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

BY

Peter C. B. Phillips and Chirok Han

COWLES FOUNDATION PAPER NO. 1455

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

2015

http://cowles.econ.yale.edu/
The true limit distributions of the Anderson–Hsiao IV estimators in panel autoregression

Peter C.B. Phillips\textsuperscript{a,b,c,d,}\textsuperscript{*}, Chirok Han\textsuperscript{e}

\textsuperscript{a}Yale University, United States
\textsuperscript{b}University of Auckland, New Zealand
\textsuperscript{c}Singapore Management University, Singapore
\textsuperscript{d}University of Southampton, United Kingdom
\textsuperscript{e}Korea University, Republic of Korea

\begin{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.
\end{abstract}

\begin{articleinfo}
\begin{articlehistory}
Received 17 September 2014
Accepted 26 November 2014
Available online 29 December 2014
\end{articlehistory}
JEL classification:
C230
C360

Keywords:
Dynamic panel
IV estimation
Levels and difference instruments

\end{articleinfo}

\section{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 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. The present work deals with the stationary case. See Han and Phillips (2013), Moon and Phillips (2004) and the references therein for recent work on the panel unit root case.

\section{Asymptotic distributions of IV estimators}
For the simple panel dynamic model $y_{it} = \alpha_i + \beta y_{i,t-1} + u_{it}$, 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

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

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

\textbf{Levels IV}

Let $y_{it} = \alpha_i + \beta y_{i,t-1} + u_{it}$ ($i = 1, \ldots, n; t = 1, \ldots, T$) with $|\beta| < 1$, $u_{it} \sim \text{i.i.d.} (0, \sigma_u^2)$, $y_{i0} = \frac{\alpha_i}{1-\beta} + \sum_{j=1}^{\infty} \rho^j u_{i,-j}$, and $\alpha_i \sim \text{i.i.d.} (0, \sigma_\alpha^2)$.

\textsuperscript{*}Phillips acknowledges support from the NSF under Grant No. SES 12-58258. Research by Han was supported by the Korean Government (2014S1A2A2027803) and Korea University (K1421311).
\textsuperscript{*}Corresponding author at: Yale University, United States.
E-mail address: peter.phillips@yale.edu (P.C.B. Phillips).

http://dx.doi.org/10.1016/j.econlet.2014.11.030
0165-1765/© 2014 Elsevier B.V. All rights reserved.
independent of $u_t$. The levels IV estimator $\hat{\beta}_l$ satisfies

$$\sqrt{n}(\hat{\beta}_l - \beta) = \frac{1}{\sqrt{n}} \left( \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} y_{it-2} - \Delta u_{it} \right) = \frac{N}{nT},$$

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

$$\sqrt{n}(\hat{\beta}_l - \beta) \Rightarrow n \rightarrow \infty \left( 0, \frac{2\sigma^2 T_1 \sigma^2 + 2\sigma^2 \left[ \frac{\sigma_1^2}{1-\beta} + \frac{\beta^2}{1+\beta} \right]}{T_1^2 (1-\beta) + \frac{\beta}{1+\beta}} \right).$$

The asymptotic variance of (5) increases with $\sigma_1^2/\sigma^2$, which is natural because the $\alpha_i$ are uninformative for the identification of $\beta$. As $T \rightarrow \infty$

$$\sqrt{n}(\hat{\beta}_l - \beta) \Rightarrow (T,n) \rightarrow \infty N(0, 2(1+\beta)).$$

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 $\sigma_1^2$. Also, note that there is a discontinuity in the limit theory as $\beta \rightarrow 1$ 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 $y_t - \Delta u_t$ 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 equivalence, algebra shows that the HP estimator $\hat{\beta}_H$ can be written in the form that relates to $\hat{\beta}_l$ with end corrections, viz.,

$$\hat{\beta}_H = \frac{1}{\sqrt{n}} \left( \sum_{t=1}^{T} \sum_{i=1}^{n} y_{it-2} \Delta y_{it} + w_1 + w_2 \right) \left( \sum_{t=1}^{T} \sum_{i=1}^{n} y_{it-2} \Delta y_{it-1} + w_1 \right),$$

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

**Difference IV**

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

$$\sqrt{n}(\hat{\beta}_d - \beta) = \frac{1}{\sqrt{n}} \left( \sum_{t=1}^{T} \sum_{i=1}^{n} y_{it-2} \Delta y_{it} \right) = \frac{N}{nT}.$$
\[
\begin{align*}
\Delta(\hat{Y}_t) &= \sqrt{n} \left[ \frac{2\sigma^2}{1 + \beta^2} + \frac{\sigma_a^2}{(1 - \beta)^2} \right] - \left( \frac{\sigma_a^2}{(1 - \beta)^2} + \frac{\sigma^2}{1 - \beta^2} \right) \\
&= -T_2 \sigma_a^2 (1 - \beta)^2 + T_2 \sigma^2 (1 - \beta) + D_{\hat{Y}} \left( \frac{T_2}{T_1} \right) (1 - \beta).
\end{align*}
\]

Fig. 1. Asymptotic variance ratio \( \text{Avar}(\hat{\beta}_1) / \text{Avar}(\hat{\beta}_2) \).

Comparing (10) with (6), it is clear that the limit variance of \( \hat{\beta}_1 \) for large \( n \) and \( T \) is smaller than that of \( \hat{\beta}_2 \) 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
The level estimator is more efficient for all cases. Larger \( \sigma \) (vergence). \( (9) \) are both sequential, \( n \to \infty \) followed by \( T \to \infty \) without restriction on the rates or the path of divergence.

For fixed \( T \), from \( (5) \) and \( (9) \), we evaluate \( \text{Avar} (\hat{\beta}) / \text{Avar} (\hat{\beta}_d) \) for various \( \beta \), \( T \) and \( \sigma^2 / \sigma^2 \). Fig. 1(a) exhibits this ratio for \( \sigma^2 / \sigma^2 = 1 \). The levels estimator is more efficient for all \( \beta \) and \( T \geq 3 \) in this case. Larger \( \sigma^2 / \sigma^2 \) 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 \( \sigma^2 / \sigma^2 \leq 4 \). For even larger \( \sigma^2 / \sigma^2 \), Fig. 1(b) considers \( \sigma^2 / \sigma^2 = 8 \). The difference AH estimator performs better than the levels estimator only for small \( T \) and large \( \beta \). Numerical evaluations show that the levels estimator is more efficient than the difference estimator for all \( \beta \) for \( T \geq 8 \).

3. Simulations

Table 1 presents the simulated variances \( (n \to 400, \beta = 0.5, 10,000 \text{ replications}) \). \( y_i = \alpha_i + \beta y_{i-1} + u_i, \alpha_i = \sigma, \sigma = \sigma, \sigma_i = 1, y_{-100} = \alpha_i (1 - \beta) + u_i / \sqrt{1 - \beta^2}, \alpha_i, u_i \sim N(0, 1) \). For fixed \( T \), from \( (5) \) and \( (9) \), we evaluate \( \text{Avar}(\hat{\beta}) / \text{Avar}(\hat{\beta}_d) \) for various \( \beta \), \( T \) and \( \sigma^2 / \sigma^2 \). Fig. 1(a) exhibits this ratio for \( \sigma^2 / \sigma^2 = 1 \). The levels estimator is more efficient for all \( \beta \) and \( T \geq 3 \) in this case. Larger \( \sigma^2 / \sigma^2 \) 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 \( \sigma^2 / \sigma^2 \leq 4 \). For even larger \( \sigma^2 / \sigma^2 \), Fig. 1(b) considers \( \sigma^2 / \sigma^2 = 8 \). The difference AH estimator performs better than the levels estimator only for small \( T \) and large \( \beta \). Numerical evaluations show that the levels estimator is more efficient than the difference estimator for all \( \beta \) for \( T \geq 8 \).

\[
(\hat{\alpha}^2 / \sigma^2)_{\beta} /
\]

\[
\text{Avar}(\hat{\beta}) = \frac{\sigma^2 e^{-2 \beta \lambda T}}{(1 - \beta \lambda)^2}
\]

\[
\text{Avar}(\hat{\beta}_d) = \frac{\sigma^2 e^{-2 \beta \lambda T}}{(1 - \beta \lambda)^2}
\]

\[
\frac{\text{Avar}(\hat{\beta})}{\text{Avar}(\hat{\beta}_d)} = \frac{(1 + \beta)^2}{T^2 (1 - \beta)} \left[ \frac{\sigma^2}{\sigma^2} + \frac{\beta}{1 + \beta} \right]
\]

which has a substantial impact on variance when \( \sigma^2 / \sigma^2 \) 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


