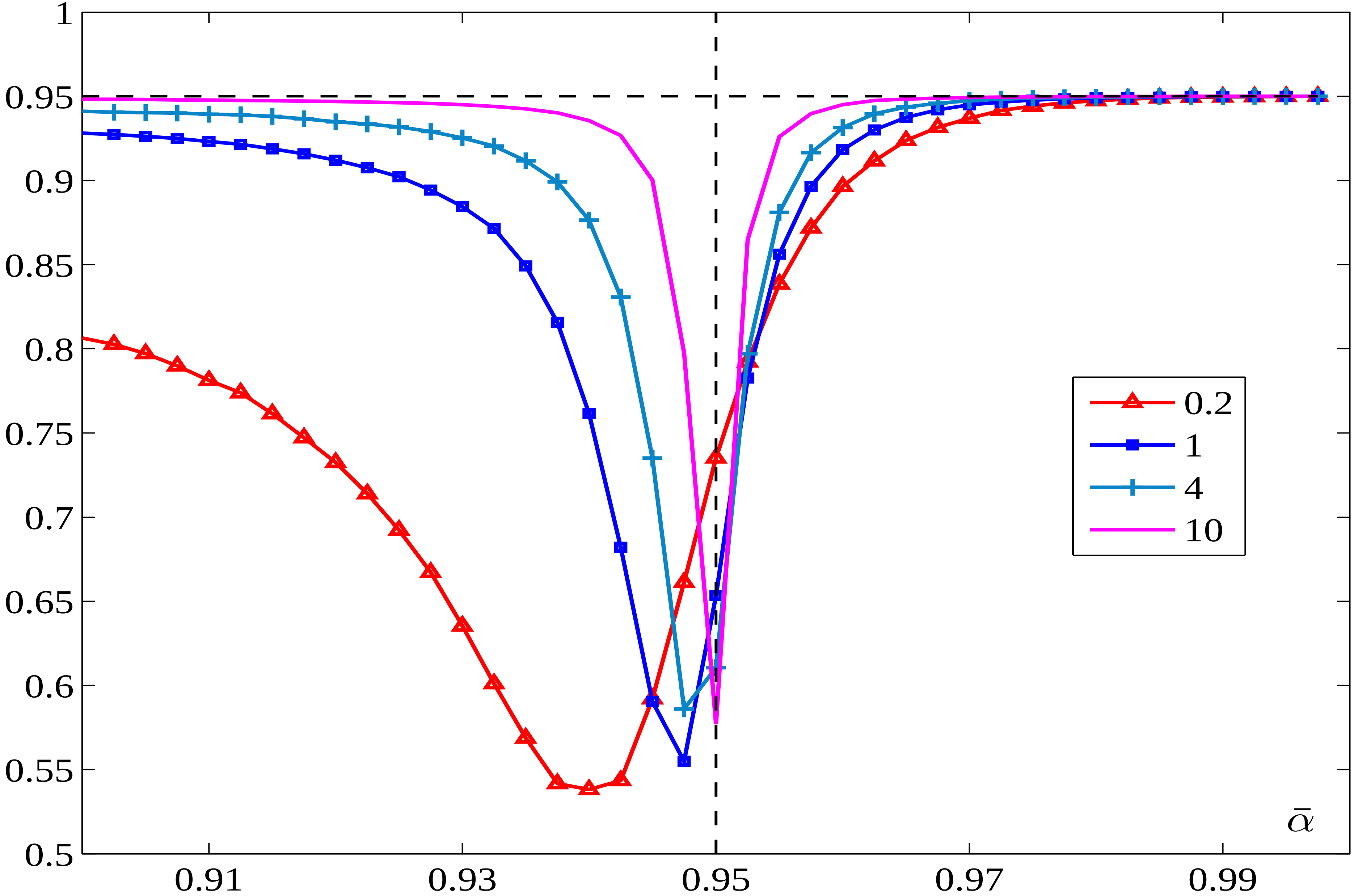

Figure 1: Median of the GMM estimator ($\lambda = 1$)

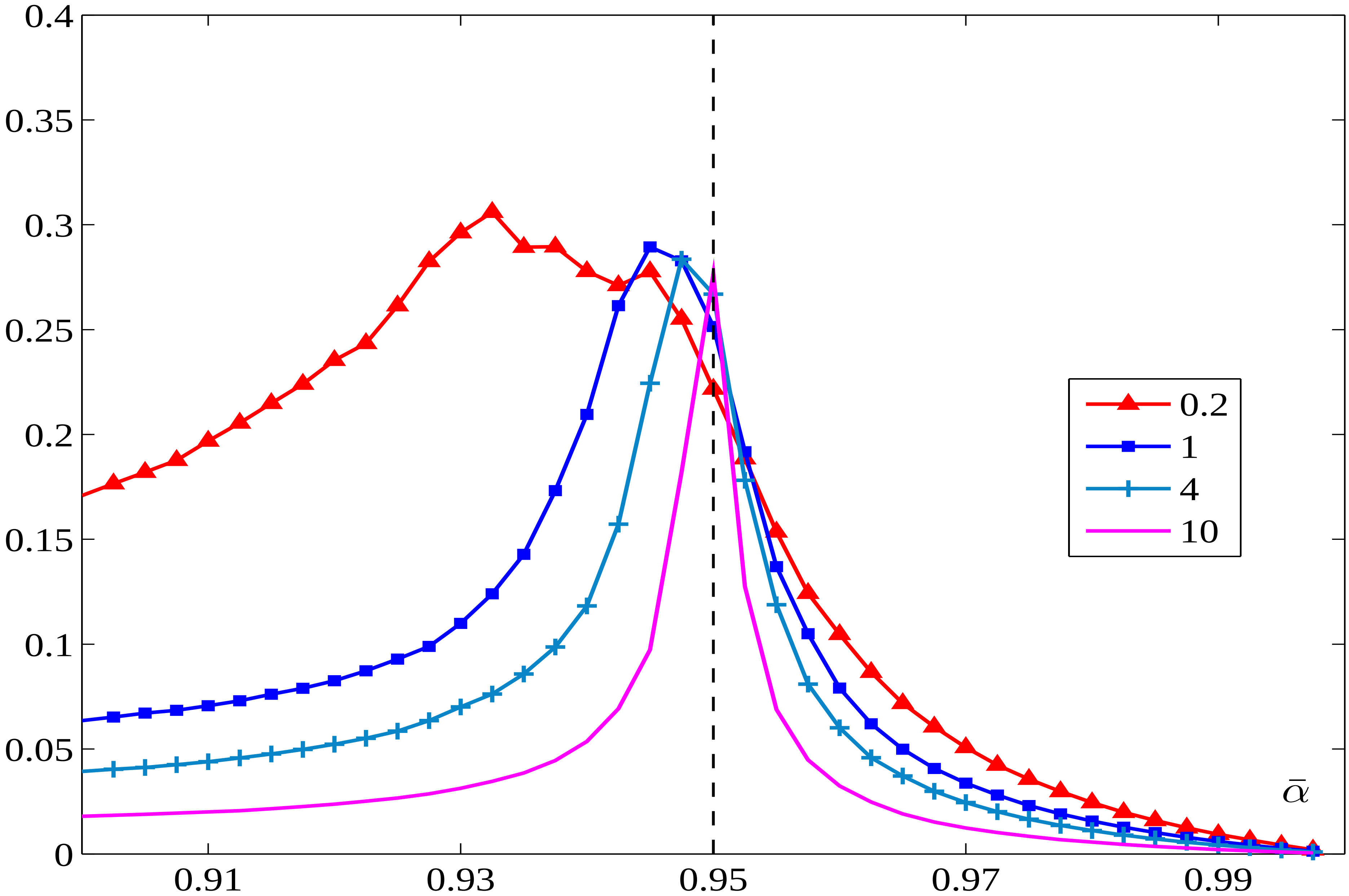

Figure 2: Interquartile range of the GMM estimator ($\lambda = 1$)

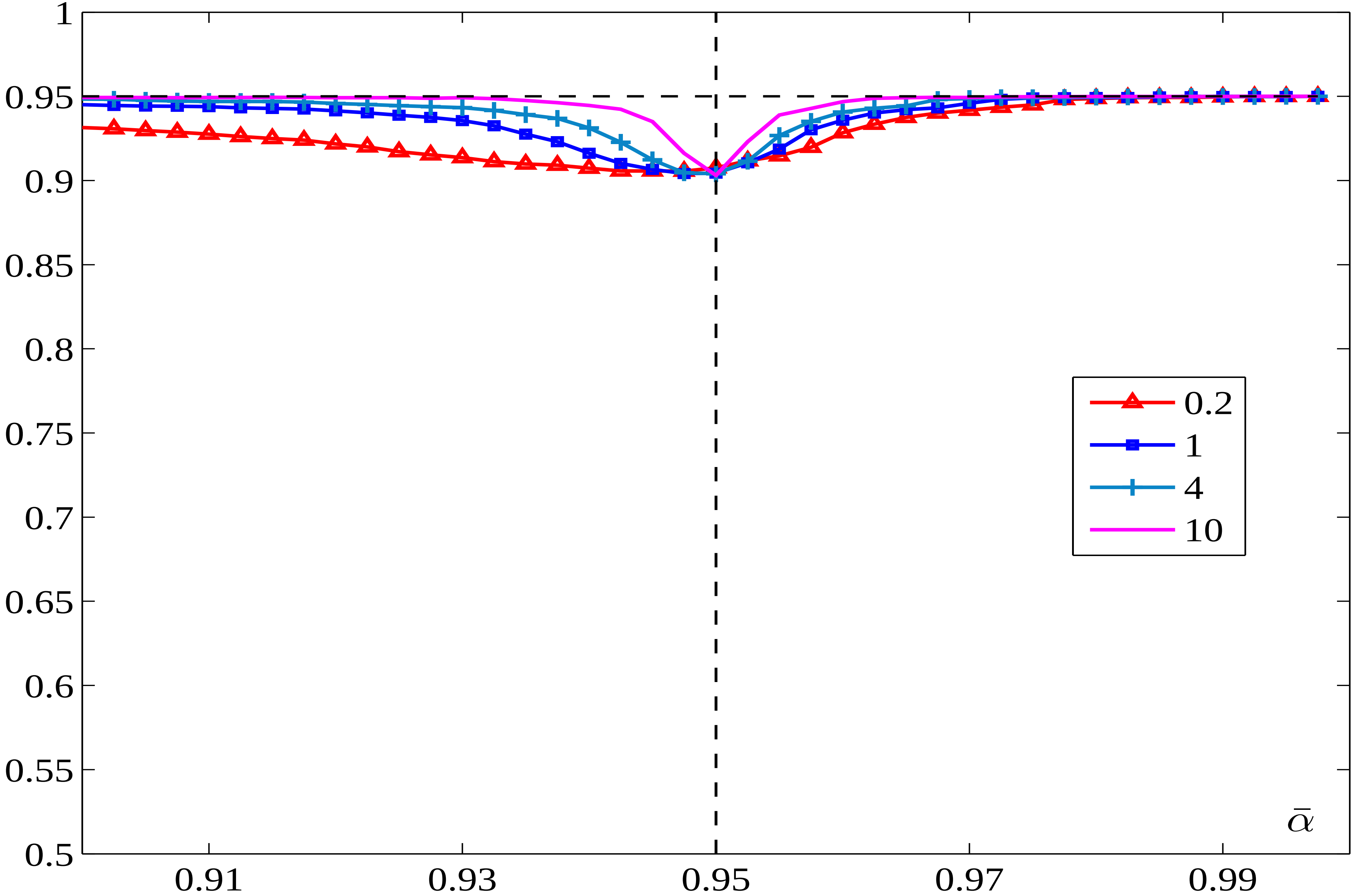

Figure 3: Median of the three times iterated LDIV estimator

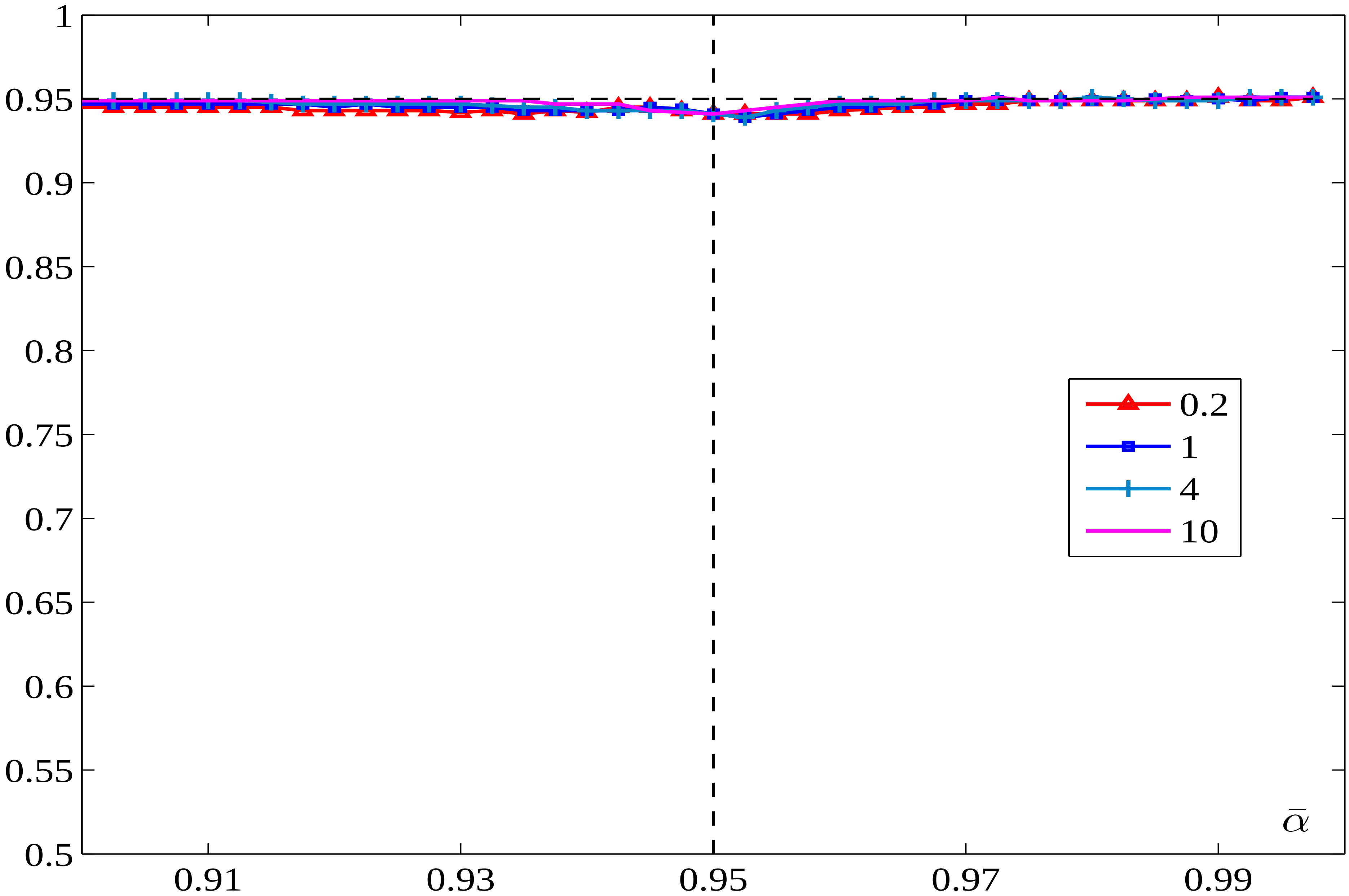

Figure 4: Median of the BCWG estimator

References

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

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

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

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