DOI: 10.6277/TER.2014.422.4

Gaussian Inference in AR(1) Models with Trend: A Note

Jhih-Gang Chen and Biing-Shen Kuo*

This paper adapts the first-difference estimator of Phillips and Han (2008) to the estimation and inference in AR(1) models with trends. With a detrending procedure, the first-difference estimator remains applicable and is shown to retain the Gaussian asymptotics. A unit root test based on the estimator is more powerful than that based on the double-difference estimator. The proposed estimator is especially useful when applied to dynamic panels.

Keywords: AR model, detrend, first difference, unit root, dynamic panel, fixed effects
JEL classification: C13, C22, C23

1 Introduction

As is well known, the conventional least squares estimator in autoregressive models may be subject to two shortcomings. First, the estimates of the autoregressive coefficients are biased downwards (Marriott and Pope, 1954; Shaman and Stine, 1988). This bias could be critical when the sample is relatively small, and will tend to reduce the prediction accuracy and inference credibility in applications. In particular, the bias problem is much more severe when there are trend components in the models (van Giersbergen, 2005). Second, the least squares estimator converges at a faster T rate when a unit root is present in the series. The associated asymptotic distribution is

*Both authors are from Department of International Business, National Chengchi University. They thank the editor and two anonymous referees for helpful comments and suggestions. The usual disclaimer applies. Financial support from National Science Council, Taiwan, is also gratefully acknowledged (under grant no. NSC 98-2410-H-004-039-MY2).

期刊資料庫

found to be nonstandard, and statistical inference has to rely on the simulated critical values (cf. Dickey and Fuller, 1979).

Recently, Phillips and Han (2008) have proposed a first-difference based approach to estimation and inference for the AR(1) models. The estimator is found to exhibit little finite-sample bias and to possess standard Gaussian asymptotics, regardless of whether or not the series contains a unit root. When an incident trend is present in the model, they suggest a double-difference approach to eliminate the nuisance parameters. However, what is being estimated is no longer the autoregressive coefficient, but a nonlinear transformation of it in this case. Although inference could be indirectly performed, a power deficiency might occur due to the nonlinear relationship as a result of double-differencing.

This paper focuses on the estimation and inference of AR(1) models with trend. To remove the nuisance parameters, we suggest detrending the series first, in combination with the first-difference procedure of Phillips and Han (2008) proposed in the trendless model. The resulting estimator is found to retain the Gaussian asymptotic properties, and to sharply display its advantage over the double-difference estimator in terms of finite-sample bias and variance. Moreover, this approach is immune to the weak instrument problem for the autoregressive coefficient in the vicinity of unity, and thus appears to be more useful in dynamic panels with fixed effects and trends.

2 Model Specification

Consider a simple AR(1) model with trend:

$$y_t = \alpha + \gamma t + u_t,\ u_t = \rho u_{t-1} + \varepsilon_t, \tag{1}$$

in which $\rho \in (-1, 1]$ and $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$. The initial condition is set to be $u_{-1} = \sum_{j=0}^{\infty} \rho^j \varepsilon_{-1-j}$ when $|\rho| < 1$, and $u_{-1} = \sum_{j=0}^{\kappa} \varepsilon_{-1-j}$ for some fixed integer κ when $\rho = 1$ (e.g., Phillips and Magdalinos, 2009).[1] Alternatively, the model above can be expressed as

$$y_t = (1-\rho)\alpha + \rho\gamma + (1-\rho)\gamma t + \rho y_{t-1} + \varepsilon_t. \tag{2}$$

[1]The initial conditions have no effect on the asymptotic theory of the difference based estimators considered in this paper due to difference transformations.

臺灣大學學術期刊資料庫

Of particular concern here is the estimation of the autoregressive coefficient, ρ. We begin with the first-difference transformation of (2):

$$\triangle y_t = (1-\rho)\gamma + \rho \triangle y_{t-1} + \triangle \varepsilon_t, \tag{3}$$

where $\triangle = 1 - \mathcal{L}$, and $\mathcal{L}$ is the backward shift operator. Directly applying least squares to (3) yields inconsistency because $\mathrm{E}[\triangle y_{t-1} \triangle \varepsilon_t] = -\sigma^2 \neq 0$. A common way to deal with this problem is by applying the instrumental variable estimation (GMM) based on the orthogonality conditions that $\mathrm{E}[y_s \triangle \varepsilon_t] = 0$, $s = t-2, t-3, \cdots$. However, when ρ is close to unity, the GMM estimator suffers from a weak instrument problem, leading to substantial finite-sample biases.

3 Difference Estimator with Detrending

When the trend coefficient (γ) is known, however, the estimation problem can be tackled in the same way as by Phillips and Han (2008) in the trendless model. Subtracting γ from both sides of (3) yields

$$\triangle y_t - \gamma = \rho(\triangle y_{t-1} - \gamma) + \triangle \varepsilon_t. \tag{4}$$

Certainly, the least-squares estimation of (4) remains inconsistent. Instead, we follow Phillips and Han (2008) to make a further transformation:

$$2(\triangle y_t - \gamma) + (\triangle y_{t-1} - \gamma) = \rho(\triangle y_{t-1} - \gamma) + \eta_t, \tag{5}$$

with $\eta_t = 2(\triangle y_t - \gamma) + (1-\rho)(\triangle y_{t-1} - \gamma)$. Now, this first-difference transformed regression produces a new moment condition whereby $\mathrm{E}[(\triangle y_{t-1} - \gamma)\eta_t] = 0$, giving the consistency. It is worth noting that this moment condition is valid even when y_t contains a unit root, and thus the weak instrument problem can be avoided.

Applying the least squares principle to (5) leads to

$$\hat{\rho}_{\mathrm{FD}} = \frac{\sum_{t=1}^{T} 2(\triangle y_{t-1} - \gamma)(\triangle y_t - \gamma)}{\sum_{t=1}^{T} (\triangle y_{t-1} - \gamma)^2} + 1, \tag{6}$$

where y_t is assumed to be observed at $t = -1, 0, \cdots, T$. The asymptotic properties of $\hat{\rho}_{\mathrm{FD}}$ are summarized as follows:

Corollary 1. For all $\rho \in (-1, 1]$, $\sqrt{T}(\hat{\rho}_{\mathrm{FD}} - \rho) \Rightarrow N(0, 2(1+\rho))$ as $T \to \infty$.

臺灣大學學術期刊資料庫

In practice, the information regarding the trend coefficient is usually unavailable, and thus $\hat{\rho}_{\mathrm{FD}}$ is infeasible. A direct way to resolve the problem is to replace γ in (6) with its consistent estimate. This is identical to applying the first-difference estimator of Phillips and Han (2008) developed in the trendless models to the detrended data $y_t - \hat{\gamma}t$. One natural choice of the consistent estimates is the sample mean of the first-difference series, given the fact that $\mathrm{E}(\triangle y_t) = \gamma$. The resulting first-difference estimator with detrending is then

$$\hat{\rho}_{\mathrm{FFD}} = \frac{\sum_{t=1}^{T} 2\left(\triangle y_{t-1} - \hat{\gamma}\right)\left(\triangle y_t - \hat{\gamma}\right)}{\sum_{t=1}^{T}\left(\triangle y_{t-1} - \hat{\gamma}\right)^2} + 1,$$

with $\hat{\gamma} = \sum_{t=0}^{T} \triangle y_t/(T+1)$.

Theorem 1. For all $\rho \in (-1, 1]$, we have $\sqrt{T}(\hat{\rho}_{\mathrm{FFD}} - \rho) \overset{d}{\equiv} \sqrt{T}(\hat{\rho}_{\mathrm{FD}} - \rho)$, where $\overset{d}{\equiv}$ denotes equality in terms of the asymptotic distribution.

The theorem states that the asymptotic distribution remains unchanged when the trend coefficient needs to be estimated. This results from the fact that the auto-covariance functions of the difference series are not altered asymptotically by the detrending procedure. Besides, the theorem also suggests a consistent estimator of variance of $\hat{\rho}_{\mathrm{FFD}}$, say $2(1 + \hat{\rho}_{\mathrm{FFD}})$, that is useful when constructing test statistics.

On the contrary, Phillips and Han (2008) adopt a double-difference method to remove both intercept and trend from the models. In detail, they consider the following transformation of (2):

$$2\triangle^2 y_t + \triangle^2 y_{t-1} = \theta \triangle^2 y_{t-1} + \tilde{\eta}_t, \quad \tilde{\eta}_t = 2\triangle^2 y_t + (1-\theta)\triangle^2 y_{t-1},$$

where

$$\theta = -\frac{(1-\rho)^2}{3-\rho}. \tag{7}$$

With the orthogonal condition that $\mathrm{E}(\triangle^2 y_{t-1}\tilde{\eta}_t) = 0$ for all $\rho \in (-1, 1]$, the least squares estimator $\hat{\theta}$ is consistent. Moreover, it is proven that $\sqrt{T}(\hat{\theta} - \theta) \Rightarrow N(0, V_\rho)$ with V_ρ being some function of ρ.[2]

[2]The explicit form of V_ρ is indeed complicated and can be found in Phillips and Han (2008).

臺灣大學學術期刊資料庫

While the nuisance parameters are eliminated, double-differencing reduces the signaling information carried by the regressor further than first-differencing, and thus leads to a deficiency when testing. Observing the relation between ρ and θ in (7), it is apparent that an $O(T^{-1/4})$ neighborhood of $\rho = 1$ corresponds to an $O(T^{-1/2})$ neighborhood of $\theta = 0$. To be specific, in the case that $\rho = 1$, $\hat{\theta}$ converges at the rate $T^{1/2}$, while $\hat{\rho}$, recovered from $\hat{\theta}$, converges at the rate $T^{1/4}$. Even though the deficiency fades as ρ departs from unity, tests based on the double-difference estimator can well be expected to be less powerful than that based on the first-difference estimator with detrending. We investigate this issue further through Monte Carlo simulations in the following section.

4 Monte Carlo Simulation

In this section, we compare the finite-sample performance of the "infeasible" first-difference estimator ($\hat{\rho}_{\mathrm{FD}}$), the "feasible" first-difference estimator with detrending ($\hat{\rho}_{\mathrm{FFD}}$), and the double-difference estimator ($\hat{\theta}$) by means of simulation. The DGP is (1) where we arbitrarily set $\alpha = \gamma = 1$ and $\varepsilon_t \overset{iid}{\sim} N(0, 1)$. The number of replications is 50, 000.

Table 1 shows the simulation results of bias, variance, and mean squared errors (MSE) of each estimator. Compared with $\hat{\rho}_{\mathrm{FD}}$, the estimation of the trend coefficient (γ) brings very little additional bias to $\hat{\rho}_{\mathrm{FFD}}$, and appears to be costless in terms of variance. As a result, the two first-difference estimators share similarity under MSE criterion.

Unlike the other estimators, $\hat{\rho}_{\hat{\theta}}$ tends to have severe bias in all cases considered. It should be noted that $\hat{\rho}_{\hat{\theta}}$ is obtained by inverting $\hat{\theta}$ through (7). The bias can thus be partly attributed to the inversion.[3] As expected, $\hat{\rho}_{\hat{\theta}}$ is also less efficient due to signal loss from differencing twice and thus tends to have the highest MSE.

Table 2 evaluates the accuracy of the normality approximation. Overall speaking, the asymptotic variance provides fairly accurate approximation to the variance of the finite-sample distribution of each estimator. The same is true to the corresponding t-statistics built on the estimators, as their respective variances are close to that of the standard normal. The last three

[3]According to simulations that are not reported here, $\hat{\theta}$ has only mild bias in all cases. The results for the bias, variance, and MSE of $\hat{\theta}$ are available from the authors on request.

臺灣大學學術期刊資料庫

Table 1: Finite Sample Performance

		10 × Bias			Variance			MSE		
ρ	T	$\hat{\rho}_{\mathrm{FD}}$	$\hat{\rho}_{\mathrm{FFD}}$	$\hat{\rho}_{\hat{\theta}}$	$\hat{\rho}_{\mathrm{FD}}$	$\hat{\rho}_{\mathrm{FFD}}$	$\hat{\rho}_{\hat{\theta}}$	$\hat{\rho}_{\mathrm{FD}}$	$\hat{\rho}_{\mathrm{FFD}}$	$\hat{\rho}_{\hat{\theta}}$
0	40	0.233	0.233	0.824	0.050	0.050	0.134	0.050	0.050	0.141
0	80	0.104	0.104	0.393	0.025	0.025	0.062	0.025	0.025	0.064
0	160	0.056	0.056	0.186	0.012	0.012	0.027	0.012	0.012	0.028
0	320	0.030	0.030	0.093	0.006	0.006	0.013	0.006	0.006	0.013
0.5	40	0.165	0.157	0.864	0.072	0.072	0.167	0.072	0.072	0.169
0.5	80	0.068	0.066	0.648	0.037	0.037	0.119	0.037	0.037	0.121
0.5	160	0.024	0.023	0.509	0.018	0.018	0.081	0.019	0.019	0.084
0.5	320	0.020	0.020	0.206	0.009	0.009	0.051	0.009	0.009	0.052
0.9	40	−0.025	−0.096	−2.034	0.090	0.091	0.136	0.090	0.091	0.177
0.9	80	−0.007	−0.037	−1.605	0.047	0.047	0.094	0.047	0.047	0.120
0.9	160	−0.003	−0.010	−1.188	0.023	0.023	0.066	0.023	0.023	0.080
0.9	320	−0.000	−0.003	−0.859	0.012	0.012	0.045	0.012	0.012	0.053
0.95	40	0.011	−0.207	−2.482	0.093	0.094	0.134	0.093	0.094	0.196
0.95	80	−0.014	−0.076	−2.044	0.048	0.048	0.093	0.048	0.048	0.135
0.95	160	−0.007	−0.022	−1.617	0.024	0.024	0.064	0.024	0.024	0.090
0.95	320	−0.003	−0.007	−1.273	0.012	0.012	0.044	0.012	0.012	0.060
1	40	0.015	−0.483	−2.955	0.096	0.096	0.133	0.096	0.098	0.221
1	80	0.004	−0.247	−2.488	0.049	0.049	0.091	0.049	0.049	0.153
1	160	0.001	−0.125	−2.080	0.025	0.025	0.064	0.025	0.025	0.107
1	320	−0.003	−0.066	−1.751	0.013	0.013	0.043	0.013	0.013	0.075

Note: DGP: $y_t = 1 + (1-\rho)t + \rho y_{t-1} + \varepsilon_t$, $t = 0, \cdots, T$, $\varepsilon_t \overset{iid}{\sim} N(0,1)$. The number of replications is 50,000. $\hat{\rho}_{\hat{\theta}}$ denotes the estimator of ρ recovered from the double-difference estimator $\hat{\theta}$, through the relationship specified in (7).

columns present the empirical sizes of the t-ratio under the null hypothesis that ρ is equal to its true value. It is evident that either t_{FD} or t_{FFD} has empirical size much closer to the nominal level (5%) than that of t_θ.

Figure 1 further displays power curves of each t-ratio under the unit root null. It is not surprising that t_{FD} and t_{FFD} have nearly the same power performance, and reveal a significant dominance over that of t_θ. These results are consistent with the prediction of the theory discussed earlier.

臺灣大學圖書館 期刊資料庫

Table 2: Normality Approximations

$\rho = 0 (H_\rho = 2),\ \theta = -0.333 (V_\rho = 1.210)$

T	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FD}})$	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FFD}})$	$T\,\mathrm{var}(\hat{\theta})$	$\mathrm{var}(t_{\mathrm{FD}})$	$\mathrm{var}(t_{\mathrm{FFD}})$	$\mathrm{var}(t_{\hat{\theta}})$	$\mathrm{size}_{\mathrm{FD}}$	$\mathrm{size}_{\mathrm{FFD}}$	$\mathrm{size}_{\hat{\theta}}$
40	1.993	1.989	1.264	0.996	0.995	1.045	2.612	2.596	2.264
80	1.997	1.996	1.232	0.999	0.998	1.018	3.368	3.366	2.982
160	1.991	1.991	1.220	0.996	0.996	1.008	3.884	3.884	3.650
320	1.983	1.983	1.207	0.992	0.992	0.998	4.078	4.078	3.910

$\rho = 0.5 (H_\rho = 3),\ \theta = -0.1 (V_\rho = 1.751)$

T	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FD}})$	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FFD}})$	$T\,\mathrm{var}(\hat{\theta})$	$\mathrm{var}(t_{\mathrm{FD}})$	$\mathrm{var}(t_{\mathrm{FFD}})$	$\mathrm{var}(t_{\hat{\theta}})$	$\mathrm{size}_{\mathrm{FD}}$	$\mathrm{size}_{\mathrm{FFD}}$	$\mathrm{size}_{\hat{\theta}}$
40	2.890	2.889	1.763	0.964	0.963	1.007	3.492	3.516	2.584
80	2.956	2.955	1.757	0.985	0.985	1.003	4.212	4.236	3.344
160	2.959	2.959	1.752	0.986	0.986	1.001	4.414	4.424	3.904
320	2.958	2.958	1.744	0.986	0.986	0.996	4.468	4.468	4.056

$\rho = 0.9 (H_\rho = 3.8),\ \theta = -0.005 (V_\rho = 1.990)$

T	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FD}})$	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FFD}})$	$T\,\mathrm{var}(\hat{\theta})$	$\mathrm{var}(t_{\mathrm{FD}})$	$\mathrm{var}(t_{\mathrm{FFD}})$	$\mathrm{var}(t_{\hat{\theta}})$	$\mathrm{size}_{\mathrm{FD}}$	$\mathrm{size}_{\mathrm{FFD}}$	$\mathrm{size}_{\hat{\theta}}$
40	3.617	3.640	1.985	0.951	0.958	0.998	4.056	4.448	2.612
80	3.732	3.737	1.988	0.982	0.984	0.999	4.692	4.822	3.350
160	3.738	3.739	1.981	0.984	0.984	0.996	4.804	4.840	3.926
320	3.746	3.747	1.974	0.986	0.986	0.992	4.818	4.850	4.046

$\rho = 0.95 (H_\rho = 3.9),\ \theta = -0.001 (V_\rho = 1.997)$

T	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FD}})$	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FFD}})$	$T\,\mathrm{var}(\hat{\theta})$	$\mathrm{var}(t_{\mathrm{FD}})$	$\mathrm{var}(t_{\mathrm{FFD}})$	$\mathrm{var}(t_{\hat{\theta}})$	$\mathrm{size}_{\mathrm{FD}}$	$\mathrm{size}_{\mathrm{FFD}}$	$\mathrm{size}_{\hat{\theta}}$
40	3.712	3.748	1.993	0.952	0.961	0.998	4.098	4.860	2.620
80	3.828	3.841	1.996	0.982	0.984	0.999	4.708	5.014	3.342
160	3.835	3.838	1.989	0.983	0.984	0.996	4.834	4.930	3.922
320	3.847	2.848	1.981	0.986	0.987	0.992	4.800	4.836	4.032

$\rho = 1 (H_\rho = 4),\ \theta = 0 (V_\rho = 2)$

T	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FD}})$	$T\,\mathrm{var}(\hat{\rho}_{\mathrm{FFD}})$	$T\,\mathrm{var}(\hat{\theta})$	$\mathrm{var}(t_{\mathrm{FD}})$	$\mathrm{var}(t_{\mathrm{FFD}})$	$\mathrm{var}(t_{\hat{\theta}})$	$\mathrm{size}_{\mathrm{FD}}$	$\mathrm{size}_{\mathrm{FFD}}$	$\mathrm{size}_{\hat{\theta}}$
40	3.850	3.831	2.007	0.963	0.958	1.004	4.242	5.826	2.661
80	3.909	3.908	1.985	0.977	0.977	0.992	4.582	5.736	3.300
160	3.952	3.952	2.004	0.988	0.988	1.002	4.946	5.778	3.880
320	4.018	4.020	2.011	1.005	1.005	1.005	5.038	5.606	4.260

Note: $H_\rho = 2(1+\rho)$ denotes the asymptotic variance shared by $\hat{\rho}_{\mathrm{FD}}$ and $\hat{\rho}_{\mathrm{FFD}}$, while V_ρ denotes that of $\hat{\theta}$. Values reported in the last three columns are the empirical sizes (in percentage) of the t-tests performed at the nominal significance level of 5%. The t-statistics are constructed as $t_{\mathrm{FD}} = \sqrt{T}(\hat{\rho}_{\mathrm{FD}} - \rho)/\sqrt{2(1+\hat{\rho}_{\mathrm{FD}})}$, $t_{\mathrm{FFD}} = \sqrt{T}(\hat{\rho}_{\mathrm{FFD}} - \rho)/\sqrt{2(1+\hat{\rho}_{\mathrm{FFD}})}$, and $t_{\hat{\theta}} = \sqrt{T}(\hat{\theta} - \theta)/\sqrt{\hat{V}_\rho}$, where $\hat{V}_\rho$ is obtained by replacing ρ with $\hat{\rho}_{\hat{\theta}}$ in V_ρ.

臺灣大學圖書館 期刊資料庫

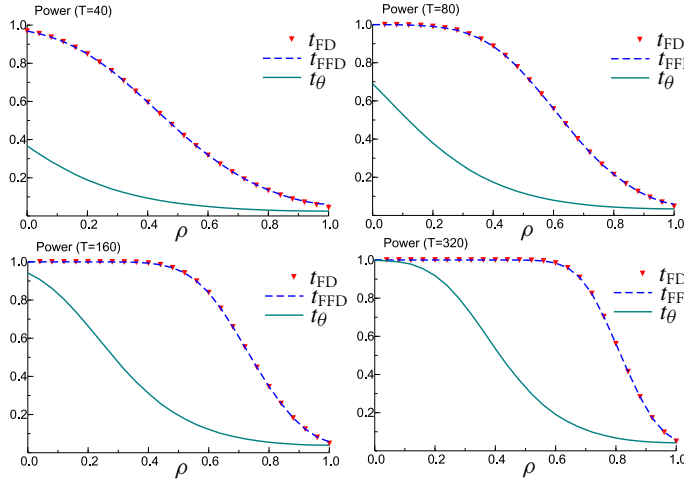


Figure 1: Power Curves of Unit Root Tests

5 Extension to Dynamic Panels

The difference based approach possesses very useful extension to dynamic panels with fixed effects and trends.[4] To illustrate, consider the panel version of (2):

$$y_{it} = (1-\rho)\alpha_i + \rho\gamma_i + (1-\rho)\gamma_i t + \rho y_{it-1} + \varepsilon_{it}, \tag{8}$$

where $\rho \in (-1, 1]$, $\varepsilon_{it} \overset{iid}{\sim} (0, \sigma^2)$, $i = 1, \cdots, n$, and $t = 0, \cdots, T$. Nickell (1981) demonstrates that the least squares dummy variable estimator of ρ is inconsistent when the time span (T) is short. The conventional way is to eliminate the fixed effects by first-differencing which leads to

$$\triangle y_{it} - \gamma_i = \rho(\triangle y_{it-1} - \gamma_i) + \triangle\varepsilon_{it}.$$

When γ_i is known, the instrumental variable method or GMM can be applied with the moment conditions that $\mathrm{E}[(y_{is} - \gamma_i t)\triangle\varepsilon_{it}]$, $s = t-2, t-3, \cdots$. Unfortunately, as noted by Blundell and Bond (1998), the moment conditions become weak near the unit circle and is even irrelevant to ρ in

[4]The proposed approach can also be extended to AR(p) models in a similar manner of Phillips and Han (2008). However, the transformation becomes computationally involved, as the autoregressive order p increases.

臺灣大學圖書館 期刊資料庫

the unit root case. As a result, the estimators may be inconsistent even when T is large and/or have nonstandard behaviors.

In contrast, we can adopt the transformation of Phillips and Han (2008), so that

$$2(\triangle y_{it} - \gamma_i) + (\triangle y_{it-1} - \gamma_i) = \rho(\triangle y_{it-1} - \gamma_i) + \eta_{it},$$

with $\eta_{it} = 2(\triangle y_{it} - \gamma_i) + (1 - \rho)(\triangle y_{it-1} - \gamma_i)$. Notably, the implied orthogonal condition, $\mathrm{E}(\triangle y_{it-1}\eta_{it}) = 0$, is strong in all of the stationary, near-integration, and unit root cases. The associated pooled least squares estimator now takes the form:

$$\hat{\rho}_{\mathrm{PFD}} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} 2(\triangle y_{it-1} - \gamma_i)(\triangle y_{it} - \gamma_i)}{\sum_{i=1}^{n}\sum_{t=1}^{T}(\triangle y_{it-1} - \gamma_i)^2} + 1. \tag{9}$$

Han and Phillips (2010) show that $\hat{\rho}_{\mathrm{PFD}}$ has Gaussian limit with a convergence rate of $\sqrt{nT}$. Specifically, as $T \to \infty$, irrespective of whether or how n passes to infinity,[5]

$$\sqrt{nT}\left(\hat{\rho}_{\mathrm{PFD}} - \rho\right) \Rightarrow N(0, 2(1 + \rho)).$$

Since the limit theory of the estimator is valid for all $\rho \in (0, -1]$, it has a direct application to panel unit root testing.

In the cases of unknown trend coefficients, as suggested by our time series results, replacing γ_i in (9) by some consistent estimator is feasible. Yet, the asymptotics of the resulting estimator would depend on the specifications of trend coefficients.

When the trend is common across series, i.e. $\gamma_i = \gamma$, the coefficient can be consistently estimated by $\hat{\gamma} = \sum_{i=1}^{n}\sum_{t=0}^{T} \triangle y_{it}/(n(T+1))$ which leads to the feasible estimator

$$\hat{\rho}^{c}_{\mathrm{PFFD}} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} 2\left(\triangle y_{it-1} - \hat{\gamma}\right)\left(\triangle y_{it} - \hat{\gamma}\right)}{\sum_{i=1}^{n}\sum_{t=1}^{T}(\triangle y_{it-1} - \hat{\gamma})^2} + 1.$$

Because $(\hat{\gamma} - \gamma)$ is $O(n^{-1/2}T^{-1/2})$, it is not difficult to obtain the following result using similar arguments in proving Theorem 1.

Corollary 2. For all $\rho \in (-1, 1]$, we have $\sqrt{nT}(\hat{\rho}^{c}_{\mathrm{PFFD}} - \rho) \overset{d}{\equiv} \sqrt{nT}(\hat{\rho}_{\mathrm{PFD}} - \rho)$ irrespective of how the composite nT goes to infinity.

[5]Han and Phillips (2010) develop asymptotic theory in the trendless case.

臺灣大學學術期刊資料庫

When the trends are heterogeneous, cross-sectional information proves not very useful for estimation of each individual trend coefficient.[6] Even though, each γ_i can be estimated the same as in the time series case, $\hat{\gamma}_i = \sum_{t=0}^{T} \triangle y_{it}/(T+1)$, and the resulting feasible estimator becomes

$$\hat{\rho}^h_{\text{PFFD}} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} 2\left(\triangle y_{it-1} - \hat{\gamma}_i\right)\left(\triangle y_{it} - \hat{\gamma}_i\right)}{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\triangle y_{it-1} - \hat{\gamma}_i\right)^2} + 1.$$

Since $(\hat{\gamma}_i - \gamma_i)$ is $O(T^{-1/2})$, the Gaussian limit for the pooled estimator remains achievable only when $T \to \infty$, and the next result follows directly.

Corollary 3. For any of the cases that

(i) Given any fixed n, as $T \to \infty$,
(ii) As $T \to \infty$ and then $n \to \infty$, or
(iii) As T and n pass to infinity jointly and $\sqrt{n/T} \to 0$,

$\sqrt{nT}(\hat{\rho}^h_{\text{PFFD}} - \rho) \Rightarrow N(0, 2(1+\rho))$ for all $\rho \in (-1, 1]$.

We conduct some simulations to examine the finite sample performance of the proposed estimators and the panel version of double difference estimator (denoted by $\tilde{\theta}$) when additional cross sectional data is available.[7] The considered DGP consists of (8) with $\alpha_i \overset{iid}{\sim} N(0,1)$, $\gamma_i = 1$, and $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$. Since $\hat{\rho}^h_{\text{PFFD}}$ appears to be more suitable for long-narrow panels as indicated by Corollary 3, the number of cross-sections is set as $n = 2$.

Table 3 shows the simulation results on bias, variance, and MSE for each estimator of interest. Similar to that in the time series cases, estimation of the trend coefficient (γ_i) brings little additional bias to both $\hat{\rho}^c_{\text{PFFD}}$ and $\hat{\rho}^h_{\text{PFFD}}$.

[6]The case with heterogeneous trends is less economically meaningful because there exists an infinite deviation of one series from others in the long-run.

[7]The panel version of double difference estimator, $\tilde{\theta}$, is simply the pooled least squares estimator for the transformed model that $2\triangle^2 y_{it} + \triangle^2 y_{it-1} = \theta \triangle^2 y_{it-1} + \tilde{\eta}_{it}$, where $\tilde{\eta}_{it} = 2\triangle^2 y_{it} + (1-\theta)\triangle^2 y_{it-1}$ and $\theta = -(1-\rho)^2/(3-\rho)$. As proven by Han and Phillips (2010), given any $T \geq 3$ and as $n \to \infty$, $\sqrt{nT}(\tilde{\theta} - \theta) \Rightarrow N(0, V_{\rho,T})$ $\forall \rho \in (-1, 1]$, where $V_{\rho,T}$ is some function of ρ and T. Besides, irrespective of whether or how n passes to infinity, as $T \to \infty$, $\sqrt{nT}(\tilde{\theta} - \theta) \Rightarrow N(0, V_\rho)$, where $V_\rho = \lim_{T\to\infty} V_{\rho,T}$ does not depends on n.

臺灣大學圖書館 期刊資料庫

However, $\hat{\rho}^h_{\text{PFFD}}$ is a little more biased than $\hat{\rho}^c_{\text{PFFD}}$ in large-ρ cases because $\hat{\rho}^c_{\text{PFFD}}$ has the information advantage that $\gamma_i = \gamma$ across sections, while $\hat{\rho}^h_{\text{PFFD}}$ does not. Even though, the three first-difference estimators perform similarly, judging by either variance or MSE. The performance of $\hat{\rho}_{\tilde{\theta}}$ remains to be not as comparable as that of each of the first-difference estimators by all standards, mostly due to signal loss from twice differencing.

On the other hand, Table 4 clearly demonstrates that the limit normality approximation to the small-sample distributions of each suggested pooled estimators and the corresponding t-statistics be quite accurate. This suggests the use of asymptotic panel unit root tests. Figure 2 displays the power curves of panel unit root t-tests associated with different pooled estimators. It is obvious that t_{PFD}, t^c_{PFFD}, and t^h_{PFFD} all shares similar power performance and fare noticeably better than t_θ. Compared with the time series results, power gains of each of the considered tests apparently come from the use of cross-sectional information. It thus proves the usefulness of the proposed estimator(s) with detrending within the dynamic panel framework.

6 Conclusion

This paper adapts the first-difference estimator of Phillips and Han (2008) to the estimation in AR(1) models with trend. The Gaussian asymptotics of the estimator remain unchanged. As expected, a unit root test based on the first-difference estimator with detrending is found to be more powerful than that based on the double-difference estimator. In this case, detrending prior to differencing is not only easy to implement, but also preserves more of the signaling information from the regressor than double-differencing. The proposed approach is especially useful when applied to dynamic panels.

臺灣大學學術期刊資料庫

Table 3: Finite Sample Performance - Panel Data Model

		10 × Bias				Variance				MSE			
ρ	T	$\hat{\rho}_{\text{PFD}}$	$\hat{\rho}^{c}_{\text{PFFD}}$	$\hat{\rho}^{h}_{\text{PFFD}}$	$\hat{\rho}_{\tilde{\theta}}$	$\hat{\rho}_{\text{PFD}}$	$\hat{\rho}^{c}_{\text{PFFD}}$	$\hat{\rho}^{h}_{\text{PFFD}}$	$\hat{\rho}_{\tilde{\theta}}$	$\hat{\rho}_{\text{PFD}}$	$\hat{\rho}^{c}_{\text{PFFD}}$	$\hat{\rho}^{h}_{\text{PFFD}}$	$\hat{\rho}_{\tilde{\theta}}$
0	40	0.116	0.116	0.116	0.420	0.025	0.025	0.025	0.064	0.025	0.025	0.025	0.066
0	80	0.061	0.061	0.061	0.195	0.012	0.012	0.012	0.027	0.012	0.012	0.012	0.028
0	160	0.029	0.029	0.029	0.092	0.006	0.006	0.006	0.013	0.006	0.006	0.006	0.013
0	320	0.016	0.016	0.016	0.047	0.003	0.003	0.003	0.006	0.003	0.003	0.003	0.006
0.5	40	0.054	0.046	0.038	0.530	0.037	0.037	0.037	0.120	0.037	0.037	0.037	0.123
0.5	80	0.031	0.029	0.027	0.523	0.018	0.018	0.018	0.082	0.018	0.018	0.018	0.084
0.5	160	0.012	0.012	0.011	0.393	0.009	0.009	0.009	0.051	0.009	0.009	0.009	0.053
0.5	320	0.007	0.007	0.007	0.245	0.004	0.004	0.004	0.028	0.005	0.005	0.005	0.028
0.9	40	0.001	−0.059	−0.118	−1.581	0.046	0.046	0.046	0.095	0.046	0.046	0.047	0.120
0.9	80	0.006	−0.009	−0.024	−1.179	0.023	0.023	0.023	0.065	0.023	0.023	0.023	0.079
0.9	160	−0.002	−0.006	−0.010	−0.870	0.012	0.012	0.012	0.046	0.012	0.012	0.012	0.053
0.9	320	0.000	−0.001	−0.002	−0.598	0.006	0.006	0.006	0.032	0.006	0.006	0.006	0.036
0.95	40	−0.006	−0.113	−0.221	−2.020	0.047	0.048	0.048	0.094	0.047	0.048	0.048	0.013
0.95	80	0.003	−0.027	−0.058	−1.607	0.024	0.024	0.024	0.064	0.023	0.024	0.024	0.090
0.95	160	−0.005	−0.012	−0.020	−1.285	0.012	0.012	0.012	0.044	0.012	0.012	0.012	0.061
0.95	320	−0.001	−0.003	−0.005	−0.997	0.006	0.006	0.006	0.031	0.006	0.006	0.006	0.041
1	40	0.007	−0.241	−0.493	−2.471	0.049	0.049	0.049	0.092	0.049	0.049	0.051	0.153
1	80	0.003	−0.121	−0.246	−2.075	0.024	0.024	0.024	0.063	0.024	0.024	0.025	0.106
1	160	0.001	−0.061	−0.124	−1.746	0.012	0.012	0.012	0.043	0.012	0.012	0.013	0.074
1	320	0.000	−0.032	−0.063	−1.456	0.006	0.006	0.006	0.030	0.006	0.006	0.006	0.051

Note: DGP: $y_{it} = (1-\rho)\alpha_i + \rho\gamma_i + (1-\rho)\gamma_i t + \rho y_{it-1} + \varepsilon_{it}$, $\alpha_i \overset{iid}{\sim} N(0,1)$, $\gamma_i = 1$, $\varepsilon_t \overset{iid}{\sim} N(0,1)$, $t = 0, \cdots, T$, and $i = 1, 2$. The number of replications is 50,000. $\hat{\rho}_{\tilde{\theta}}$ denotes the estimator of ρ recovered from $\tilde{\theta}$ through the relationship specified in (7).

臺灣大學學術期刊資料庫

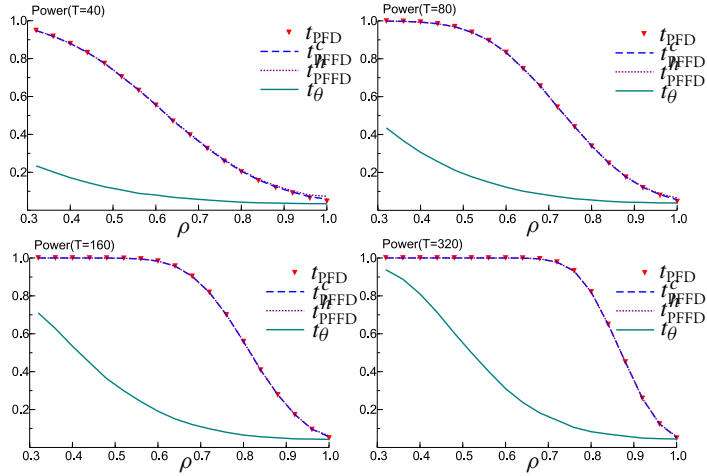


Figure 2: Power Curves of Panel Unit Root Tests

Appendix

Proof of Theorem 1. Rewrite the first-difference estimator with detrending as

$$\hat{\rho}_{\mathrm{FFD}} = \frac{\sum_{t=1}^{T} 2\left(\triangle y_{t-1} - \gamma + \gamma - \hat{\gamma}\right)\left(\triangle y_t - \gamma + \gamma - \hat{\gamma}\right)}{\sum_{t=1}^{T}\left(\triangle y_{t-1} - \gamma + \gamma - \hat{\gamma}\right)^2} + 1.$$

Note that $\triangle y_t$ is covariance stationary and $\mathrm{E}(\triangle y_t) = \gamma$. The central limit theorem for stationary process (CLT; Anderson, 1971, p. 429) implies that $(\hat{\gamma} - \gamma) = O(T^{-1/2})$. Then it is trivial to obtain $\sqrt{T}(\hat{\rho}_{\mathrm{FFD}} - \rho) \overset{d}{\equiv} \sqrt{T}(\hat{\rho}_{\mathrm{FD}} - \rho)$, where $\overset{d}{\equiv}$ denotes equality in terms of the asymptotic distribution.

臺灣大學學術期刊資料庫

Table 4: Normality Approximations - Panel Data Model

$\rho = 0 (H_\rho = 2),\ \theta = -0.333 (V_\rho = 1.210)$

T	T var$(\hat{\rho}_{\text{PFD}})$	T var$(\hat{\rho}^c_{\text{PFFD}})$	T var$(\hat{\rho}^h_{\text{PFFD}})$	T var$(\tilde{\theta})$	var(t_{PFD})	var(t^c_{PFFD})	var(t^h_{PFFD})	var$(t_{\tilde{\theta}})$	size$_{\text{PFD}}$	size$^c_{\text{PFFD}}$	size$^h_{\text{PFFD}}$	size$_{\tilde{\theta}}$
40	2.007	2.005	2.004	1.254	1.054	1.053	1.052	1.090	3.794	3.788	3.780	3.594
80	1.991	1.991	1.990	1.224	1.020	1.020	1.020	1.037	4.076	4.072	4.064	3.752
160	2.012	2.012	2.012	1.229	1.018	1.018	1.018	1.029	4.298	4.298	4.296	4.132
320	2.021	2.021	2.021	1.223	1.017	1.017	1.017	1.018	4.464	4.462	4.464	4.214

$\rho = 0.5 (H_\rho = 3),\ \theta = -0.1 (V_\rho = 1.751)$

T	T var T var$(\hat{\rho}_{\text{PFD}})$	T var T var$(\hat{\rho}^c_{\text{PFFD}})$	T var T var$(\hat{\rho}^h_{\text{PFFD}})$	T var T var$(\tilde{\theta})$	var var(t_{PFD})	var var(t^c_{PFFD})	var var(t^h_{PFFD})	var var$(t_{\tilde{\theta}})$	size$_{\text{PFD}}$	size$^c_{\text{PFFD}}$	size$^h_{\text{PFFD}}$	size$_{\tilde{\theta}}$
40	2.946	2.945	2.944	1.770	1.031	1.031	1.031	1.059	4.534	4.572	4.606	3.722
80	2.953	2.953	2.952	1.756	1.009	1.009	1.009	1.025	4.622	4.634	4.654	3.956
160	2.981	2.981	2.981	1.771	1.006	1.006	1.006	1.021	4.670	4.672	4.682	4.262
320	3.017	3.017	3.017	1.775	1.012	1.012	1.012	1.017	4.814	4.814	4.814	4.482

$\rho = 0.9 (H_\rho = 3.8),\ \theta = -0.005 (V_\rho = 1.990)$

T	T var$(\hat{\rho}_{\text{PFD}})$	T var$(\hat{\rho}^c_{\text{PFFD}})$	T var$(\hat{\rho}^h_{\text{PFFD}})$	T var$(\tilde{\theta})$	var(t_{PFD})	var(t^c_{PFFD})	var(t^h_{PFFD})	var$(t_{\tilde{\theta}})$	size$_{\text{PFD}}$	size$^c_{\text{PFFD}}$	size$^h_{\text{PFFD}}$	size$_{\tilde{\theta}}$
40	3.689	3.700	3.710	1.996	1.019	1.022	1.025	1.055	4.908	5.246	5.602	3.802
80	3.711	3.714	3.717	1.984	1.001	1.002	1.003	1.022	4.806	4.946	5.064	4.072
160	3.734	3.734	3.735	2.001	0.995	0.995	0.995	1.018	4.912	4.952	4.996	4.280
320	3.780	3.780	3.780	2.011	1.001	1.001	1.001	1.017	5.024	5.044	5.074	4.472

to be continued

國立臺灣大學圖書館 期刊資料庫

continued

$\rho = 0.95(H_\rho = 3.9),\ \theta = -0.001(V_\rho = 1.997)$

T	$T\ \mathrm{var}(\hat{\rho}_{\mathrm{PFD}})$	$T\ \mathrm{var}(\hat{\rho}^c_{\mathrm{PFFD}})$	$T\ \mathrm{var}(\hat{\rho}^h_{\mathrm{PFFD}})$	$T\ \mathrm{var}(\tilde{\theta})$	$\mathrm{var}(t_{\mathrm{PFD}})$	$\mathrm{var}(t^c_{\mathrm{PFFD}})$	$\mathrm{var}(t^h_{\mathrm{PFFD}})$	$\mathrm{var}(t_{\tilde{\theta}})$	$\mathrm{size}_{\mathrm{PFD}}$	$\mathrm{size}^c_{\mathrm{PFFD}}$	$\mathrm{size}^h_{\mathrm{PFFD}}$	$\mathrm{size}_{\tilde{\theta}}$
40	3.785	3.803	3.819	2.004	1.019	1.024	1.028	1.055	5.028	5.610	6.250	3.790
80	3.809	3.817	3.824	1.991	1.001	1.003	1.005	1.022	4.880	5.108	5.330	4.078
160	3.831	3.833	3.835	2.009	0.995	0.995	0.996	1.018	4.954	5.032	5.128	4.286
320	3.875	3.876	3.876	2.018	1.000	1.000	1.000	1.017	5.008	5.048	5.072	4.458

$\rho = 1(H_\rho = 4),\ \theta = 0(V_\rho = 2)$

T	$T\ \mathrm{var}(\hat{\rho}_{\mathrm{PFD}})$	$T\ \mathrm{var}(\hat{\rho}^c_{\mathrm{PFFD}})$	$T\ \mathrm{var}(\hat{\rho}^h_{\mathrm{PFFD}})$	$T\ \mathrm{var}(\tilde{\theta})$	$\mathrm{var}(t_{\mathrm{PFD}})$	$\mathrm{var}(t^c_{\mathrm{PFFD}})$	$\mathrm{var}(t^h_{\mathrm{PFFD}})$	$\mathrm{var}(t_{\tilde{\theta}})$	$\mathrm{size}_{\mathrm{PFD}}$	$\mathrm{size}^c_{\mathrm{PFFD}}$	$\mathrm{size}^h_{\mathrm{PFFD}}$	$\mathrm{size}_{\tilde{\theta}}$
40	3.886	3.883	3.876	1.993	1.020	1.019	1.017	1.048	4.880	6.140	6.210	3.760
80	3.884	3.885	3.886	1.979	0.995	0.996	0.996	1.014	4.772	5.622	5.804	3.860
160	3.974	3.977	3.977	1.996	1.006	1.007	1.007	1.011	4.988	5.620	5.634	4.200
320	3.995	3.995	3.998	2.005	1.005	1.005	1.006	1.009	4.994	5.432	5.558	4.452

Note: $H_\rho = 2(1+\rho)$ stands for the same asymptotic variance for $\hat{\rho}_{\mathrm{PFD}}$, $\hat{\rho}^c_{\mathrm{PFFD}}$ and $\hat{\rho}^h_{\mathrm{PFFD}}$, as $T \to \infty$, while V_ρ represents that of $\tilde{\theta}$. Numbers reported in the last four columns are the empirical sizes (in percentage) of the t-tests computed at the nominal significance level of 5%. The t-statistics are constructed as $t_{\mathrm{PFD}} = \sqrt{nT}(\hat{\rho}_{\mathrm{PFD}} - \rho)/\sqrt{2(1+\hat{\rho}_{\mathrm{PFD}})}$, $t^c_{\mathrm{PFFD}} = \sqrt{nT}(\hat{\rho}^c_{\mathrm{PFFD}} - \rho)/\sqrt{2(1+\hat{\rho}^c_{\mathrm{PFFD}})}$, $t^h_{\mathrm{PFFD}} = \sqrt{nT}(\hat{\rho}^h_{\mathrm{PFFD}} - \rho)/\sqrt{2(1+\hat{\rho}^h_{\mathrm{PFFD}})}$, and $t_{\tilde{\theta}} = \sqrt{nT}(\tilde{\theta} - \theta)/\sqrt{\hat{V}_\rho}$, where $\hat{V}_\rho$ is obtained by replacing ρ with $\hat{\rho}_{\tilde{\theta}}$ in V_ρ.

臺灣大學學術期刊資料庫

References

Anderson, Thedorc Wilbur (1971), *The Statistical Analysis of Time Series*, NewYork: Wiley.

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

Dickey, David Alan and Wayne Arthur Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Han, Chirok and Peter Charles Bonest Phillips (2010), "GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity," *Econometric Theory*, 26, 119–151.

Marriott, F. H. C. and J. A. Pope (1954), "Bias in the Estimation of Autocorrelations," *Biometrika*, 41, 390–402.

Nickell, Stephen (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

Phillips, Peter Charles Bonest and Chirok Han (2008), "Gaussian Inference in AR(1) Time Series with or Without a Unit Root," *Econometric Theory*, 24, 631–650.

Phillips, Peter Charles Bonest and Tassos Magdalinos (2009), "Unit Root And Cointegrating Limit Theory when Initialization is in the Infinite Past," *Econometric Theory*, 25, 1682–1715.

Shaman, Paul and Robert A. Stine (1988), "The Bias of Autoregressive Coefficient Estimators," *Journal of American Statistical Association*, 83, 842–848.

van Giersbergen, Noud P. A. (2005), "On the Effect of Deterministic Terms on the Bias in Stable AR Models," *Economics Letters*, 89, 75–82.

投稿日期: 2011年7月25日, 接受日期: 2012年5月21日

臺灣大學學術期刊資料庫

有時間趨勢之一階自我迴歸模型的常態推論

陳致綱
政治大學國際經營與貿易學系
郭炳伸
政治大學國際經營與貿易學系

本文將 Phillips and Han (2008) 之一階差分估計式推展至有時間趨勢之一階自我迴歸模型的估計與推論上。藉由簡易的去除時間趨勢過程, 一階差分估計式仍具有漸近常態分配的性質; 同時, 據以建構之單根檢定相較於以雙重差分估計式爲基礎之單根檢定更有檢定力。本文提出的方法在固定效果動態追蹤資料模型下更具應用價值。

關鍵詞: 自我迴歸模型, 去除趨勢, 一階差分, 單根, 動態追蹤資料, 固定效果
JEL 分類代號: C13, C22, C23

臺灣大學學術期刊資料庫