A Note On Mean Squared Prediction Error Under The Unit Root Model With Deterministic Trend

Shu-Hui Yu, National University of Kaohsiung

shuhui@nuk.edu.tw

Joint Meeting of the 2011 Taipei International Statistical Symposium and 7th Conference of the Asian Regional Section of the IASC

12/19/2011

Outline

- Problem Description
- Literature Review
  - Least Squares estimators
  - Mean Squared Prediction Error (MSPE) in Stationary Model
- Main Results
  - MSPE under the random walk model
- Simulation Studies
- Conclusions

Problem Description

- Assume observations $y_1, \ldots, y_n$ are generated from AR(1) model with linear time trend,

$$y_t = \alpha + \delta(t-1) + \rho y_{t-1} + \varepsilon_t,$$

where $-1 < \rho \le 1,\ -\infty < \alpha, \delta < \infty$ are unknown coefficients and ε_t are unobservable random disturbance with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2$ for all *t*.

- Goal: Investigating how the value of ρ influences the mean squared prediction error (MSPE) $E(y_{n+1} - \hat{y}_{n+1})^2$ of $\hat{y}_{n+1}$ where $\hat{y}_{n+1} = \mathbf{x}_n' \hat{\theta}_n$ denotes the least squares predictor of y_{n+1}, $\mathbf{x}_t = (1, t, y_t)'$ and $\hat{\theta}_n$ satisfies $(\sum_{t=1}^{n-1} \mathbf{x}_t \mathbf{x}_t')\hat{\theta}_n = \sum_{t=1}^{n-1} \mathbf{x}_t y_{t+1}$.

Literature Review on LSE

- Testing whether $\rho = 1$ based on the least squares estimators of $\theta = (\alpha, \delta, \rho)'$ and σ^2, see Phillips and Perron (1988, Biometrika), Hamilton (1994), and Ng and Perron (2001, Econometrica).
- Case 1: No constant term or time trend in the regression; true process is a random walk.

$$y_t = \rho y_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d. with mean zero and variance σ^2.

$$T(\hat{\rho}_T - 1) \xrightarrow{L} \frac{(1/2)\{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2 dr},$$

$\hat{\rho}_T$: the OLS estimate.

Literature Review on LSE

- Case 2: Constant term but no time trend included in the regression; true process is a random walk,

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t .$$

$$\begin{bmatrix} T^{1/2}\hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int [W(r)]^2 dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ (1/2)\{[W(1)]^2 - 1\} \end{bmatrix}$$

where $(\hat{\alpha}_T, \hat{\rho}_T)'$: the OLS estimates. The second element states that

$$T(\hat{\rho}_T - 1) \xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)dr}{\int [W(r)]^2 dr - \left[\int W(r)dr\right]^2} .$$

Literature Review on LSE

- Case 3: Constant term but no time trend included in the regression; true process is random walk with drift,

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t,$$

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} N(\mathbf{0}, \sigma^2 Q^{-1}), \quad Q = \begin{bmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{bmatrix},$$

where $(\hat{\alpha}_T, \hat{\rho}_T)'$: the OLS estimates. The second element states that

$$T^{3/2}(\hat{\rho}_T - 1) \xrightarrow{L} Normal.$$

Literature Review on LSE

- Case 4: Constant term and time trend included in the regression; true process is random walk with or without drift,

$$
\begin{aligned}
y_t &= \alpha + \delta t + \rho y_{t-1} + \varepsilon_t \\
&= (1-\rho)\alpha + \rho[y_{t-1} - \alpha(t-1)] + (\delta + \rho\alpha)t + \varepsilon_t \\
&\equiv \alpha^* + \rho^* \xi_{t-1} + \delta^* t + \varepsilon_t,
\end{aligned}
$$

$$
\begin{bmatrix} T^{1/2}\hat{\alpha}_T^* \\ T(\hat{\rho}_T^* - 1) \\ T^{3/2}(\hat{\delta}_T^* - \alpha_0) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr & \frac{1}{2} \\ \int W(r)dr & \int [W(r)]^2 dr & \int rW(r)dr \\ \frac{1}{2} & \int rW(r)dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ (1/2)\{[W(1)]^2 - 1\} \\ W(1) - \int W(r)dr \end{bmatrix}
$$

where $(\hat{\alpha}_T^*, \hat{\rho}_T^*, \hat{\delta}_T^*)'$: the OLS estimates.

Literature Review on MSPE

- In the case of $|\rho|<1$, Corollary 1 of Ing (2003, ET) has shown that the second-order MSPE satisfies

$$\lim_{n\to\infty} n[E(y_{n+1}-\hat{y}_{n+1})^2-\sigma^2] = \lim_{n\to\infty} nE(y_{n+1}-\hat{y}_{n+1}-\varepsilon_{n+1})^2$$

$$= \lim_{n\to\infty} nE[\mathbf{x}_n'(\hat{\theta}_n-\theta)]^2 = 5\sigma^2,$$

which reflects the joint effect of $\mathbf{x}_n$ and $\hat{\theta}_n-\theta$ on MSPE.

- Note that Ing (2003) assume that observations come from

$$y_t = \sum_{i=0}^{q} \beta_i t^i + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t,$$

where q is a known nonnegative integer, β_i's are unknown real numbers.

Model Setting

- Assume observations $y_1, \ldots, y_n$ are generated from AR(1) model with linear time trend,

$$y_t = \alpha + \delta(t-1) + \rho y_{t-1} + \varepsilon_t,$$

where $-1 < \rho \le 1$, $-\infty < \alpha, \delta < \infty$ are unknown coefficients and ε_t are unobservable random disturbance with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2$ for all *t*.

Main Results

- What is the corresponding result in the case of $\rho = 1$?
- We show that when $\rho = 1$

$$\lim_{n\to\infty} n\{E(y_{n+1} - \hat{y}_{n+1})^2 - \sigma^2\} = \begin{cases} 6\sigma^2, & \delta = 0, \\ 9\sigma^2, & \delta \neq 0. \end{cases}$$

$$[\lim_{n\to\infty} n[E(y_{n+1} - \hat{y}_{n+1})^2 - \sigma^2] = 5\sigma^2,\ \text{Ing (2003)}]$$

- This result not only quantifies the excess MSPE due to the unit root as compared with Ing's result, it also indicates that in the presence of a unit root, the second-order MSPE can vary depending upon whether the linear time trend exists or not.

Assumptions

- (C1) $\{\varepsilon_t\}$ is a sequence of independent random variables with $E(\varepsilon_t)=0$ and $E(\varepsilon_t^2)=\sigma^2>0$ for all t.
- (C2) Denote the distribution function of

$$(m_2-m_1)^{-1/2}\sum_{i=m_1+1}^{m_2}\varepsilon_i$$

by $F_{m_2,m_1}(\cdot)$, where $m_2>m_1\geq 0$. There exist positive numbers ϕ, η and K and a positive integer ν such that

$$\sup_{m_2-m_1\geq\nu}|F_{m_2,m_1}(x)-F_{m_2,m_1}(y)|<K|x-y|^{\phi},$$

provided $|x-y|<\eta$.
- Some additional moment conditions on ε_t.
- Initial condition $y_0=0$.

Discussion on Assumption (C2)

- Smoothness condition seems to be indispensable.
- To see this, assume that $y_t = y_{t-1} + \varepsilon_t, \; y_0 = 0,$ and

$$\varepsilon_t = \begin{cases} 0, & \text{with prob. } 1/2, \\ 1, & \text{with prob. } 1/4, \\ -1, & \text{with prob. } 1/4, \end{cases}$$

then $y_t = \sum_{i=1}^{t} \varepsilon_i$ and

$$P(\tfrac{1}{n^2}\sum y_t^2 = 0) = P(\varepsilon_1 = 0, \ldots, \varepsilon_n = 0) = (1/2)^n.$$

Therefore, the normalized Fisher's information number does not converge since

$$P((\tfrac{1}{n^2}\sum y_t^2)^{-1} = \infty) > 0, \text{ for each } n.$$

Discussion on Assumption (C2)

- As argued in Remark 1 of Ing (2001, JTSA), (C2) is quite general and can accommodate many time series applications. In particular, when ε_t's are i.i.d., (C2) is fulfilled by most random disturbances have density functions.
- Several examples are shown in Ing (2001, JTSA).

Fisher Information Matrix

- To establish the previous result, we reliant on negative moment bounds for the minimum eigenvalues of the normalized Fisher information matrix,

$$E\{\lambda_{\min}^{-q}(\hat{S}_n)\} = O(1), \quad q > 1, \tag{1}$$

where

$$\hat{S}_n = \begin{cases} n^{-1}\sum_{j=1}^{n-1} G_n^* Q^* \mathbf{x}_j \mathbf{x}_j' Q^{*\prime} G_n^{*\prime}, & \text{if } \delta = 0, \\ n^{-1}\sum_{j=1}^{n-1} G_n Q \mathbf{x}_j \mathbf{x}_j' Q' G_n', & \text{if } \delta \neq 0, \end{cases}$$

$$G_n^* = diag(1, n^{-1}, n^{-1/2}\sigma^{-1}), \; G_n = diag(1, n^{-1}, n^{-2}),$$

$$Q^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\alpha & 1 \end{pmatrix},$$

Fisher Information Matrix

- Q is identical to Q^*, except that $-\alpha$ is replaced by $-\alpha+(\delta/2)$.
- Note that
  - In case $\delta=0$, $G_n^* Q^* \mathbf{x}_j = (1, j/n, n^{-1/2}\sigma^{-1}S_j)'$
  - In case $\delta\neq 0$, $G_n Q \mathbf{x}_j = (1, j/n, (\delta/2)(j^2/n^2)+S_j/n^2)'$
  - the first two components are non-random

Main Results

- Lemma 1. Assume (C1), (C2), $\rho = 1,\ \delta = 0$ and

$$\sup_{t \geq 1} E|\varepsilon_t|^r < \infty, \quad \text{for some } r > 2q.$$

Then (1) follows.

- Lemma 2. Assume (C1), (C2), $\rho = 1,\ \delta \neq 0$ and

$$\sup_{t \geq 1} E|\varepsilon_t|^r < \infty, \quad \text{for some } r > 4q.$$

Then (1) follows.

Technical Difficulties

- Lemma 1 of Ing, Sin and Yu (2010, ET), assuming observations come from ARIMA(p,1,0) without deterministic term

$$(1+\alpha_1 B+\cdots+\alpha_p B^p)(1-B)y_t=\varepsilon_t,$$

implies that (1) holds with $\hat{S}_n = n^{-1}\sum_{t=p+1}^{n-1}\mathbf{s}_{n,t}\mathbf{s}'_{n,t}$, where $\mathbf{s}_{n,t}=\tilde{G}_n\tilde{Q}\mathbf{y}_t$, with $\mathbf{y}_t=(y_t,\cdots,y_{t-p})'$, $\tilde{G}_n = diag(1,\cdots,1,n^{-1/2})$, and Q being the $(p+1)\times(p+1)$ matrix satisfying

$$\tilde{Q}\mathbf{y}_t=(y_t-y_{t-1},\cdots,y_{t-p+1}-y_{t-p},y_t)'.$$

Technical Difficulties

- This result is based on the property that the distribution of $\mathbf{a}'\mathbf{s}_{n,t}$ obey a certain Lipschitz condition for all direction $\mathbf{a}=(a_1,\cdots,a_{p+1})'\in R^{p+1}$ with $\|\mathbf{a}\|^2=\sum_{i=1}^{p+1}a_i^2=1$, which is entailed by a smoothness condition on ε_t similar to (C2).
- However, in this case, due to the non-randomness of the first two components of $G_n^*Q^*\mathbf{x}_j$ or $G_nQ\mathbf{x}_j$, for any unit vector $\mathbf{a}\in R^3$ with zero third component, the distribution of $\mathbf{a}'G_n^*Q^*\mathbf{x}_j$ cannot be of Lipschitz type, even if (C2) is imposed on ε_t. A substantial modification is needed.

Fisher Information Matrix

- Note that
  - In case $\delta = 0$, $G_n^* Q^* \mathbf{x}_j = (1, j/n, n^{-1/2}\sigma^{-1} S_j)'$
  - In case $\delta \neq 0$, $G_n Q \mathbf{x}_j = (1, j/n, (\delta/2)(j^2/n^2) + S_j/n^2)'$
  - the first two components are non-random

Expressions for the MSPE

- Theorem 1. Assume (C1), (C2), $\rho = 1,\ \delta = 0$ and $\sup_{t\geq 1} E|\varepsilon_t|^r < \infty$ for some $r > 8$. Then

$$\lim_{n\to\infty} nE\{(y_{n+1} - \hat{y}_{n+1})^2 - \sigma^2\} = E(\Lambda^2)\sigma^2,$$

where $\Lambda = \mathbf{v}_2'\Xi^{-1}\mathbf{v}_1$, with $\mathbf{v}_2 = (1,1,W(1))'$,

$$\mathbf{v}_1 = (W(1), W(1) - \int_0^1 W(r)dr, (1/2)(W^2(1) - 1))',$$

$$\Xi = \begin{pmatrix} 1 & 1/2 & \int_0^1 W(r)dr \\ 1/2 & 1/3 & \int_0^1 rW(r)dr \\ \int_0^1 W(r)dr & \int_0^1 rW(r)dr & \int_0^1 W^2(r)dr \end{pmatrix}.$$

Expressions for the MSPE

- Theorem 2. Assume (C1), (C2), $\rho = 1$, $\delta \neq 0$ and $\sup_{t \geq 1} E|\varepsilon_t|^r < \infty$ for some $r > 12$. Then

$$\lim_{n\to\infty} nE\{(y_{n+1} - \hat{y}_{n+1})^2 - \sigma^2\} = \mathbf{v}'_\delta \Xi_\delta^{-1} \mathbf{v}_\delta \sigma^2 = 9\sigma^2,$$

where $\mathbf{v}_\delta = (1,1,\delta/2)'$ and

$$\Xi_\delta = \begin{pmatrix} 1 & 1/2 & \delta/6 \\ 1/2 & 1/3 & \delta/8 \\ \delta/6 & \delta/8 & \delta^2/20 \end{pmatrix}.$$

- Normalized LSE

$$n^{1/2}(Q'G'_n)^{-1}(\hat{\theta}_n - \theta) \to N(\mathbf{0}, \Xi_\delta^{-1}\sigma^2)$$

which does vary with δ.

Growth Rate

- Theorem 3. Let $y_1,\dots,y_n$ be generated from model with $\alpha=\delta=0,\ \rho=1$ and ε_t be i.i.d. $N(0,\sigma^2)$ random variables. Then

$$E(\Lambda^2)=\underset{n\to\infty}{p\lim}\frac{\log\det(\sum_{t=1}^{n-1}\mathbf{x}_t\mathbf{x}_t')}{\log n}=6.$$

- Remark 1. Kim, Leybourne and Newbold (2004, JTSA) had obtained an approximate value of $\mathrm{var}(\Lambda)$, 6, through simulation.
- Remark 2. Comparing Theorem 1 with Theorem 3, we have

$$\lim_{n\to\infty}\frac{n\{E(y_{n+1}-\hat{y}_{n+1})^2-\sigma^2\}}{\sigma^2}=\underset{n\to\infty}{p\lim}\frac{\log\det(\sum_{t=1}^{n-1}\mathbf{x}_t\mathbf{x}_t')}{\log n}.$$

Growth Rate

- Remark 3. Remark 2 is still valid for $\delta \neq 0$. Moreover, when $\delta \neq 0$,

$$\frac{\log \det(\sum_{t=1}^{n-1} \mathbf{x}_t \mathbf{x}_t')}{\log n} = 9 + \frac{\log \det(\Xi_\delta)}{\log n} + o_p(1/\log n),$$

which shows that the impact of δ on the order of growth of $\det(\sum_{t=1}^{n-1} \mathbf{x}_t \mathbf{x}_t')$ is asymptotically negligible.

Simulation Studies

- Empirical estimates of $n\{E(y_{n+1}-\hat{y}_{n+1})^2-\sigma^2\}$ with $\alpha\in\{0,1\}$, $\delta\in\{0,1\}$, $\rho\in\{0,0.8,1\}$ and ε_t's being i.i.d. $N(0,1)$. Replication number =10,000 .

Simulation Studies

	n	$\alpha = \delta = 0$	$\alpha = 1, \delta = 0$	$\alpha = 0, \delta = 1$	$\alpha = \delta = 1$
$\rho = 0$	100	4.89	5.09	4.96	4.98
	500	5.08	5.17	4.97	5.13
	1000	4.96	5.01	5.01	5.03
	∞	5	5	5	5
$\rho = 0.8$	100	5.30	5.18	4.65	4.75
	500	5.12	5.13	4.85	4.79
	1000	5.10	5.08	4.95	5.02
	∞	5	5	5	5
$\rho = 1$	100	6.01	5.86	9.20	9.24
	500	5.95	6.01	9.10	9.15
	1000	5.98	5.97	9.07	9.05
	∞	6	6	9	9

Conclusions

- For stationary time series, the MSPE of the least squared predictor is usually proportional to the number of estimated parameters. This, however, is not true for most nonstationary time series.
- This paper takes the first step towards developing a systematic view of the MSPE under nonstationarity.
- In particular, by establishing moment bunds for the inverse of the minimum eigenvalues of the normalized Fisher information matrix, we show that the MSPE can be linked to the growth rate of the Fisher information.