Département de sciences économiques

9817

# The FCLT with Dependent Errors: an Helicopter Tour of the Quality of the Approximation

*PERRON, Pierre*
*MALLET, Sylvie*

Université de Montréal

**Département de sciences économiques**
Université de Montréal
Faculté des arts et des sciences
C.P. 6128, succursale Centre-Ville
Montréal (Québec) H3C 3J7
Canada
http://www.sceco.umontreal.ca
SCECO-information@UMontreal.CA
Téléphone : (514) 343-6539
Télécopieur : (514) 343-7221

Ce cahier a également été publié par le Centre interuniversitaire de recherche en économie quantitative (CIREQ) sous le numéro 3497.

*This working paper was also published by the Center for Interuniversity Research in Quantitative Economics (CIREQ), under number 3497.*

ISSN 0709-9231

**CAHIER 9817**

# THE FCLT WITH DEPENDENT ERRORS : AN HELICOPTER TOUR OF THE QUALITY OF THE APPROXIMATION

Pierre PERRON[1] and Sylvie MALLET[2]

[1] Centre de recherche et développement en économique (C.R.D.E.) and Département de sciences économiques, Université de Montréal, and Department of Economics, Boston University

[2] Transports Canada

December 1997

---

This paper is drawn from the second author's M.Sc. essay at the Université de Montréal [Mallet (1995)]. The first author acknowledges financial support from the Social Sciences and Humanities Research Council (SSHRC) of Canada, the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Fonds pour la Formation de chercheurs et l'aide à la recherche (FCAR) of Québec.

## RÉSUMÉ

Cette note examine la pertinence de l'approximation en échantillon fini donnée par le théorème de la limite centrale fonctionnelle quand les erreurs sont dépendantes. Nous comparons la distribution des sommes partielles échelonnées de quelques donnés avec la distribution du processus de Wiener sur lequel elle converge. Notre modèle est intentionnellement très simple, compte tenu qu'il considère les donnés générées à partir d'un processus ARMA(1,1). Cependant, c'est suffisant pour obtenir des conclusions intéressantes sur les éléments particuliers qui causent des approximations inadéquates même en larges échantillons.



Mots clés : processus de Wiener, pont brownien, fonction de répartition, corrélation temporelle, approximation asymptotique



## ABSTRACT

This note investigates the adequacy of the finite-sample approximation provided by the Functional Central Limit Theorem (FCLT) when the errors are allowed to be dependent. We compare the distribution of the scaled partial sums of some data with the distribution of the Wiener process to which it converges. Our setup is purposely very simple in that it considers data generated from an ARMA(1,1) process. Yet, this is sufficient to bring out interesting conclusions about the particular elements which cause the approximations to be inadequate in even quite large sample sizes.



Key words : Wiener process, Brownian bridge, distribution function, serial correlation, asymptotic approximation

## 1 Introduction

The Central Limit Theorem (CLT) allowing data to be heterogenous and/or dependent and its generalization to the Functional Central Limit Theorem (FCLT) are by now standard tools in econometrics. The FCLT states that, under suitable conditions on the data generating process of some series $x_t$, we have the following convergence results as the sample size, $T$, increases:

$$X_T(r) \equiv \sigma^{-1}T^{-1/2}\sum_{t=1}^{[Tr]} x_t \Rightarrow W(r) \tag{1}$$

for $r \in [0,1]$, where $W(r)$ is the standard Wiener process, $\Rightarrow$ denotes weak convergence in distribution (usually from the space $D[0,1]$ to $C[0,1]$ under the sup or the Skorohod metrics) and $\sigma^2 = \lim_{T\to\infty} T^{-1}E(\sum_{t=1}^{T} x_t)^2$ which in the case of a stationary process is equivalent to $(2\pi)$ times the spectral density function at frequency zero of the series $x_t$. There are a wide variety of sufficient conditions available in the literature which ensure that the result (1) holds. For example, a class of linear processes with martingale difference innovations (Phillips and Solo, 1992), mixing conditions as used in Phillips and Perron (1988) (Herrndorf (1984)), and so on. For a comprehensive review, see Davidson (1994).

This asymptotic result has been widely used in time series analysis, in particular in relation to the asymptotic distribution of estimators and test statistics related to integrated (autoregressive unit root) and cointegrated processes. Yet, despite its popularity, little has been documented on how well the distribution of the Wiener process $W(r)$ approximates the finite sample distribution of $X_T(r)$ in finite samples, especially in the context where serial correlation is present. In this note, we try to partially fill this gap using simple $ARMA(1,1)$ processes for the data $x_t$. The framework is kept very simple to better highlight the importance of the nature of the serial correlation for the adequacy (or lack of it) of the asymptotic approximation. We furthermore assume that the innovations of the $ARMA$ processes are *i.i.d.* $N(0,1)$. In that case, both $X_T(r)$ and $W(r)$ are normally distributed with mean 0 and to the extent that their distributions differ they do so only via a different variance. Hence, a comparison of the relative variances of $X_T(r)$ and $W(r)$ is enough to assess the extent to which the approximation is adequate or not.

The rest of this note is structured as follows. Section 2 discusses the experimental design and states some required results concerning the distributions of $X_T(r)$ and $W(r)$ in relation to the data-generating processes used. Section 3 discusses the special case with *i.i.d.* data. Section 4 presents the results in the general case. Section 5 concludes.

## 2 The Experimental Design

What is of interest here is to highlight the nature of the serial correlation which can cause the approximation to be adequate or not in finite samples. To that extent, a useful starting point is to consider a simple $ARMA(1,1)$ process for the $x_t$'s, i.e.

$$x_t = \alpha x_{t-1} + e_t + \theta e_{t-1}, \tag{2}$$

where the sequence of innovations $e_t$ is assumed to be $i.i.d.$ $N(0,1)$ and $x_0 = 0$. To ensure that the process satisfies the conditions such that (1) holds, we impose $|\alpha| < 1$ and $|\theta| < 1$. In this case, $X_T(r)$ is normally distributed with mean 0 as is $W(r)$. Hence, the only difference between the distribution of $X_T(r)$ and $W(r)$ relates to the variance of each variable. As is well known, $Var(W(r)) = r$. Using the fact that

$$\sigma^2 = \lim_{T\to\infty} T^{-1}E[\sum_{t=1}^{T} x_t]^2 = \frac{(1+\theta)^2}{(1-\alpha)^2},$$

tedious but straightforward algebra shows that

$$\begin{aligned} Var(X_T(r)) \quad &= \quad \frac{(1-\alpha)^2}{(1+\theta)^2}\left\{\frac{[Tr]}{T} + 2\frac{(\theta+\alpha)}{(1-\alpha)}[\frac{([Tr]-1)}{T} - \frac{\alpha(1-\alpha^{([Tr]-1)})}{T(1-\alpha)}] \right. \\ & \left. + \frac{(\theta+\alpha)^2}{(1-\alpha)^2}[\frac{([Tr]-1)}{T} - 2\frac{\alpha(1-\alpha^{([Tr]-1)})}{T(1-\alpha)} + \frac{\alpha^2(1-\alpha^{2([Tr]-1)})}{T(1-\alpha^2)}]\right\}. \end{aligned}$$

As special cases, we have that with an $MA(1)$ process:

$$Var(X_T(r)) = \frac{([Tr]-1)}{T} + \frac{1}{T(1+\theta)^2},$$

and with an $AR(1)$ process:

$$Var(X_T(r)) = \frac{[Tr]}{T} - \frac{2\alpha(1-\alpha^{[Tr]})}{T(1-\alpha)} + \frac{\alpha^2(1-\alpha^{2[Tr]})}{T(1-\alpha^2)}.$$

It is easily seen, from each expression, that as $T \to \infty$, $Var(X_T(r)) \to r = Var(W(r))$.

To assess, the "approximation error" we use the percentage relative deviation of $Var(X_T(r))$ and $Var(W(r))$, that is

$$e(T,r,\alpha,\theta) = 100\frac{(Var(X_T(r)) - Var(W(r)))}{Var(W(r))} = 100\frac{(Var(X_T(r)) - r)}{r}. \tag{3}$$

As highlighted in the argument of the error function $e(\cdot)$, this error depends on the parameters $T$, $r$, $\alpha$ and $\theta$.

## 3 The Case with an *i.i.d.* Process

When the process is *i.i.d.* $N(0,1)$, we have

$$e(T,r,0,0) = 100(\frac{[Tr]}{T} - r)/r = 100(\frac{[Tr]}{Tr} - 1). \tag{4}$$

It is to be noted that the error is non-positive for all $T$ and $r$, so that the variance of the Wiener process $W(r)$ over-estimates the variance of $X_T(r)$. An interesting feature of the error function (4) is that it can easily be seen that the error is not, in general, a monotonically decreasing function of the sample size. Take for example the case $r = .5$, the error is 0 for an even-valued sample size but is $100 * (.5/T)$ for any odd-valued sample size. There is, hence, a seesaw pattern in the mean-squared error which increases in importance as $r$ decreases. The error caused by this feature can be quite large. Take for example the case $r = .1$. The errors are as follows: 0% at $T = 20$, $-4.76\%$ at $T = 21$ up to $-31.03\%$ at $T = 29$, back to 0 at $T = 30$, $-3.23\%$ at $T = 31$ and up to $-23.08\%$ at $T = 39$.

In Figure 1, we present the percentage error function for values of $T$ ranging from 10 to 100 (in steps of 1) and for $r$ between 0.05 and 1.0 (in steps of .05). The seesaw pattern is seen to be quite important when $T$ and $r$ are small. When $r$ increases, it rapidly becomes negligeable. However, when $r$ is small (e.g. .1), the non-monotonic pattern remains important even when $T$ is as large as 100.

A closely related asymptotic result concerns the demeaned process

$$x_t^* = x_t - T^{-1}\sum_{t=1}^{T} x_t.$$

It follows from the basic functional central limit theorem (1), that

$$X_T^*(r) \equiv \sigma^{-1}T^{-1/2}\sum_{t=1}^{[Tr]} x_t^* \Rightarrow W(r) - rW(1) \equiv W^*(r),$$

where $W^*(r)$ is a standard Brownian Bridge which starts and ends at 0. It is easy to see that $var(W^*(r)) = r(1-r)$, and with *i.i.d.* $N(0,1)$ data,

$$Var(X_T^*(r)) = \frac{[Tr]}{T}(1 - \frac{[Tr]}{T}),$$

so that, in this case the error function is

$$e^*(T,r,0,0) = 100 * \left(\frac{[Tr]}{T}(1 - \frac{[Tr]}{T}) - r(1-r)\right)/(r(1-r)).$$

In Figure 2, we present the corresponding percentage error function again for values of $T$ ranging from 10 to 100 (in steps of 1) and for $r$ between 0.05 and 1.0 (in steps of .05). The behavior of this error function is rather different than from that of $e(T, r, 0, 0)$ presented in Figure 1. Here when $r < .5$, the error is non-positive while it is non-negative when $r > .5$. The error and the accompanying seesaw pattern can be quite important when $T$ is small and $r$ is close to either 0 or 1. When $r$ approaches .5, the error approaches 0, faster as $T$ is large.

This feature due to the truncation argument holds as well in the general $ARMA$ case. We shall, however, not dwell upon it and choose values of $T$ and $r$ such that $[Tr] = Tr$ to concentrate of the effect of the serial correlation.

## 4 Numerical Results for ARMA(1,1) Processes

We present results of the error function (3) for a wide range of values for $\alpha$ and $\theta$ and values of $T$ selected to be $T = 30$, 100, 500 and $5,000$. The values of $r$ are from .1 to .9 in steps of .1. The pairs $\{r, T\}$ selected avoid the non-monotonic behavior of the error caused by the truncation $[Tr]$ relative to $Tr$.

Figure 3 presents the error function $e(T, r, 0, \theta)$ corresponding to the case of an $MA(1)$ process. The graphs are constructed using values of $\theta$ between $-0.95$ and 0.95 in steps of .05. For any value of the MA parameter, the error function is positive and decreasing as $r$ increases. A notable feature is the fact that when the MA component is close to $-1$, the error is very large reaching as high a 14,000% for $T = 30$ when $r = .1$ and $\theta = -.95$. Even, when $r = 1$, it is of the order of 2,000%. As expected, the error function decreases as $T$ increases but very slowly when the MA parameter is large and negative. Even at $T = 5000$, the error with $\theta = -.95$ is between 10% and 80% depending on the value of $r$.

Figure 4 presents a similar graph for the case of an AR(1) process again for values of $\alpha$ between $-0.95$ and 0.95 in steps of .05. Here the results are quite different. First, the errors are nowhere near as large as in the MA case. The largest error is of the order of 160% when $T = 30$. An interesting feature is that the error is negative when $\alpha$ is positive and vice versa. The error declines in absolute value as $r$ increases. Also, the errors are larger for negative values of $\alpha$. The largest errors occur when $\alpha$ is either close to $-1$ or $+1$ in which case the rate of decrease as $T$ increases is rather slow. The rate of decrease is slower when $\alpha$ is positive, so that when $T$ is either 500 or 5000, the error is now larger (in absolute value) for positive than for negative values of $\alpha$.

We now consider the behavior of the error function as a function of the MA and AR parameters jointly. We consider the case $r = 1$ for illustration since the qualitative results are unchanged using other values of $r$. What transpires from Figure 5 is the fact that the error function is totally dominated by the region where both the MA and the AR coefficients are large and negative. In the extreme, with $\alpha = \theta = -.95$, the error is of the order of 50,000% when $T = 30$ and is still as high as 300% when $T = 5,000$. The error in this region is so large that we fail to see variations in the error function for other parameter configurations. To that effect, we present in Figure 6, the error function in the region where the AR coefficient $\alpha$ is positive. One can see that there are still important variations when the MA coefficient is negative. For example, with $T = 30$, $\alpha = 0$ and $\theta = -.95$, the error is sill around 1,400% and decreases to 0% when $\alpha = .95$ and $\theta = -.95$ (in which case the process is *i.i.d.* since the roots cancel). The error decreases quite rapidly but when $T = 5,000$ it is still around 8% when $\alpha = 0$ and $\theta = -.95$. Hence, we see that the error function is highest, by a wide margin when the MA coefficient is close to $-1$. The effect induced by the presence of an autoregressive component is to reduce the error function when $\alpha$ is positive and to increase it when $\alpha$ is negative reaching extremely high values as both the MA and AR coefficients get close to $-1$.

To better assess the effect on the error function of variations in the AR coefficient, Figure 7 presents the results for the parameter space where the MA coefficient is constrained to be non-negative. What is first of interest is the fact that variations in the MA coefficient from 0 to 0.95 basically induce no variation in the error function (except somewhat when the AR coefficient takes a high negative value in which case an increase in the MA coefficient decreases somewhat the error). Hence, the behavior of the error function as $\alpha$ varies is basically the same as that presented in Figure 3 for $r = 1$.

In summary, the results show that when the data exhibit little serial correlation, the Wiener process is a good approximation to the finite sample distribution of the scaled partial sums $X_T(r)$, except for some non-monotonicities caused by truncation which, however, dissipate rather quickly as the sample size increases. The presence of positive serial correlation of the moving-average type, even if quite large, does not induce a deterioration of the approximation. However, the presence of a negative moving average component causes serious distortions to the approximation. Such distortions become very important when the moving average coefficient approaches the boundary $-1$. The distortions are mitigated somewhat when a positive autoregressive component is present as well but exacerbated to a great extent with a negative autoregressive component. When both the MA and AR coefficients are close to $-1$ the error can reach as much as 50,000% when $T$ is small and reduces only very slowly as $T$ increases; the error remaining important even if $T$ is as large as 5,000.

When only an autoregressive component is present, the errors can be large for sample sizes like $T = 100$ but become small as $T$ increases further.

We have performed the same experiments with the demeaned process $X_T^*(r)$ in relation to its asymptotic approximation $W^*(r)$, the Brownian Bridge. The qualitative results are very similar when $r = .5$. The difference is that the magnitude of the errors decreases as $r$ approaches either 0 or 1, but the general shape remains the same. The interested reader may consult the full set of results in Mallet (1995).

## 5 Conclusion

Our analysis, though very simple, yields nevertheless interesting implications about the quality of the approximation provided by the class of Functional Central Limit Theorem when dependent errors are allowed. The main qualitative result is that care should be applied when using such an approximation when the data are heavily serially correlated. The problem is especially severe when a negative moving average component is present and even more so when it is accompanied by a negative autoregressive component.

The fact that the approximation is poor is such cases has clear implications for the adequacy of the asymptotic distribution of statistics that are based on the partial sums $X_T(r)$. A leading example is the least-squares estimate in a first-order autoregression. This case was extensively analyzed by Perron (1996) where it is shown that the asymptotic distribution provides a poor approximation to the finite sample distribution when the errors have a strongly negative MA component or a strongly positive or negative AR component. An alternative asymptotic framework that treats the coefficient as local to its relevant boundary was developed in Nabeya and Perron (1994) and found to yield quite accurate approximations. Another example is the class of unit root tests developed by Phillips (1987) and extended in Phillips and Perron (1988). Since these tests are directly based on estimates from a first-order autoregression, they inherit the same problems. This issue is addressed in details in Perron and Ng (1996) where the local framework of Nabeya and Perron (1994) is used to motivate alternative tests that suffer less from size distortions in the presence of substantial correlation in the data. Of course, given the widespread application of the FCLT in deriving asymptotic approximations, the results documented here have implications for many other estimators and test statistics.

## References


[1] Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press.

[2] Herrndorf, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *Annals of Probability* **12**, 141-153.

[3] Mallet, S. (1995): *Évaluation de la qualité du théorème limite central fonctionnel en présence d'erreurs dépendantes*, Unpublished M.Sc. Essay, Université de Montréal.

[4] Nabeya, S. and P. Perron (1994): "Local Asymptotic Distribution Related to the AR(1) Model with Dependent Errors," *Journal of Econometrics* **62**, 229-264.

[5] Perron, P. (1996): "The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors," *Journal of Econometrics* **70**, 317-350.

[6] Perron, P. and S. Ng (1996): "Useful Modifications to Unit Root Tests with Dependent Errors and their Local Asymptotic Properties," *Review of Economic Studies* **63**, 435-463.

[7] Phillips, P.C.B. (1987): "Time Series Regression with Unit Roots," *Econometrica* **55**, 277-302.

[8] Phillips, P.C.B. and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika* **75**, 335-346.

[9] Phillips, P.C.B. and V. Solo (1992), "Asymptotics for Linear Processes," *Annals of Statistics* **20**, 971-1001.

Figure 1: Relative Error in the i.i.d. Case; Wiener Process
(Negative of the value)

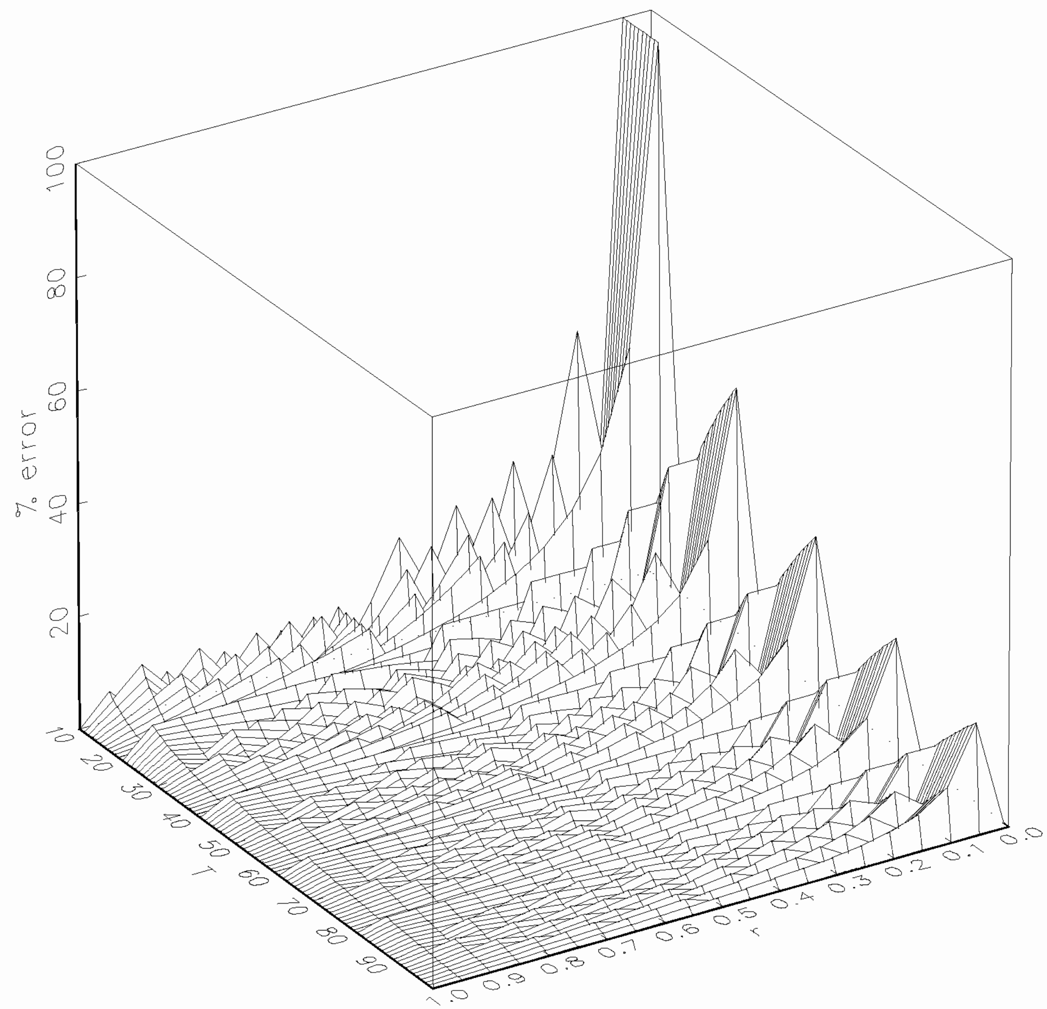

Figure 2: Relative Error in the i.i.d. Case; Brownian Bridge

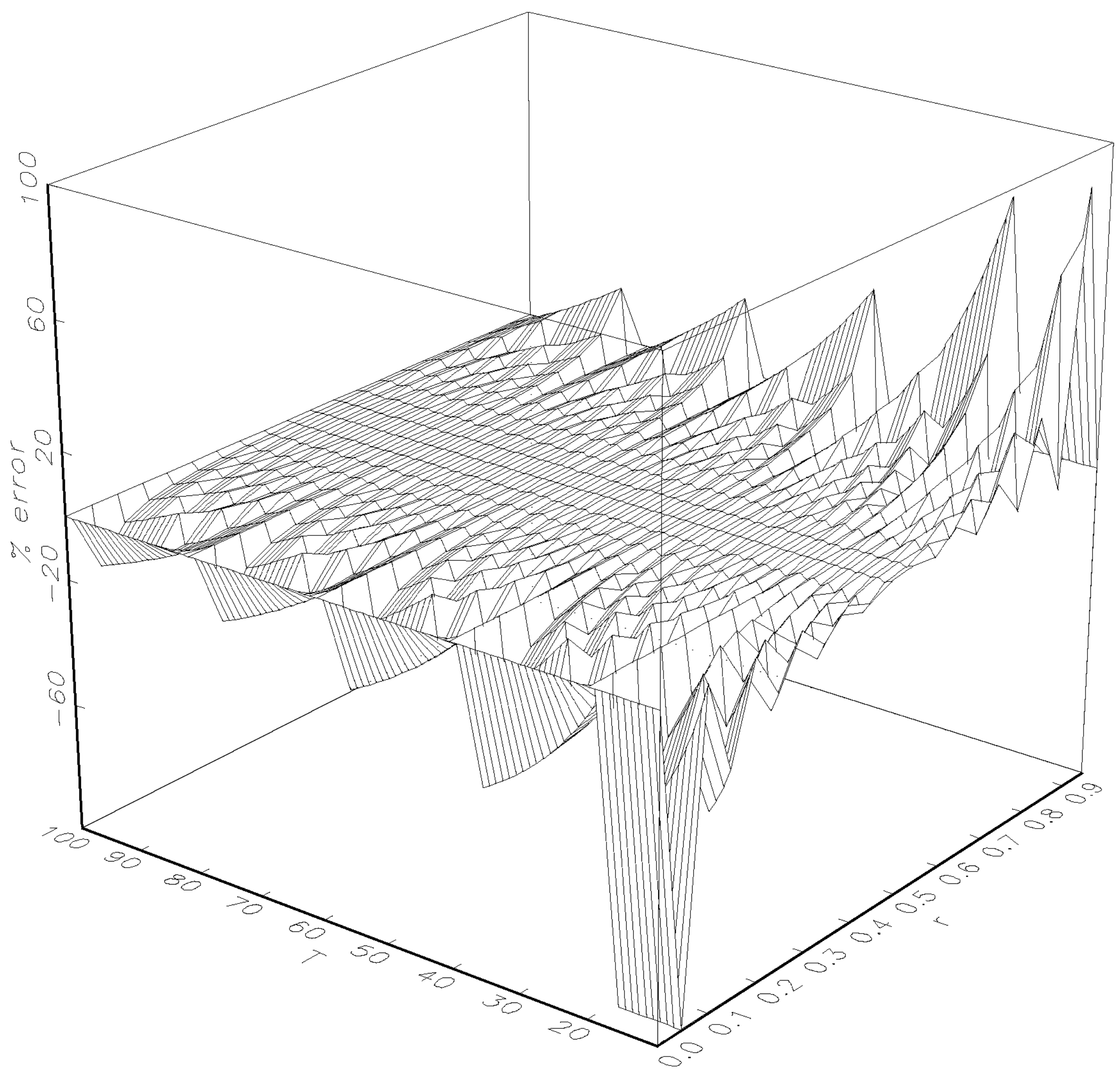

Figure 3: The Error Function in the MA Case

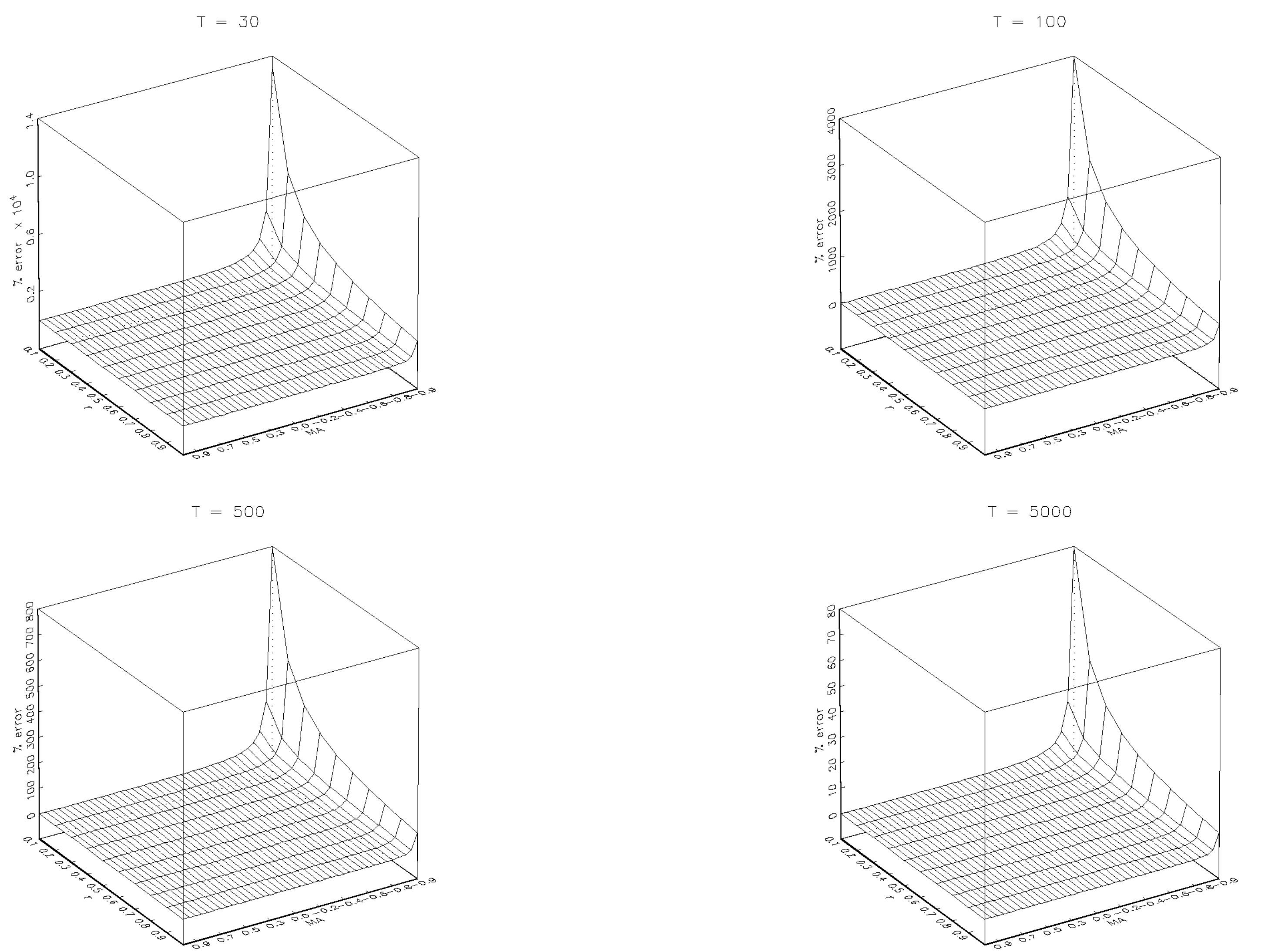

Figure 4: The Error Function in the AR Case

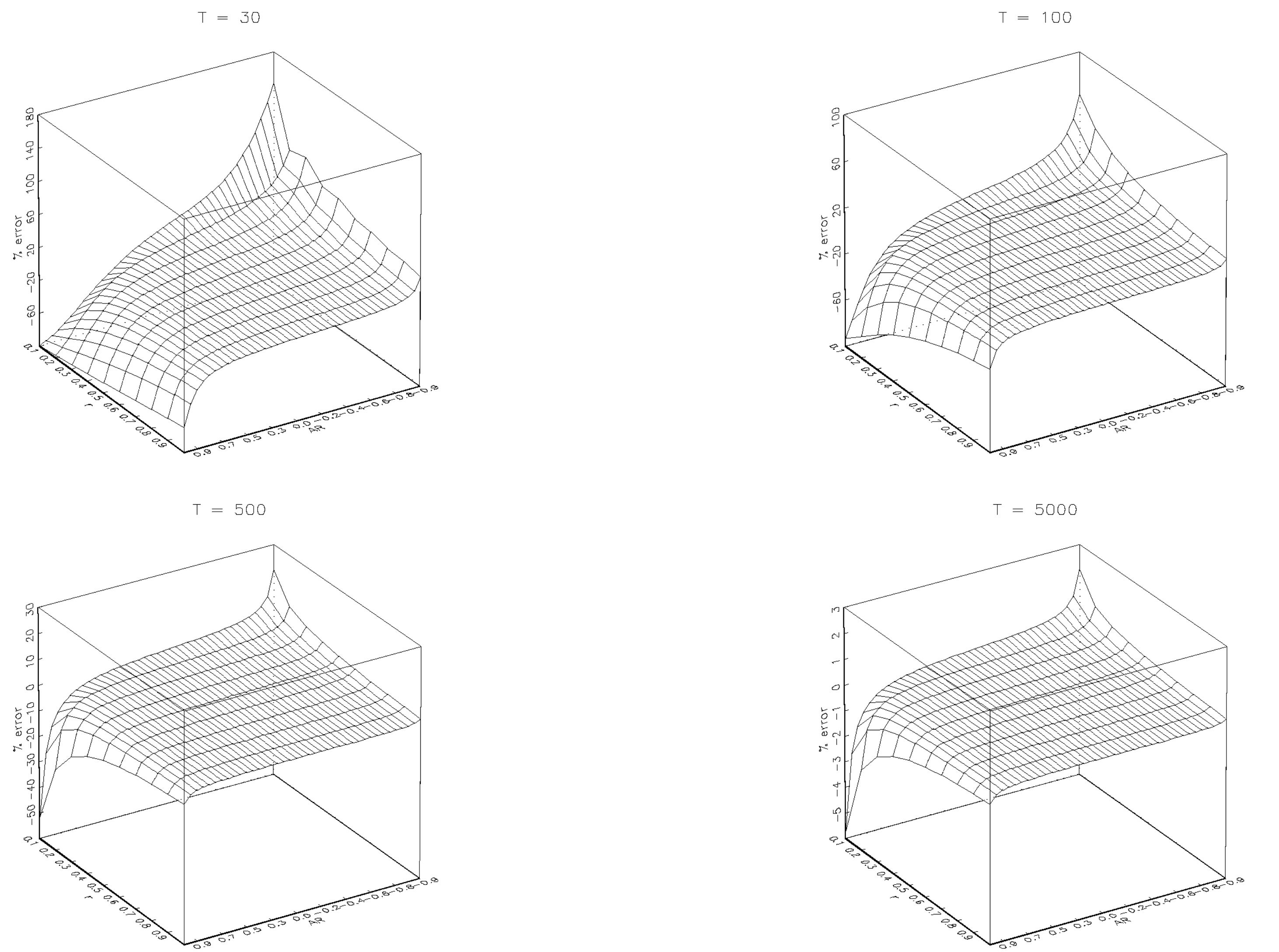

Figure 5: The Error Function in the ARMA(1,1) Case

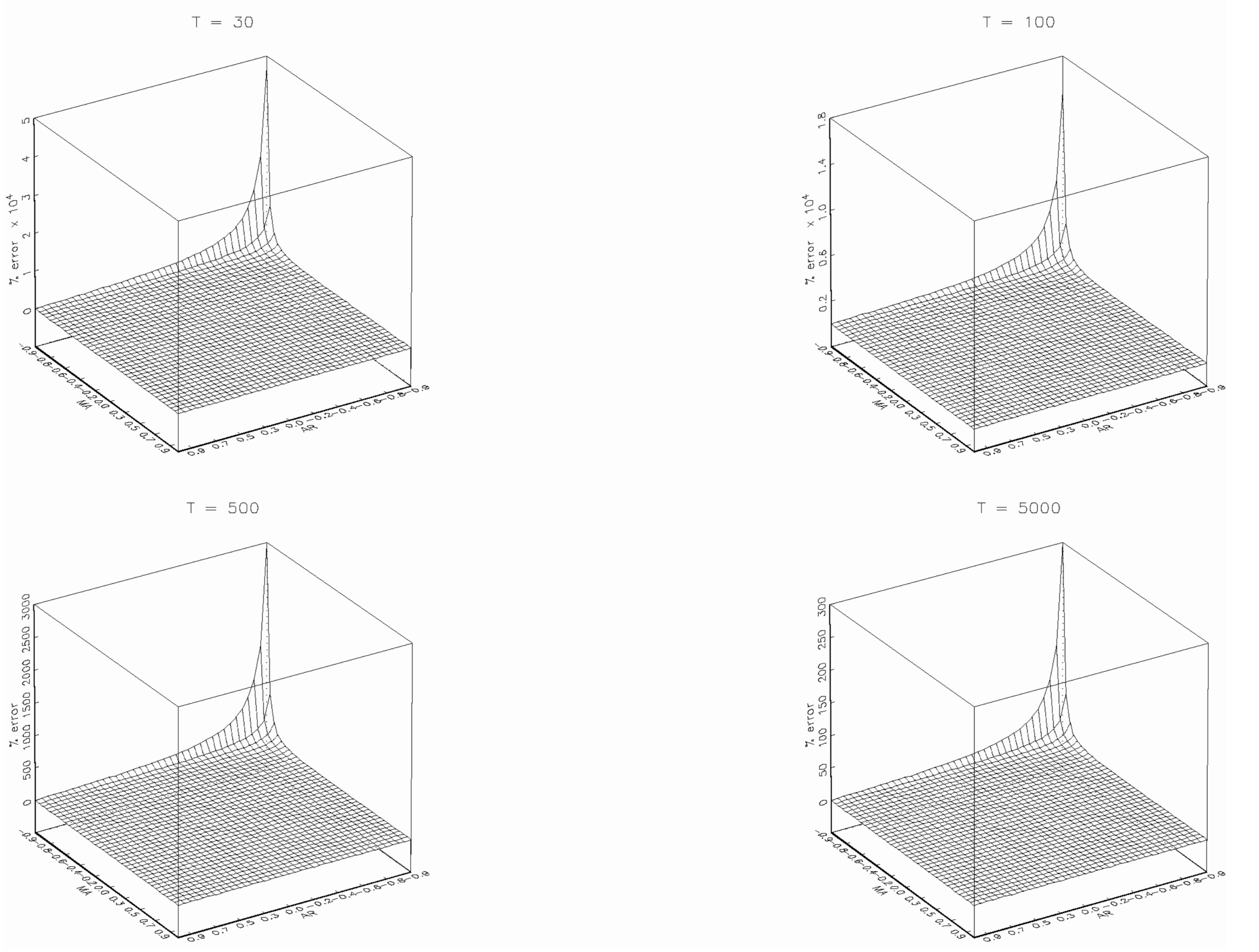

Figure 6: The Error Function in the ARMA(1,1) Case
(Positive AR coefficients)

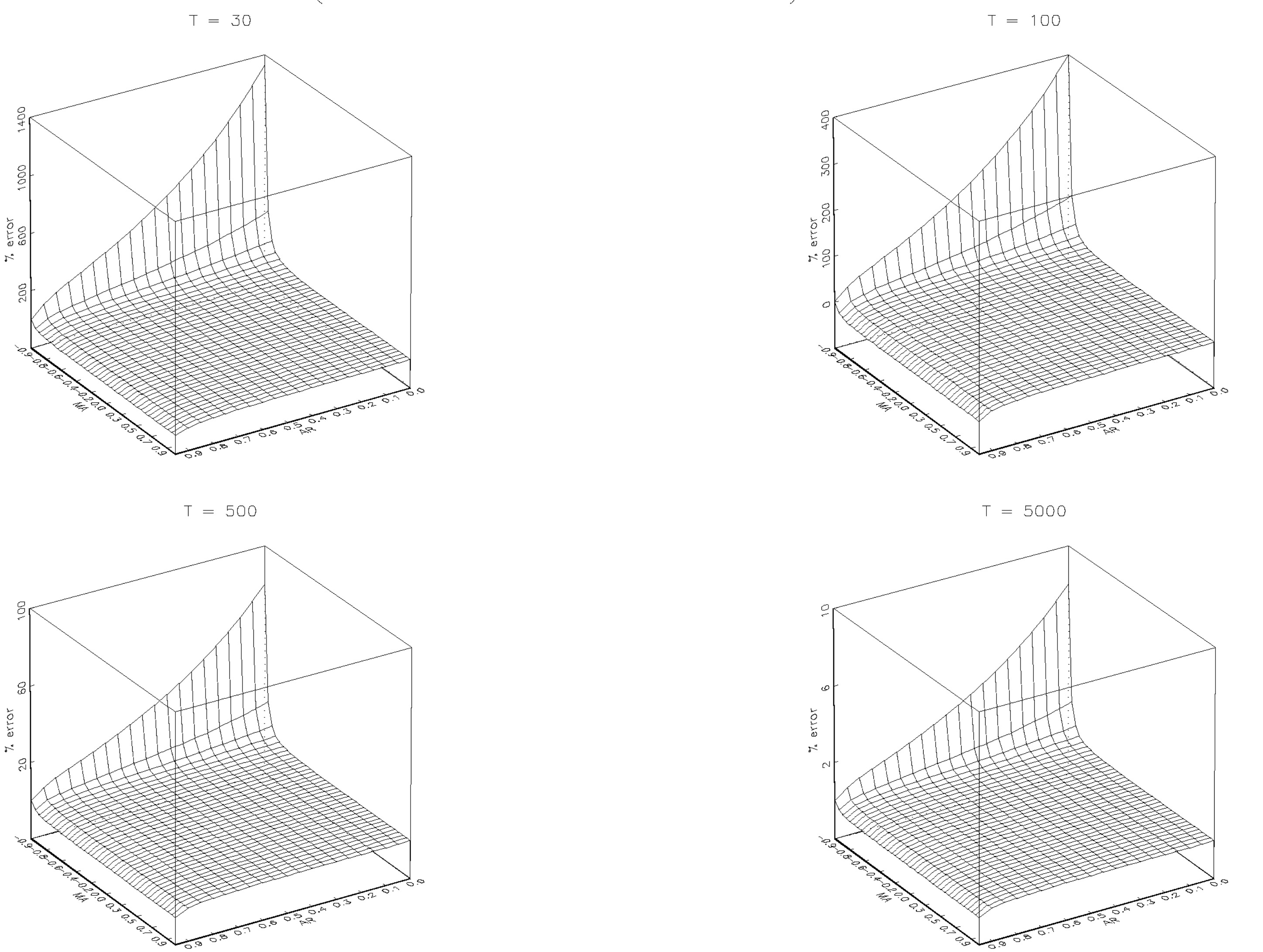

Figure 7: The Error Function in the ARMA(1,1) Case (Positive MA coefficients)