

Department of Economics, Hitotsubashi University

Discussion Paper Series #2014-19

**Asymptotic Distribution of the Conditional Sum of Squares Estimator
Under Moderate Deviation From a Unit Root in MA(1)**

Ryota Yabe

December 2014

Asymptotic Distribution of the Conditional Sum of Squares Estimator Under Moderate Deviation From a Unit Root in MA(1)

Ryota Yabe*

Abstract

This paper considers the conditional sum of squares estimator (CSSE) for the moderate deviation MA(1) process that has the parameter of the MA(1) with the distance between the parameter and unity being larger than $O(T^{-1})$. We show that the asymptotic distribution of the CSSE is normal, even though the process belongs to the local-to-unity class. The convergence rate continuously changes from an invertible order to a noninvertible one. In this sense, the moderate deviation process in MA(1) has a continuous bridge property like the AR process.

Keywords. Moving average; Noninvertible moving average; Unit root; local to unity; Moderate Deviations; Conditional sum of squares estimation;

1 Introduction

A moderate deviation process is the local-to-unity process that has a larger distance from unity than the near unit root one. An asymptotic property of the moderate deviation in the AR(1) process has been variously explored to investigate the asymptotic properties in the local-to-unity class. In a stationary region, Giraitis and Phillips (2006) and Phillips and Magdalinos (2007) showed that an ordinary least squares estimator uniformly has an asymptotically normal distribution. Moreover, its convergence rate continuously changes from a stationary process to a unit root one. These studies showed that the moderate deviation process and the stationary one share some similar asymptotic properties.

A (near) unit root MA process that belongs to the local-to-unity-class is deeply connected with the test of the null hypothesis of stationarity and has been studied several researchers. See Kwiatkowski et al. (1992), Saikkonen and

*E-mail address:k131014a@r.hit-u.ac.jp

Luukkonen (1993b) and Saikkonen and Luukkonen (1993a) for the typical studies of the null hypothesis stationarity test. In this paper, we consider the moderate deviation MA(1) process that appears in empirical studies. It is well known that the ARIMA-based model for macroeconomic data often has the largest MA root close to one, so that the usual asymptotic theory does not work well in finite samples. The MA process with the root close to one has been studied by modeling the MA coefficient as a moderate deviation. For example, Pantula (1991) and Nabeya and Perron (1994) have theoretically shown that the unit root AR process with the moderate deviation MA(1) error term does not follow the conventional unit root asymptotic theory. Ng and Perron (2001) has proposed the AR unit root test by correcting information criteria for the special case of moderate deviation.

There are a lot of studies of estimation theory for a noninvertible process. However, derivation of the asymptotic distribution of the maximum likelihood estimator (MLE) has been unsolved. This difficulty comes from the asymptotic properties of the MLE. The MLE of the unit root MA(1) process asymptotically has the probability mass at unity with positive probability and a so-called pile-up effect. Because of the pile-up effect that was found by Sargan and Bhargava (1983) and Anderson and Takemura (1986), the usual limit theory such as the central limit theorem is inapplicable, which was also observed by Davis and Dunsmuir (1996) in a near unit root process. There has never been a study that has succeeded in deriving the asymptotic distribution of the MLE in the local-to-unity class. However, the result of Yabe (2012) has conjectured that an estimator such as the MLE in a moderate deviation process is asymptotically normal like an invertible one. This study gives the asymptotic distribution of the CSSE for the moderate deviation MA(1) process.

The CSSE is known to coincide with the MLE in the case where the initial value of the disturbance is zero. See Harvey (1993) for the details of the CSSE in an MA(1) process. We concentrate on the zero initial value case, because the asymptotic properties are different for the nonzero initial value one. We show that the CSSE satisfies consistency and its convergence order continuously changes from an invertible process to a noninvertible one like the AR(1) moderate deviation process. The asymptotic distribution of the CSSE is normal even though the process belongs to the local-to-unity class and is the same as the one in the invertible process.

The rest of this paper is organized as follows: In Section 2, we introduce the model and assumptions. In Section 3, we give the definition of the CSSE and derive its asymptotic distribution. Section 4 provides some concluding remarks. All proofs of the theoretical results are presented in the Appendix.

Throughout this chapter, we use the following notations: \rightarrow_p and \Rightarrow to denote convergence in probability and weak convergence, respectively, as the sample size T goes to infinity.

2 Model, Assumptions and CSSE

We assume that the MA(1) data generating process $\{y_t\}$ is given by

$$y_t = \epsilon_t - \rho_T \epsilon_{t-1}, \quad (t = 1, \dots, T). \quad (1)$$

We impose the following regularity conditions on $\{\epsilon_t\}$.

Assumption 1. $\{\epsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean 0 and variance σ^2 . We also assume that $\epsilon_0 = 0$ and $y_t = 0$ almost sure (a.s.) for $t \leq 0$.

Although we adopt the i.i.d. assumption for simplicity, the assumption can be extended to the martingale difference. The initial value condition for ϵ_t is crucial, because the nonzero initial value causes different asymptotic properties in the (near) unit root MA(1) model. See Tanaka (1996) for the details of the effect of the initial-value assumption on the MLE and the test statistic. Moreover, as noted by Harvey (1993), the CSSE does not coincide with the MLE in the nonzero-initial-value case. Though the analysis of different initial-value assumptions is interesting, further study in this direction is beyond the scope of our work.

The moderate deviation MA root can be written as $\rho_T = 1 - c/k_T$, where k_T is a deterministic sequence to go to infinity and $c > 0$, and is assumed to satisfy the following condition.

Assumption 2. The coefficient ρ_T depending on T satisfies $\rho_T \rightarrow 1$, $T(1 - \rho_T) = cT/k_T \rightarrow \infty$ and $\rho_T \in (-1, 1)$.

From the definition of the coefficient, note that $\{y_t\}$ is invertible in a finite sample and belongs to the local-to-unity class. Though $\{y_t\}$ is a triangular array $\{y_{t,T}\}$, we omit the index T to simplify the notation.

We introduce the CSSE that is defined as the solution of the following minimization problem:

$$\hat{\rho} = \arg \min_{\rho \in [-1, 1]} Q_T(\rho), \quad (2)$$

where

$$Q_T(\rho) = \frac{1}{T} \sum_{t=1}^T \left(\sum_{l=0}^{t-1} \rho^l y_{t-l} \right)^2.$$

$Q_T(\rho)$ is equivalent to the sum of squares of the disturbance terms at $\rho = \rho_T$. The solution of this minimizing problem is restricted in the region $[-1, 1]$ to satisfy identification. Note that in the zero initial value case, the objective function $Q_T(\rho)$ is a monotonic transformation of the concentrated likelihood function, and so the CSSE coincides with the MLE

even in finite samples. However, in the nonzero-initial-value case, the CSSE is not equal to the MLE. See Harvey (1993) and Box et al. (2013) for the details about properties of CSSE in an invertible process.

3 Main Result

In this section we derive the asymptotic distribution of the CSSE. First, we show consistency of the CSSE in the following theorem.

Theorem 1. *Under Assumptions 1 and 2, the CSSE $\hat{\rho}$ is a consistent estimator.*

This theorem shows that the CSSE satisfies consistency for invertible and noninvertible processes. Since ρ_T converges to unity, the CSSE converges to 1 in probability. The minimizer defined in (2) satisfies the following first-order condition:

$$0 = \frac{\partial Q_T(\hat{\rho})}{\partial \rho}. \quad (3)$$

Applying the mean value theorem to (3) yields

$$\hat{\rho} - \rho_T = \left(\frac{d^2}{d\rho_T^2} Q_T(\rho^+) \right)^{-1} \left(\frac{d}{d\rho_T} Q_T(\rho_T) \right), \quad (4)$$

where ρ^+ exists between ρ_T and $\hat{\rho}$,

$$\frac{d}{d\rho} Q_T(\rho) = \frac{2}{T} \sum_{t=1}^T \left(\sum_{l=0}^{t-1} \rho^l y_{t-l} \right) \left(\sum_{l=0}^{t-1} l \rho^{l-1} y_{t-l} \right) \quad (5)$$

and

$$\frac{d^2}{d\rho^2} Q_T(\rho) = \frac{2}{T} \sum_{t=1}^T \left[\left(\sum_{l=0}^{t-1} l \rho^{l-1} y_{t-l} \right)^2 + \left(\sum_{l=0}^{t-1} \rho^l y_{t-l} \right) \left(\sum_{l=0}^{t-1} l(l-1) \rho^{l-2} y_{t-l} \right) \right]. \quad (6)$$

Note that $\hat{\rho}$ is consistent, so that ρ^+ also converges to ρ_T in probability. The following lemmas give the asymptotic results of the first and second derivatives of the objective functions given in (5) and (6), respectively.

Lemma 1. *Under Assumptions 1 and 2, the asymptotic distribution of the normalized first derivative of the objective function $Q_T(\rho)$ given in (5) at the true value ρ_T is given by:*

$$\sqrt{T(1 - \rho_T^2)} \frac{d}{d\rho_T} Q_T(\rho) \Big|_{\rho=\rho_T} \Rightarrow N(0, 4\sigma^4).$$

Lemma 2. *Under Assumptions 1 and 2, the normalized second derivative of the objective function $Q_T(\rho)$ given in (6) at the true value ρ_T converges to a constant in probability. That is,*

$$(1 - \rho_T^2) \frac{d^2}{d\rho_T^2} Q_T(\rho) \Big|_{\rho=\rho_T} \rightarrow_p 2\sigma^2.$$

The following corollary is immediately given by Lemma 2 and the continuous mapping theorem.

Corollary 1. *Lemma 2 holds even if the second derivative of $Q_T(\rho)$ is evaluated at ρ^+ , where ρ^+ is a convex combination of $\hat{\rho}$ and ρ_T .*

Theorem 2. *Under Assumptions 1 and 2, the normalized CSSE by $\sqrt{T/(1 - \rho_T^2)}$ converges in distribution to a standard normal random variable. That is,*

$$\sqrt{\frac{T}{1 - \rho_T^2}} (\hat{\rho} - \rho_T) \Rightarrow N(0, 1).$$

This theorem shows that the convergence order of the distance between the CSSE and the true value under moderate deviation, $\hat{\rho} - \rho_T = O_p((Tk_T)^{-1/2})$, continuously changes from the order of the invertible process, $O_p(T^{-1/2})$, to that of the (near) noninvertible one, $O_p(T^{-1})$. We can clarify that the moderate deviation process in MA(1) is the continuous bridge between invertible and noninvertible processes. Since Theorem 2 also holds in the invertible process, we can conclude that the moderate deviation and invertible processes share some asymptotic properties.

4 Concluding Remarks

In this paper, we give the asymptotic distribution of the CSSE in the MA(1) model under moderate deviation. The fact that the convergence rate continuously changes from the invertible process to the unit root one means that the continuous bridge property of moderate deviation can be observed in the MA(1) process like the AR process.

Our work is useful for developing an AR unit root test for the ARIMA model when the largest root of MA is close to unity. Although Ng and Perron (2001) has dealt with the special case where $k_T = \sqrt{T}$, our results should be applicable for more general cases by using a generalized least squares (GLS) technique. Such an analysis will be presented in future work.

Acknowledgments

The author is grateful to Professors Katsuto Tanaka, Eiji Kurozumi, Katsumi Shimotsu and Toshio Honda for valuable comments. The financial support from the Grant-in-Aid for the Japan Society for the Promotion of Science (JSPS) Fellows is gratefully acknowledged.

References

- Anderson, T. W. and Takemura, A. (1986). Why Do Noninvertible Estimated Moving Averages Occur? *Journal of Time Series Analysis*, 7(4):235–254.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2013). *Time Series Analysis*. Forecasting and Control. John Wiley & Sons.
- Davis, R. A. and Dunsmuir, W. T. M. (1996). Maximum Likelihood Estimation for MA(1) Processes with a Root on or near the Unit Circle. *Econometric Theory*, 12(01):1–29.
- Giraitis, L. and Phillips, P. (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis*, 27(1):51–60.
- Harvey, A. C. (1993). *Time Series Models*. Hemel Hempstead: Harvester Wheatsheaf.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3):159–178.
- Nabeya, S. and Perron, P. (1994). Local asymptotic distribution related to the AR(1) model with dependent errors. *Journal of Econometrics*, 62(2):229–264.
- Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554.
- Pantula, S. G. (1991). Asymptotic Distributions of Unit-Root Tests When the Process Is Nearly Stationary. *Journal of Business & Economic Statistics*, 9(1):63–71.
- Phillips, P. C. B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1):115–130.

- Saikkonen, P. and Luukkonen, R. (1993a). Point Optimal Tests for Testing the Order of Differencing in Arima Models. *Econometric Theory*, 9(3):343–362.
- Saikkonen, P. and Luukkonen, R. (1993b). Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models. *Journal of the American Statistical Association*, 88(422):596–601.
- Sargan, J. D. and Bhargava, A. (1983). Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root Lies on the Unit Circle. *Econometrica*, 51(3):799.
- Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. Wiley.
- Yabe, R. (2012). Limiting distribution of the score statistic under moderate deviation from a unit root in MA (1). *Journal of Time Series Analysis*, 33(4):533–541.

Appendix

For derivation of the asymptotic properties, we prove the following preliminary Lemma 1 on the asymptotic order of the coefficient power series.

Lemma A 1. *The following relations hold under Assumption 2 as $T \rightarrow \infty$.*

(a) For all $a > 0$, $b > 0$,

$$T^a \rho_T^{bT} = o(1).$$

(b)

$$\frac{1}{Tk_T^3} \sum_{t=1}^T \sum_{l=0}^{t-1} 2l(2l-1) \rho_T^{2l-2} = \frac{1}{c^3} + o(1).$$

Proof of Lemma A1

(a) Taking logarithms and using the Taylor expansion for sufficiently large T , we can show that:

$$a \log T + bT \log \rho_T = a \log T + bT \log(1 - c/k_T) = a \log T - bcT/k_T + o(Tk_T^{-1}) \rightarrow -\infty. \blacksquare$$

(b)

$$\begin{aligned}
\frac{1}{Tk_T^3} \sum_{t=1}^T \sum_{l=0}^{t-1} 2l(2l-1)\rho_T^{2l-2} &= \frac{1}{Tk_T^3} \frac{d^2}{d\rho_T^2} \sum_{t=1}^T \sum_{l=0}^{t-1} \rho_T^{2l} \\
&= \frac{1}{Tk_T^3} \left(\frac{2T}{(1-\rho_T^2)^2} + \frac{8T\rho_T^2}{(1-\rho_T^2)^3} - \frac{2}{(1-\rho_T^2)^2} - \frac{8\rho_T^2}{(1-\rho_T^2)^3} - \frac{12\rho_T^2}{(1-\rho_T^2)^3} - \frac{24\rho_T^4}{(1-\rho_T^2)^4} \right) + o(1) \\
&= 1/c^3 + o(1),
\end{aligned}$$

where the second last equality holds by (a) of Lemma A1. ■.

Proof of Theorem 1

The definition of the CSSE in equation (2) implies the following inequality:

$$P(Q_T(\hat{\rho}) \leq Q_T(\rho_T)) = 1.$$

The left-hand side is also bounded above by:

$$P(Q_T(\hat{\rho}) \leq Q_T(\rho_T) \text{ and } \hat{\rho} \in \Theta_T) + P(Q_T(\hat{\rho}) \leq Q_T(\rho_T) \text{ and } \hat{\rho} \notin \Theta_T) \leq P(\hat{\rho} \in \Theta_T) + P(Q_T(\hat{\rho}) \leq Q_T(\rho_T) \text{ and } \hat{\rho} \notin \Theta_T)$$

for all open interval Θ_T in $[-1, 1]$. Without loss of generality, we assume that $\Theta_T = \{\rho \mid |\rho - \rho_T| < c^*, c^* > 0\} \cap [-1, 1]$, where c^* is an arbitrary positive constant. Consistency of $\hat{\rho}$ is implied by proving $P(Q_T(\hat{\rho}) \leq Q_T(\rho_T) \text{ and } \hat{\rho} \notin \Theta_T) \rightarrow 0$.

Since the weak law of large numbers implies that:

$$Q_T(\rho_T) = \frac{1}{T} \sum_{t=1}^T \left(\sum_{l=0}^{t-1} \rho_T^l y_{t-l} \right)^2 = \frac{1}{T} \sum_{t=1}^T \epsilon_t^2 \rightarrow_p \sigma^2,$$

it is sufficient to show that

$$P\left(\min_{\rho \notin \Theta_T} \frac{1}{\sigma^2} Q_T(\rho) > 1\right) \rightarrow 1. \tag{A.7}$$

Without loss of generality, we assume that $\sigma^2 = 1$. For all $\rho \notin \Theta_T$, we can obtain

$$\sum_{l=0}^{t-1} \rho^l y_{t-l} = \epsilon_t + \sum_{l=0}^{t-2} \rho^l (\rho - \rho_T)^{l+1} \epsilon_{t-1-l}.$$

Consider the case where $\rho < 0$. Since $(\rho - \rho_T) < -1$ for sufficiently large T , $Q_T(\rho)$ diverges with a exponential order.

This shows (A.7). We consider the case where $\rho \geq 0$. Then, the objective function $Q(\rho)$ can be written as:

$$Q_T(\rho) = 1 + \frac{2}{T} \sum_{t=2}^T \epsilon_t \sum_{l=0}^{t-2} \rho^l (\rho - \rho_T)^{l+1} \epsilon_{t-1-l} + \frac{1}{T} \sum_{t=2}^T \left(\sum_{l=0}^{t-2} \rho^l (\rho - \rho_T)^{l+1} \epsilon_{t-1-l} \right)^2 + o_p(1). \quad (\text{A.8})$$

Now, we show that the second term is asymptotically negligible and the third term is bounded below by a positive constant in probability. First, we consider the second term. By applying Chebyshev's inequality, if we show that the expectation of the squared second term converges to 0, the second term converges in probability to 0. The expectation of the squared second term is given by:

$$\begin{aligned} \frac{1}{T^2} \sum_{s=2}^T \sum_{t=2}^T \sum_{j=0}^{s-2} \sum_{l=0}^{t-2} \rho^{l+j} (\rho - \rho_T)^{l+j+2} \mathbb{E}(\epsilon_s \epsilon_t \epsilon_{t-1-l} \epsilon_{s-1-j}) &= \frac{(\rho - \rho_T)^2}{T^2} \sum_{t=2}^T \sum_{l=0}^{t-2} \rho^{2l} (\rho - \rho_T)^{2l} \\ &= \frac{(\rho - \rho_T)^2}{T^2} \sum_{t=2}^T \frac{1 - \rho^{2(t-1)} (\rho - \rho_T)^{2(t-1)}}{1 - \rho^2 (\rho - \rho_T)^2} \\ &< \frac{1}{T} \frac{1}{1 - \rho^2 (\rho - \rho_T)^2} < \frac{1}{T} \frac{1}{1 - \rho^2} \leq \frac{1}{T} \frac{1}{1 - c^{*2}} = O(T^{-1}). \end{aligned}$$

Thus, the second term asymptotically converges in probability to 0. We can also show that the third term is bounded below by c^{*2} in the limit such that:

$$\begin{aligned} \frac{1}{T} \sum_{t=2}^T \left(\sum_{l=0}^{t-2} \rho^l (\rho - \rho_T)^{l+1} \epsilon_{t-1-l} \right)^2 &= \frac{1}{T} \sum_{t=2}^T \left((\rho - \rho_T)^2 \epsilon_{t-1}^2 + 2(\rho - \rho_T) \epsilon_{t-1} \sum_{l=1}^{t-2} \rho^l (\rho - \rho_T)^{l+1} \epsilon_{t-1-l} \right. \\ &\quad \left. + \left(\sum_{l=1}^{t-2} \rho^l (\rho - \rho_T)^{l+1} \epsilon_{t-1-l} \right)^2 \right) \\ &\geq \frac{1}{T} \sum_{t=2}^T \left((\rho - \rho_T)^2 \epsilon_{t-1}^2 + 2(\rho - \rho_T) \epsilon_{t-1} \sum_{l=1}^{t-2} \rho^l (\rho - \rho_T)^{l+1} \epsilon_{t-1-l} \right) \\ &\geq \frac{c^{*2}}{T} \sum_{t=2}^T \epsilon_{t-1}^2 + o_p(1) \rightarrow_p c^{*2} > 0. \end{aligned}$$

Therefore, we obtain the required result. ■

Proof of Lemma 1 $\sum_{l=0}^{t-1} \rho_T^l y_{t-l}$ and $\sum_{l=0}^{t-1} l \rho_T^{l-1} y_{t-l}$ can be written such as:

$$\sum_{l=0}^{t-1} \rho_T^l y_{t-l} = (\epsilon_t - \rho_T \epsilon_{t-1}) + (\rho_T \epsilon_{t-1} - \rho_T^2 \epsilon_{t-2}) + \cdots + \rho_T^{t-1} \epsilon_1 = \epsilon_t \quad (\text{A.9})$$

and

$$\begin{aligned}
\sum_{l=0}^{t-1} l \rho_T^{l-1} y_{t-l} &= y_{t-1} + 2\rho_T y_{t-2} + \cdots + (t-1)\rho_T^{t-2} y_1 \\
&= (\epsilon_{t-1} - \rho_T \epsilon_{t-2}) + 2\rho_T(\epsilon_{t-2} - \rho_T \epsilon_{t-3}) + \cdots + (t-1)\rho_T^{t-2} \epsilon_1 \\
&= \epsilon_{t-1} + \rho_T \epsilon_{t-2} + \cdots + \rho_T^{t-2} \epsilon_1 \\
&= x_{t-1} = \rho_T x_{t-2} + \epsilon_{t-1}.
\end{aligned} \tag{A.10}$$

From (A.10), note that $\{x_t\}$ follows the moderate deviation AR(1) process. By substituting (A.9) and (A.10) into (5), the first derivative of the CSS objective function can be represented as

$$\frac{\partial Q_T(\rho_T)}{\partial \rho_T} = 2 \sum_{t=1}^T \epsilon_t x_{t-1}. \tag{A.11}$$

By applying (b) in Phillips and Magdalinos (2007) or Lemma 1 in Giraitis and Phillips (2006) to (A.11), we obtain the required result. ■

Proof of Lemma 2 The first term of $(1 - \rho_T^2)d^2/d\rho^2 Q_T(\rho)$ can be expressed as

$$(1 - \rho_T^2) \frac{2}{T} \sum_{t=2}^T x_{t-1}^2, \tag{A.12}$$

Applying Lemma 2 of Giraitis and Phillips (2006) yields that the right-hand side of (A.12) converges in probability to $2\sigma^2$. Thus, it is enough to show that the second term of $(1 - \rho_T^2)d^2/d\rho^2 Q_T(\rho)$ is asymptotically negligible. Since

$$\begin{aligned}
\sum_{l=0}^{t-1} l(l-1)\rho_T^{l-2} y_{t-l} &= 2y_{t-2} + 6\rho_T y_{t-3} + 12\rho_T^2 y_{t-4} + \cdots + (t-1)(t-2)\rho_T^{t-3} y_1 \\
&= 2[\epsilon_{t-2} + 2\rho_T \epsilon_{t-3} + \cdots + (t-2)\epsilon_1] \\
&= 2 \sum_{l=2}^{l=t-1} (l-1)\rho_T^{l-2} \epsilon_{t-l},
\end{aligned}$$

the second term is expressed by the following equation:

$$\frac{2(1 - \rho_T^2)}{T} \sum_{t=1}^T \left(\sum_{l=0}^{t-1} \rho_T^l y_{t-l} \right) \left(\sum_{l=0}^{t-1} l(l-1)\rho_T^{l-2} y_{t-l} \right) = \frac{4(1 - \rho_T^2)}{T} \sum_{t=1}^T \epsilon_t \sum_{l=2}^{t-1} (l-1)\rho_T^{l-2} \epsilon_{t-l}. \tag{A.13}$$

To prove this Lemma, it is sufficient to show that the variance of (A.13) converges to 0 like the proof of Theorem 1 by Chebyshev's inequality. The expectation of the squares of the second term is given by:

$$\frac{16(1 - \rho_T^2)^2}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{l=2}^{t-1} \sum_{k=2}^{t-1} (l-1)(k-1) \rho_T^{l+k-4} \mathbb{E}(\epsilon_t \epsilon_s \epsilon_{t-l} \epsilon_{s-k}) = \frac{16(1 - \rho_T^2)^2 \sigma^4}{T^2} \sum_{t=1}^T \sum_{l=2}^{t-1} (l-1)^2 \rho_T^{2(l-2)}.$$

Thus, the required result can be obtained by applying Lemma A1 (b) to the above equation. ■

Proof of Theorem 2

Applying Lemmas 1 and 2, corollary 1 and the continuous mapping theorem to (4) immediately yields Theorem 2.

■