TRUE LIMIT DISTRIBUTIONS OF THE ANDERSON-HSIAO IV ESTIMATORS IN PANEL AUTOREGRESSION

By

Peter C. B. Phillips and Chirok Han

December 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1963

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

The True Limit Distributions of the Anderson-Hsiao IV Estimators in Panel Autoregression*

Peter C. B. Phillips

Yale University, University of Auckland,

Singapore Management University & University of Southampton

Chirok Han

Korea University

September 2014

Abstract

This note derives the correct limit distributions of the Anderson Hsiao (1981) levels and differences instrumental variable estimators, provides comparisons showing that the levels IV estimator has uniformly smaller variance asymptotically as the cross section (n) and time series (T) sample sizes tend to infinity, and compares these results with those of the first difference least squares (FDLS) estimator.

Keywords: Dynamic panel, IV estimation, Levels and difference instruments.

JEL classification: C230, C360

1 Introduction

In pioneering work on dynamic panel models, Anderson and Hsiao (1981, AH hereafter) developed two consistent instrumental variable (IV) estimators for the common slope coefficient in first

*Phillips acknowledges support from the NSF under Grant No. SES 12-58258. Research by Han was supported by Korea University (K1421311).

order panel autoregression. These estimators used lagged levels and lagged differences as instruments and they form the core of much later work on GMM approaches to inference in dynamic panels. This note corrects the AH limit theory and provides an interesting asymptotic equivalence between their levels IV estimator and the first difference least squares estimator of Phillips and Han (2008) and Han and Phillips (2010). The levels IV estimator is shown to have asymptotically uniformly smaller variance than the difference IV estimator when $(n, T) \to \infty$. For fixed T, the levels estimator is also more efficient except when T is very small.

2 Asymptotic distributions of IV estimators

For the simple panel dynamic model $y_{it} = \alpha_i + \beta y_{it-1} + u_{it}$, after eliminating the nuisance fixed effects by first-differencing the equation, AH (1981, Section 8) propose using lagged variables in levels or differences as potential instrumental variables. The resulting levels and difference IV estimators are

$$\hat{\beta}_l = \frac{\sum_{i=1}^n \sum_{t=2}^T y_{it-2} \Delta y_{it}}{\sum_{i=1}^n \sum_{t=2}^T y_{it-2} \Delta y_{it-1}} \text{ and } \hat{\beta}_d = \frac{\sum_{i=1}^n \sum_{t=3}^T \Delta y_{it-2} \Delta y_{it}}{\sum_{i=1}^n \sum_{t=3}^T \Delta y_{it-2} \Delta y_{it-1}},$$

where $\Delta y_{it} = y_{it} - y_{it-1}$. We provide the correct asymptotics for these two estimators under stationarity.

Levels IV

Let $y_{it} = \alpha_i + \beta y_{it-1} + u_{it}$ $(i = 1, ..., n; t = 1, ..., T)$ with $|\beta| < 1$, $u_{it} \sim_{iid} (0, \sigma^2)$, $y_{i0} = \frac{\alpha_i}{1-\beta} + \sum_{j=0}^\infty \rho^j u_{i,-j}$, and $\alpha_i \sim_{iid} (0, \sigma_\alpha^2)$ independent of u_{it}. The levels IV estimator $\hat{\beta}_l$ satisifies

$$\sqrt{n}(\hat{\beta}_l - \beta) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{t=2}^T y_{it-2} \Delta u_{it}}{\frac{1}{n} \sum_{i=1}^n \sum_{t=2}^T y_{it-2} \Delta y_{it-1}} =: \frac{N_{nT}^l}{D_{nT}^l}. \tag{1}$$

Since $y_{it} = \frac{\alpha_i}{1-\beta} + \sum_{j=0}^\infty \rho^j u_{t-j}$ is stationary,

$$\mathbb{E}\left(y_{it}^2\right) = \frac{\sigma_\alpha^2}{(1-\beta)^2} + \frac{\sigma^2}{1-\beta^2}, \ \ \mathbb{E}\left(y_{it} y_{it-j}\right) = \frac{\sigma_\alpha^2}{(1-\beta)^2} + \frac{\beta^{|j|}\sigma^2}{1-\beta^2}. \tag{2}$$

Hence, as $n \to \infty$,

$$\begin{aligned}
D_{nT}^{l} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=3}^{T} y_{it-2}\Delta y_{it-1} \\
&\to {}_{a.s.}\mathbb{E}\left(\sum_{t=2}^{T} y_{it-2}\Delta y_{it-1}\right) = \sum_{t=2}^{T}\left[\mathbb{E}\left(y_{it-2}y_{it-1}\right) - \mathbb{E}\left(y_{it-2}^{2}\right)\right] \\
&= T_1\left[\mathbb{E}\left(y_{it-1}y_{it-2}\right) - \mathbb{E}\left(y_{it-2}^{2}\right)\right] = \frac{T_1\left(\beta-1\right)\sigma^2}{\left(1-\beta\right)^2} = -\frac{T_1\sigma^2}{1+\beta},
\end{aligned} \tag{3}$$

where $T_k = T - k$. By partial summation $\Delta(y_{it-2}u_{it}) = y_{it-2}\Delta u_{it} + (\Delta y_{it-2})u_{it-1}$ for $t \geq 3$ and we have

$$\begin{aligned}
\sum_{t=3}^{T} y_{it-2}\Delta u_{it} &= y_{iT-2}u_{iT} - y_{i0}u_{i2} - \sum_{t=3}^{T}(\Delta y_{it-2})u_{it-1}, \\
y_{i0}\Delta u_{i2} &= y_{i0}u_{i2} - y_{i0}u_{i1},
\end{aligned}$$

and adding gives

$$\sum_{t=2}^{T} y_{it-2}\Delta u_{it} = y_{iT-2}u_{iT} - y_{i0}u_{i1} - \sum_{t=3}^{T}(\Delta y_{it-2})u_{it-1}. \tag{4}$$

Since $(\Delta y_{it-2})u_{it-1}$ is a martingale difference sequence and

$$\mathbb{E}(\Delta y_{it})^2 = 2\mathbb{E}(y_{it}^2) - 2\mathbb{E}(y_{it}y_{it-1}) = \frac{2\sigma^2\left(1-\beta\right)}{\left(1-\beta^2\right)} = \frac{2\sigma^2}{1+\beta},$$

we have

$$\begin{aligned}
\mathbb{E}\left(\sum_{t=2}^{T} y_{it-2}\Delta u_{it}\right)^2 &= \mathbb{E}\left[y_{iT-2}u_{iT} - y_{i0}u_{i1} - \sum_{t=3}^{T}(\Delta y_{it-2})u_{it-1}\right]^2 \\
&= 2\sigma^2\left[\frac{\sigma_\alpha^2}{\left(1-\beta\right)^2} + \frac{\sigma^2}{1-\beta^2}\right] + \sigma^2\sum_{t=3}^{T}\mathbb{E}\left(\Delta y_{it-2}\right)^2 \\
&= 2\sigma^2\left[\frac{\sigma_\alpha^2}{\left(1-\beta\right)^2} + \frac{\sigma^2}{1-\beta^2}\right] + 2\sigma^2 T_2\frac{\sigma^2}{1+\beta} \\
&= 2\sigma^2\left[\frac{\sigma_\alpha^2}{\left(1-\beta\right)^2} + \frac{\sigma^2}{1-\beta^2} - \frac{\sigma^2}{1+\beta}\right] + 2\sigma^2 T_1\frac{\sigma^2}{1+\beta} \\
&= 2\sigma^2\left[\frac{\sigma_\alpha^2}{\left(1-\beta\right)^2} + \frac{\sigma^2\beta}{1-\beta^2}\right] + 2\sigma^2 T_1\frac{\sigma^2}{1+\beta}.
\end{aligned}$$

Since $\mathbb{E}(\sum_{t=2}^{T} y_{it-2}\Delta u_{it}) = 0$, the numerator of (1) satisfies the CLT

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=2}^{T} y_{it-2}\Delta u_{it} \Rightarrow N\left(0, 2\sigma^2 T_1\frac{\sigma^2}{1+\beta} + 2\sigma^2\left[\frac{\sigma_\alpha^2}{\left(1-\beta\right)^2} + \frac{\sigma^2\beta}{1-\beta^2}\right]\right)$$

and the denominator $D_{nT}^{l} \to_{a.s.} -T_1 \frac{\sigma^2}{1+\beta}$, we have

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_l - \beta) \quad & \underset{n\to\infty}{\Rightarrow} N\left(0, \frac{2\sigma^2 T_1 \frac{\sigma^2}{1+\beta} + 2\sigma^2 \left[\frac{\sigma_\alpha^2}{(1-\beta)^2} + \frac{\sigma^2\beta}{1-\beta^2}\right]}{\left(-T_1 \frac{\sigma^2}{1+\beta}\right)^2}\right) \\
& = N\left(0, \frac{2(1+\beta)}{T_1} + \frac{2(1+\beta)^2}{T_1^2(1-\beta)}\left[\frac{\sigma_\alpha^2/\sigma^2}{1-\beta} + \frac{\beta}{1+\beta}\right]\right). \qquad (5)
\end{aligned}$$

The asymptotic variance of (5) increases with σ_α^2/σ^2, which is natural because the α_i are uninformative for the identification of β. As $T \to \infty$

$$\sqrt{nT}(\hat{\beta}_l - \beta) \underset{(T,n)\to\infty}{\Rightarrow} N(0, 2(1+\beta)), \qquad (6)$$

which is the same limit distribution in the stationary case as the GMM estimator in Han and Phillips (2010, HP). Importantly, in (6) there is no dependence in the limit variance on σ_α^2. Also, note that there is a discontinuity in the limit theory as $\beta \to 1$ because in that case $\hat{\beta}_l$ is only $\sqrt{T}$ consistent and has a limit Cauchy distribution (Phillips, 2014).

The expression in (5) differs from AH's result (8.4). The error in AH seems to arise because $y_{it-2}\Delta u_{it}$ is mistaken as a martingale difference sequence and the asymptotic variance actually involves cross product terms and a (finite T) long run variance. The above demonstration simply avoids this calculation by using partial summation to put the sum $\sum_{t=2}^{T} y_{it-2}\Delta u_{it}$ into a more convenient form.

The equivalence of the AH and HP estimators for large T is unexpected, because the AH estimator is derived under weaker orthogonality conditions $\mathbb{E}(y_{it-2}\Delta u_{it}) = 0$ whereas the HP estimator requires covariance stationarity also. To explore the equivalance, algebra shows that the HP estimator $\hat{\beta}_{fd}$ can be written in the form that relates to $\hat{\beta}_l$ with end corrections, viz.,

$$\hat{\beta}_{fd} = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T} y_{it-2}\Delta y_{it} + w_1 + w_2}{\sum_{i=1}^{n}\sum_{t=2}^{T} y_{it-2}\Delta y_{it-1} + w_1}, \qquad (7)$$

where $w_1 = \frac{1}{2}\sum_{i=1}^{n}(y_{i0}^2 - y_{iT-1}^2)$ and $w_2 = \sum_{i=1}^{n}(y_{i0}\Delta y_{i1} - y_{iT-1}\Delta y_{iT})$. The stationarity requirement for $\hat{\beta}_{fd}$ affects w_1 and w_2. When $E(y_{it}^2)$ is stable, the terms w_1 and w_2 terms are dominated by the leading terms and are therefore negligible for large T, leading to the asymptotic equivalence of $\hat{\beta}_{fd}$ and $\hat{\beta}_l$.

Difference IV

The difference IV estimator $\hat{\beta}_d$ satisifies

$$\sqrt{n}(\hat{\beta}_d - \beta) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{t=3}^T \Delta y_{it-2} \Delta u_{it}}{\frac{1}{n} \sum_{i=1}^n \sum_{t=3}^T \Delta y_{it-2} \Delta y_{it-1}} =: \frac{N_{nT}^d}{D_{nT}^d}.$$

In the stationary case, the denominator satisfies

$$\begin{aligned} D_{nT}^d &= \frac{1}{n} \sum_{i=1}^n \sum_{t=3}^T \Delta y_{it-2} \Delta y_{it-1} \to_{a.s.} \mathbb{E}\left(\sum_{t=3}^T \Delta y_{it-2} \Delta y_{it-1} \right) \\ &= \sum_{t=3}^T \left[\mathbb{E}\left(y_{it-1} y_{it-2}\right) - \mathbb{E}\left(y_{it-2}^2\right) - \mathbb{E}\left(y_{it-1} y_{it-3}\right) + \mathbb{E}\left(y_{it-2} y_{it-3}\right) \right] \\ &= T_2 \left[2\left(\frac{\beta\sigma^2}{1-\beta^2} + \frac{\sigma_\alpha^2}{(1-\beta)^2} \right) - \left(\frac{\sigma_\alpha^2}{(1-\beta)^2} + \frac{\sigma^2}{1-\beta^2} \right) - \left(\frac{\beta^2\sigma^2}{1-\beta^2} + \frac{\sigma_\alpha^2}{(1-\beta)^2} \right) \right] \\ &= -T_2 \frac{\sigma^2 (1-\beta)^2}{1-\beta^2} = -T_2 \sigma^2 \left(\frac{1-\beta}{1+\beta} \right) = D_{nT}^l \left(\frac{T_2}{T_1} \right) (1-\beta). \end{aligned}$$

Similar to (4), we have

$$\sum_{t=3}^T \Delta y_{it-2} \Delta u_{it} = (\Delta y_{iT-2}) u_{iT} - (\Delta y_{i1}) u_{i2} - \sum_{t=4}^T (\Delta^2 y_{it-2}) u_{it},$$

where $\Delta^2 y_{it-2} = \Delta y_{it-2} - \Delta y_{it-3}$. Because $\mathbb{E}(\Delta y_{it})^2 = \frac{2\sigma^2}{1+\beta}$, $\mathbb{E}(\Delta^2 y_{it-2})^2 = \frac{2(3-\beta)\sigma^2}{1+\beta}$, $\mathbb{E}(\Delta y_{it-1} \Delta y_{it-2}) = -\frac{\sigma^2(1-\beta)}{1+\beta}$, and $(\Delta^2 y_{it-2}) u_{it}$ is a martingale difference, the numerator satisfies $N_{nT}^d \Rightarrow N(0, V_T)$ with

$$\begin{aligned} V_T &= \left[2\mathbb{E}(\Delta y_{it})^2 + T_3 \mathbb{E}(\Delta^2 y_{it-2})^2 \right] \sigma^2 = \left[\frac{4\sigma^2}{1+\beta} + T_3 \cdot \frac{2(3-\beta)\sigma^2}{1+\beta} \right] \sigma^2 \\ &= \frac{2\sigma^4}{1+\beta} [2 + T_3(3-\beta)] = \frac{2\sigma^4}{1+\beta} [T_2(3-\beta) - (1-\beta)]. \end{aligned} \tag{8}$$

Then

$$\begin{aligned} \sqrt{n}(\hat{\beta}_d - \beta) &= \frac{N_{nT}^d}{D_{nT}^d} \underset{n\to\infty}{\Rightarrow} N\left(0, \frac{2T_2\left(\frac{3-\beta}{1+\beta}\right) - 2\left(\frac{1-\beta}{1+\beta}\right)}{\left[-T_2 \left(\frac{1-\beta}{1+\beta}\right) \right]^2} \right) \\ &= N\left(0, \frac{2(1+\beta)(3-\beta)}{T_2 (1-\beta)^2} - \frac{2}{T_2^2} \left(\frac{1+\beta}{1-\beta} \right) \right), \end{aligned} \tag{9}$$

which differs from AH's result (8.3) and leads to the following sequential limit theory

$$\sqrt{nT}(\hat{\beta}_d - \beta) \underset{(T,n)\to\infty}{\Rightarrow} N\left(0, \frac{2(1+\beta)(3-\beta)}{(1-\beta)^2} \right). \tag{10}$$

Efficiency Comparison

Comparing (10) with (6), it is clear that the limit variance of $\hat{\beta}_l$ for large n and T is smaller than that of $\hat{\beta}_d$ for all $\beta \in (-1, 1)$ since $\frac{3-\beta}{(1-\beta)^2} > 0$. It is easy to show, as in Phillips (2014) using the methods of Phillips and Moon (1999), that the convergences in (10) with (6) are both sequential ($n \to \infty$ followed by $T \to \infty$) and joint ($n, T \to \infty$ without restriction on the rates or the path of divergence).

For fixed T, from (5) and (9), we evaluate $Avar(\hat{\beta}_l)/Avar(\hat{\beta}_d)$ for various β, T and σ_α^2/σ^2. Figure 1(a) exhibits this ratio for $\sigma_\alpha^2/\sigma^2 = 1$. The levels estimator is more efficient for all β and $T \geq 3$ in this case. Larger σ_α^2/σ^2 ratios are more favorable to $\hat{\beta}_d$ but still the levels estimator is more efficient unless T is very small. Numerical evaluations suggest that the levels estimator is better than the difference estimator for all $\beta \in (-1, 1)$ for all T if $\sigma_\alpha^2/\sigma^2 \leq 4$. For even larger σ_α^2/σ^2, Figure 1(b) considers $\sigma_\alpha^2/\sigma^2 = 8$. The difference AH estimator performs better than the levels estimator only for small T and large β. Numerical evaluations show that the levels estimator is more efficient than the difference estimator for all β for $T \geq 8$.

3 Simulations

Table 1 presents the simulated variances (times n) and the asymptotic variances in (5) and (9) for $\beta = 0.5$. For the levels IV estimator, denoted AH(L), and difference IV estimator, denoted AH(D), the simulated variances are close to the asymptotic except for very small T, for which a larger n would be required due to possible correlation between the numerator and the denominator. For all settings except $T = 5$ and $\sigma_\alpha^2/\sigma^2 = 8$, AH(L) is more efficient than AH(D), reflecting the asymptotic theory. (See Figure 1 for the asymptotic variances.) For $\sigma_\alpha^2/\sigma^2 = 1$, the variance of the HP estimator is uniformly (and considerably) smaller than that of AH(L) for small T; and for large T, HP and AH(L) perform similarly, just as the large-T asymptotic equivalence suggests. But when $\sigma_\alpha^2/\sigma^2 = 8$, the discrepancy is larger and the HP estimator is markedly superior to AH(L). From (5), the large n asymptotic variance of AH(L) involves the additional term

$$\frac{2\,(1+\beta)^2}{T_1^2(1-\beta)} \left[\frac{\sigma_\alpha^2/\sigma^2}{1-\beta} + \frac{\beta}{1+\beta} \right],$$

which has a substantial impact on variance when σ_α^2/σ^2 is large. In this event, much larger values of T are required for the variance of AH(L) to be close to that of HP.

References

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components, Journal of the American Statistical Association, 76, 598–606.

Han, C. and P. C. B. Phillips (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity, Econometric Theory, 26, 119–151.

Han, C. and P. C. B. Phillips (2013). First difference MLE and dynamic panel estimation, Journal of Econometrics, 175, 35–45.

Moon, H. R. and P. C. B. Phillips, (2004). GMM estimation of autoregressive roots near unity with panel data, Econometrica, 72, 467–522.

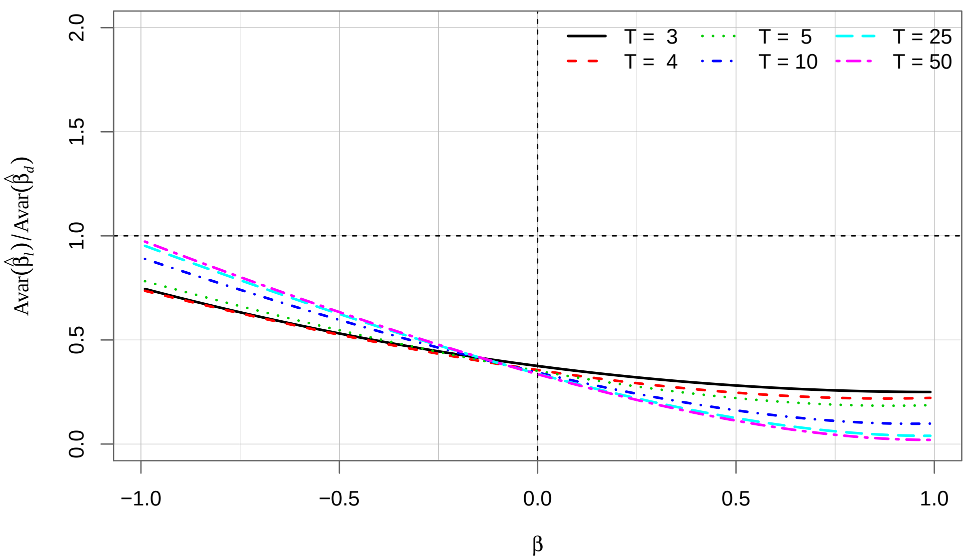
Phillips, P. C. B, (2014). Dynamic panel GMM with roots near unity, Working Paper, Yale University.

Phillips, P. C. B. and C. Han (2008). Gaussian inference in AR(1) time series with or without a unit root, Econometric Theory, 24, 631–650.

Phillips, P.C.B. and H.R. Moon (1999). Linear regression limit theory for nonstationary panel data, Econometrica, 67, 1057–1111.

Figure 1: Asymptotic variance ratio $Avar(\hat{\beta}_l)/Avar(\hat{\beta}_d)$

(a) $\sigma_\alpha^2/\sigma^2 = 1$



(b) $\sigma_\alpha^2/\sigma^2 = 8$

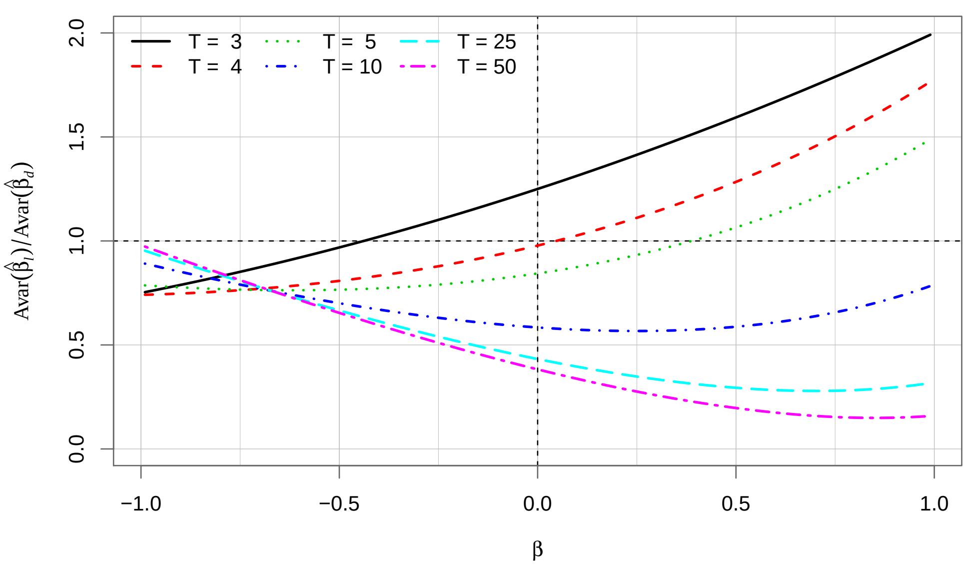


Table 1: Simulated $n\times$ variances ($n = 400$, $\beta = 0.5$, 10,000 replications)

$$y_{it} = \alpha_i + \beta y_{it-1} + u_{it},\ \alpha_i = \sigma_\alpha a_i,\ u_i = \sigma\varepsilon_{it},\ \sigma = 1,$$

$$y_{i,-100} = \alpha_i/(1-\beta) + u_{i0}/\sqrt{1-\beta^2},\ a_i, \varepsilon_{it} \sim_{iid} N(0,1)$$

(a) $\sigma_\alpha^2/\sigma^2 = 1$

T	HP	AH(L)		AH(D)	
5	0.7564	2.1299	(2.0625)	9.7448	(9.3333)
10	0.3308	0.5943	(0.5926)	3.7375	(3.6562)
20	0.1579	0.2203	(0.2161)	1.6798	(1.6481)
40	0.0771	0.0929	(0.0907)	0.7889	(0.7853)
80	0.0378	0.0415	(0.0413)	0.3833	(0.3836)
160	0.0190	0.0200	(0.0197)	0.1924	(0.1896)

(b) $\sigma_\alpha^2/\sigma^2 = 8$

T	HP	AH(L)		AH(D)	
5	0.7564	11.0850	(9.9375)	9.7448	(9.3333)
10	0.3308	2.2150	(2.1481)	3.7375	(3.6562)
20	0.1579	0.5809	(0.5651)	1.6798	(1.6481)
40	0.0771	0.1792	(0.1736)	0.7889	(0.7853)
80	0.0378	0.0621	(0.0615)	0.3833	(0.3836)
160	0.0190	0.0253	(0.0247)	0.1924	(0.1896)

Note: The numbers in parentheses are asymptotic variances.