Why do Doctoral Students Choose the Wrong Central Limit Theorem When Data are Dependent?

Timothy Falcon Crack* and Olivier Ledoit†

November 11, 2008

*Corresponding Author, Professor of Finance, University of Otago, Department of Finance and Quantitative Analysis, PO Box 56, Dunedin, New Zealand, tcrack@otago.ac.nz

†Research Associate, Institute for Empirical Research in Economics, University of Zurich, oledoit@iew.uzh.ch

Why do Doctoral Students Choose the Wrong Central Limit Theorem When Data are Dependent?

Abstract

Doctoral students in Finance and Economics at top US schools consistently choose inappropriate central limit theorems (CLTs) when data are dependent. This is a serious problem, given the sustained high level of research interest in persistence and dependence in financial time series. We attribute the students' poor choices of CLTs to several causes. These include a lack of clarity in the presentation of simple CLTs in standard text books and a lack of access to worked examples applying advanced CLTs to problems with dependent data. We address these deficiencies by discussing appropriate use of simple CLTs and by illustrating the use of a CLT for dependent data in the derivation of the asymptotic distribution of the sample variance of a Gaussian AR(1) process. We also take this opportunity to show how students can use a simple Monte-Carlo simulation to check their derivation of an asymptotic distribution.

1 INTRODUCTION

Our experience is that doctoral students in finance and economics at top US schools consistently choose inappropriate central limit theorems when data are dependent. Although we have not conducted a formal experiment, our observations are far from anecdotal; they were made over a half-dozen years and apply to a sample of roughly 100 doctoral students in finance and economics from three top-tier US schools. The students making these poor choices are not a fresh cohort with little exposure to formal econometrics. Rather, they are typically second-year students who have already completed two or more excellent doctoral-level econometrics courses. This presents a serious problem because there is a sustained high level of research interest in persistence and dependence in financial data (stretching, for example, from Fama [1965], through Lo and MacKinlay [1988], to Bajgrowicz and Scaillet [2008] and beyond). Therefore, our students need to know how to apply central limit theorems correctly when data exhibit dependence.

2 THE PROBLEM

Over three years as a teaching assistant at MIT for both MIT and Harvard Finance and Economics doctoral students and another three years as a professor teaching a Financial Economics doctoral course at Indiana University, the first author observed that not one of the roughly 100 doctoral students in these classes was able to solve an econometric problem that required the use of a central limit theorem for dependent data. In about 95 percent of cases, the students applied the Lindberg-Levy or Lindberg-Feller Central Limit Theorems—which require independent data. In the remaining cases, the students were unable to construct any solution.

These students had typically already completed two or more excellent doctoral-level econometrics courses. Unfortunately, the standard texts for these courses do not always make clear the assumptions under which the simple central limit theorems contained therein apply, and they may, by their very nature, fail to contain more advanced central limit theorems.

For example, looking at the Lindberg-Levy and Lindberg-Feller Central Limit Theorems in Greene [2008], it is not at all clear that they do not apply to dependent data (see Theorems D.18A and D.19A in Greene [2008, pp. 1054–1055]). Only very careful reading of earlier material in the book, combined with some intelligent inference, reveals the full assumptions of these theorems. The assumptions for these two theorems are, however, clearly stated in more advanced books (see White [1984 and 2001, Theorems 5.2 and 5.6]; Feller [1968, p. 244] and Feller [1971, p. 262]; and Davidson [1997, Theorems 23.3 and 23.6]).

Unfortunately, even where the assumptions for the simple central limit theorems do appear clearly and where the more advanced central limit theorems for dependent data are present, it is nearly impossible to find any worked example showing the application of the more advanced theorems to concrete problems. This deficiency makes the general area the of application of advanced central limit theorems to cases of dependent data poorly accessible to many doctoral students.

We believe that the best way to address this problem is by combining our discussion of the issues with a worked example using a central limit theorem for dependent data in a simple case. So, in what follows, we derive the asymptotic distribution of the sample variance of a Gaussian AR(1) process using a central limit theorem from White [1984, 2001]. We also derive the asymptotic distribution of the sample mean for the process. This latter derivation does not need a central limit theorem, but the result is needed for the asymptotic distribution of the sample variance.

We have found that a Monte-Carlo simulation of the process and of the asymptotic distributions of the sample estimators aids student understanding significantly. We therefore present MATLAB code for a Monte-Carlo simulation, and we plot the resulting theoretical and simulated empirical asymptotic distributions.

Section 3 presents the Gaussian AR(1) process and our asymptotic distributional results for sample mean and variance. Section 4 discusses the Monte-Carlo simulation. Section 5 concludes. MATLAB code for the Monte-Carlo appears in Appendix A. Details of the derivations appear in Appendix B.

3 ASYMPTOTIC DISTRIBUTIONS

We assume that the random variable X_t follows a Gaussian AR(1) process:

$$X_t = \mu + \rho(X_{t-1} - \mu) + \epsilon_t, \tag{1}$$

where $\epsilon_t \sim \text{IID}\,\mathcal{N}(0, \sigma_\epsilon^2)$, "IID" means independent and identically distributed, and "$\mathcal{N}(a, b)$" denotes a Normal distribution with mean a and variance b. The only other assumption we make in the paper is that $|\rho| < 1$ (so that X_t is stationary).

The functional form of (1) is the simplest example of a non-IID data-generating process. By restricting our attention to an AR(1), we minimize the complexity of the dependence in the data while still being able to demonstrate the use of a central limit theorem for dependent data. Our asymptotic results may be derived without our assumption of Gaussian increments (e.g., using theorems in Fuller [1996, Section 6.3] or Brockwell and Davis [1991, Section 6.4]). The Gaussian specification of the problem allows, however, for a cleaner pedagogical illustration using an elegant central limit theorem from White [1984, 2001]. It also allows for a cleaner specification of the Monte-Carlo simulation we perform.

The Gaussian AR(1) process X_t is stationary and ergodic by construction (see the proof of Lemma 4 in Appendix B). Stationarity and ergodicity are strictly weaker than the IID assumption of the classical theorems in probability theory (e.g., the Lindberg-Levy and Lindberg-Feller Central Limit Theorems). Thus, these theorems do not apply. Stationarity and ergodicity are sufficient, however, for us to derive asymptotic results analogous to those available in the case where X_t is IID.

Let $\hat{\mu}$, and $\hat{\sigma}^2$ denote the usual sample mean and variance of the X_t's,

$$\hat{\mu} \equiv \frac{1}{n}\sum_{t=1}^{n} X_t\,, \quad \text{and} \quad \hat{\sigma}^2 \equiv \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \hat{\mu})^2. \tag{2}$$

The following two lemmas and theorem give the asymptotic distribution of the sample mean $\hat{\mu}$ of the Gaussian AR(1) process.

Lemma 1 *We have the following exact distributional result for a Gaussian AR(1):*

$$\sqrt{n}(\hat{\mu}-\mu)+\left[\frac{\rho}{1-\rho}\cdot\frac{X_n-X_0}{\sqrt{n}}\right]\sim\mathcal{N}\left(0,\frac{\sigma_\epsilon^2}{(1-\rho)^2}\right). \tag{3}$$

Proof: See Appendix B.

Lemma 2 *The following probability limit result holds for the second term on the left-hand side of (3):*

$$plim\left[\frac{\rho}{1-\rho}\cdot\frac{X_n-X_0}{\sqrt{n}}\right]=0. \tag{4}$$

Proof: See Appendix B.

Theorem 1 *We have the following asymptotic distributional result for the sample mean of a Gaussian AR(1) process:*[1]

$$\sqrt{n}(\hat{\mu}-\mu)\overset{A}{\sim}\mathcal{N}\left(0,\frac{\sigma^2(1+\rho)}{1-\rho}\right), \tag{5}$$

where σ^2 is the variance of X_t.

Proof: Apply Lemma 2 to (3) in Lemma 1 to deduce the asymptotic Normality of $\sqrt{n}(\hat{\mu} - \mu)$. Then use the stationarity of X_t (recall $|\rho| < 1$) to replace σ_ϵ^2 by $\sigma^2(1-\rho^2)$, thus completing the proof. This proof does not require a central limit theorem, but one is needed in the proof of Lemma 4. □

The following two lemmas and theorem give the asymptotic distribution of the sample variance $\hat{\sigma}^2$ of the Gaussian AR(1) process.

Lemma 3 *We may rewrite the term* $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ *as follows:*

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) = \sqrt{n}(s^2 - \sigma^2) - \sqrt{n}\left(s^2 - \left(\frac{n-1}{n}\right)\hat{\sigma}^2\right) + \frac{\hat{\sigma}^2}{\sqrt{n}}, \tag{6}$$

where $s^2 \equiv \frac{1}{n}\sum_{t=1}^{n}(X_t - \mu)^2$.

Proof: Direct algebraic manipulation and cancellation of terms. □

Lemma 4 *The following asymptotic distributional and probability limit results hold for the three terms on the right-hand side of (6):*

$$\begin{aligned} \sqrt{n}(s^2 - \sigma^2) &\overset{A}{\sim} \mathcal{N}\left(0, \frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)}\right), & (7)\\ plim\left[\sqrt{n}\left(s^2 - \left(\frac{n-1}{n}\right)\hat{\sigma}^2\right)\right] &= 0, \quad and & (8)\\ plim\left[\frac{\hat{\sigma}^2}{\sqrt{n}}\right] &= 0. & (9) \end{aligned}$$

Proof: This is the most difficult derivation. It requires a central limit theorem for dependent data. See Appendix B.

Theorem 2 *We have the following asymptotic distributional result for the sample variance of a Gaussian AR(1) process:*

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \overset{A}{\sim} \mathcal{N}\left(0, \frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)}\right). \tag{10}$$

Proof: Apply the three results in Lemma 4 to the three right-hand side terms, respectively, appearing in Lemma 3, and deduce the result directly. □

The asymptotic results for $\hat{\mu}$ in (5) of Theorem 1 and for $\hat{\sigma}^2$ in (10) of Theorem 2 have elegant interpretations. The higher is the degree of positive autocorrelation ρ, the larger is the standard error of both $\hat{\mu}$ and $\hat{\sigma}^2$—higher positive ρ means fewer effectively independent observations of X_t. Similarly, the higher is the degree of negative autocorrelation, then the larger is the standard error of $\hat{\sigma}^2$. We leave the reader with a small challenge: Deduce the qualitative explanation for why larger negative autocorrelation *reduces* the standard error of $\hat{\mu}$.

4 MONTE-CARLO SIMULATION

When students incorrectly use central limit theorems for independent data, they invariably conclude that the variance on the left-hand side of (10) is $2\sigma^4$ rather than $\frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)}$. You may then ask your students to perform a Monte-Carlo simulation of the Gaussian AR(1) process with $\rho \neq 0$, so that they can demonstrate for themselves that they have statistically significantly underestimated the true standard error.

A portion of our MATLAB code for the Monte-Carlo simulation appears in Appendix A. We choose the values $\mu = 0$, $\rho = 0.90$, and $\sigma_\epsilon = 0.50$. Figures 1 and 2 compare the realized empirical distribution to the theoretical results for both the asymptotic distribution of $\hat{\sigma}^2$ and the actual large sample distribution of $\hat{\sigma}^2$ (they are, of course, scaled versions of each other). We do not show the analogous results for $\hat{\mu}$.

Two pedagogical purposes are served by the Monte-Carlo simulation. First, our experience is that when a doctoral student simulates the process, repeatedly collects the asymptotic sample statistics, and then forms a distribution, he or she only then attains a clear concrete notion of what an asymptotic distribution actually is. Second, by comparing the realized asymptotic distribution to the derived theoretical one, the students understand the power of a Monte-Carlo in attempting to confirm or deny the consistency of a difficult analytical result—each of Figures 1 and 2 clearly distinguish between the competing asymptotic distributions.

Insert Figures 1 and 2 about here

5 CONCLUSIONS

In our experience, finance and economics doctoral students at top schools are unfamiliar with the use of central limit theorems for dependent data. They consistently and inappropriately use central limit theorems for independent data in cases where the data are clearly dependent. These students need exposure to the use of central limit theorems for dependent data. We meet this need by showing how to use a central limit theorem for dependent data to derive the asymptotic distribution of the sample estimator of the variance of a Gaussian AR(1) process. We also present a Monte-Carlo simulation to aid student understanding of asymptotic distributions and to illustrate the use of a Monte-Carlo in attempting to confirm or deny an analytical result.

A MATLAB MONTE-CARLO CODE

```
clear;rho=0.90;sigmae=0.50;mu=0;sigma=sigmae/sqrt(1-rho^2);

N=500000;NUMBREPS=10000; rseed=20081103; randn('seed',rseed);

collect=[ ];

for J=1:NUMBREPS

Y=[]; epsilon=randn(N,1); xpf=epsilon*sigmae;

bpf=1; apf=[1 -rho]; Y=filter(bpf,apf,xpf);

collect=[collect' [mean(Y) var(Y)]']';

end

asymeanv=0; asyvarv=2*(sigma^4)*(1+rho^2)/(1-rho^2);

asymeanv1=0; asyvarv1=2*(sigma^4); v=sqrt(N)*(collect(:,2)-sigma^2);

hpdf=[];mynormpdf=[];[M,X]=hist(v,250);M=M';X=X';dx=min(diff(X));

hpdf=M/(sum(M)*dx);

mynormpdf=(1/(sqrt(2*pi)*sqrt(asyvarv))).*exp(

-0.5*((X-asymeanv)/sqrt(asyvarv)).^2);

mynormpdf1=(1/(sqrt(2*pi)*sqrt(asyvarv1))).*exp(

-0.5*((X-asymeanv1)/sqrt(asyvarv1)).^2);

plot(X,[hpdf mynormpdf mynormpdf1],'k')

xlabel('Asymptotic Sample Variance of the Gaussian AR(1)');
```

B DERIVATIONS

Proof of Lemma 1: Rewrite the left-hand side of (3) in terms of the residual ϵ_t (the exact distribution of which is known).

$$
\begin{aligned}
& \sqrt{n}(\hat{\mu} - \mu) + \left[\frac{\rho}{1-\rho} \cdot \frac{X_n - X_0}{\sqrt{n}}\right] \\
= \quad & \frac{\sqrt{n}}{(1-\rho)} \left[(1-\rho)(\hat{\mu} - \mu) + \rho\left(\frac{X_n - X_0}{n}\right)\right] \\
= \quad & \frac{\sqrt{n}}{(1-\rho)} \left[(\hat{\mu} - \mu) - \rho\left[(\hat{\mu} - \mu) - \left(\frac{X_n - X_0}{n}\right)\right]\right] \\
= \quad & \frac{\sqrt{n}}{(1-\rho)} \left[\frac{1}{n}\sum_{t=1}^{n}(X_t - \mu) - \rho\left[\frac{1}{n}\sum_{t=1}^{n}(X_t - \mu) - \left(\frac{X_n - X_0}{n}\right)\right]\right] \\
= \quad & \frac{1}{\sqrt{n}(1-\rho)} \left[\sum_{t=1}^{n}(X_t - \mu) - \rho\left[\sum_{t=1}^{n}(X_t - \mu) - (X_n - X_0)\right]\right] \\
= \quad & \frac{1}{\sqrt{n}(1-\rho)} \left[\sum_{t=1}^{n}(X_t - \mu) - \rho\sum_{t=1}^{n}(X_{t-1} - \mu)\right] \\
= \quad & \frac{1}{\sqrt{n}(1-\rho)} \sum_{t=1}^{n}\left[(X_t - \mu) - \rho(X_{t-1} - \mu)\right] \\
= \quad & \frac{1}{\sqrt{n}(1-\rho)} \sum_{t=1}^{n}\epsilon_t,
\end{aligned}
$$

where the last line uses the definition of ϵ_t implicit within (1). We may now use $\epsilon_t \sim$ IID $\mathcal{N}(0, \sigma_\epsilon^2)$ to deduce

$$\frac{1}{\sqrt{n}(1-\rho)} \sum_{t=1}^{n} \epsilon_t \sim \mathcal{N}\left(0, \frac{\sigma_\epsilon^2}{(1-\rho)^2}\right),$$

thus proving the lemma. $\square$

Proof of Lemma 2: Let "var(·)," "cov(·, ·)," and "corr(·, ·)," denote the unconditional variance, covariance, and correlation operators, respectively. Let σ^2 denote $\mathrm{var}(X_t)$. The term $\rho(X_n - X_0)/[(1-\rho)\sqrt{n}\,]$ is shown to have variance of order $O(1/n)$ as follows:

$$\begin{aligned}
\mathrm{var}\left[\frac{\rho}{1-\rho} \cdot \frac{(X_n - X_0)}{\sqrt{n}}\right] &= \frac{1}{n}\left(\frac{\rho}{1-\rho}\right)^2 \mathrm{var}(X_n - X_0) \\
&= \frac{1}{n}\left(\frac{\rho}{1-\rho}\right)^2 [\mathrm{var}(X_n) + \mathrm{var}(X_0) \\
&\qquad - 2\,\mathrm{cov}(X_n, X_0)] \\
&= \frac{1}{n}\left(\frac{\rho}{1-\rho}\right)^2 \left[\sigma^2 + \sigma^2 - 2\,\mathrm{corr}(X_n, X_0)\sigma\sigma\right] \\
&\le \frac{4\sigma^2}{n}\left(\frac{\rho}{1-\rho}\right)^2. \qquad (11)
\end{aligned}$$

This derivation assumes $|\rho| < 1$ (so that stationarity of X_t gives $\mathrm{var}(X_n) = \mathrm{var}(X_0) = \sigma^2$). We also use $\mathrm{corr}(X_n, X_0) \ge -1$ at the last step.

Tchebychev's Inequality (Greene [2008, p. 1040]) says that for random variable V and small $\delta > 0$,

$$P(|V - E(V)| > \delta) \leq \frac{\text{var}(V)}{\delta^2}.$$

We may apply Tchebychev's Inequality to $V_n \equiv \rho(X_n - X_0)/[(1-\rho)\sqrt{n}]$, and use (11) to find

$$\begin{aligned} P\left(\left|\frac{\rho}{1-\rho} \cdot \frac{(X_n - X_0)}{\sqrt{n}}\right| > \delta\right) &\leq \frac{\text{var}(V_n)}{\delta^2} \\ &\leq \frac{4\sigma^2}{n\delta^2}\left(\frac{\rho}{1-\rho}\right)^2. \end{aligned}$$

Thus, for any $\delta > 0$, we have $\lim_{n\to\infty} P(|V_n| > \delta) = 0$. That is, $\text{plim}\, V_n = 0$, thus proving the lemma. $\square$

Proof of Lemma 4: We demonstrate each of Equations (7), (8), and (9) in turn. We begin with the proof of the asymptotic result in (7):

$$\sqrt{n}(s^2 - \sigma^2) \overset{A}{\sim} \mathcal{N}\left(0, \frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)}\right),$$

where $s^2 \equiv \frac{1}{n}\sum_{t=1}^{n}(X_t - \mu)^2$, and $\sigma^2 = \text{var}(X_t)$. To derive this result, we apply the following central limit theorem for non-IID data adapted directly from White [1984].

Theorem (from White [1984, Theorem 5.15, p. 118]) *Let $\mathcal{F}_t$ be the sigma-algebra generated by the entire current and past history of a stochastic variable Z_t; let $\mathcal{R}_{t,j}$ be the revision made in forecasting Z_t when information becomes available at time $t-j$, that is, $\mathcal{R}_{t,j} \equiv E(Z_t|\mathcal{F}_{t-j}) - E(Z_t|\mathcal{F}_{t-j-1})$; let $\bar{Z}_n$ denote the sample mean of $Z_1, \ldots, Z_n$; and let $\bar{\sigma}_n^2 \equiv var(\sqrt{n}\bar{Z}_n)$. Then, if the sequence $\{Z_t\}$ satisfies the following conditions: 1. $\{Z_t\}$ is stationary;*[2] *2. $\{Z_t\}$ is ergodic; 3. $E(Z_t^2) < \infty$; 4. $E(Z_0|\mathcal{F}_{-m}) \stackrel{q.m.}{\to} 0$ as $m \to \infty$; and 5. $\sum_{j=0}^{\infty} \left[var\left(\mathcal{R}_{0,j}\right)\right]^{1/2} < \infty$, we obtain the results $\bar{\sigma}_n^2 \to \bar{\sigma}^2$, as $n \to \infty$, and if $\bar{\sigma}^2 > 0$, then $\frac{\sqrt{n}\bar{Z}_n}{\bar{\sigma}} \stackrel{A}{\sim} \mathcal{N}(0,1)$.*

We apply the theorem to $Z_t \equiv (X_t - \mu)^2 - \sigma^2$. With this definition of Z_t, we obtain $\bar{Z}_n = (1/n)\sum_{t=1}^{n} Z_t = s^2 - \sigma^2$, and, thus, $\sqrt{n}\bar{Z}_n = \sqrt{n}(s^2 - \sigma^2)$. However, before we can apply the theorem, we must check that its five conditions are satisfied, and we must calculate $\lim_{n\to\infty} \bar{\sigma}_n^2 = \lim_{n\to\infty} \mathrm{var}(\sqrt{n}\bar{Z}_n)$. We begin by checking the five conditions.

Condition 1: We have assumed $|\rho| < 1$. Thus, our Gaussian AR(1) process X_t is stationary. Stationarity of X_t yields stationarity of Z_t immediately (by definition of Z_t).

Condition 2: White [2001, p. 48] uses Ibragimov and Linnik [1971, pp. 312–313] to deduce that a Gaussian AR(1) with $|\rho| < 1$ is strong mixing. White [2001, p. 48] then uses Rosenblatt [1978] to state that strong mixing plus stationarity (recall $|\rho| < 1$) implies ergodicity. It follows that X_t is ergodic. This yields ergodicity of Z_t immediately (by definition of Z_t).

Condition 3: We note first that since ϵ_t is Gaussian, then so too is X_t [Hamilton 1994, p. 118]. It is well known that if $X_t \sim \mathcal{N}(\mu, \sigma^2)$, then $E[(X_t - \mu)^4] = 3\sigma^4$. It follows that

$$\begin{aligned} E(Z_t^2) &= E[((X_t - \mu)^2 - \sigma^2)^2] \\ &= E[(X_t - \mu)^4 - 2\sigma^2(X_t - \mu)^2 + \sigma^4] \\ &= 3\sigma^4 - 2\sigma^4 + \sigma^4 = 2\sigma^4 < \infty. \end{aligned} \quad (12)$$

Condition 4: To show that $E(Z_0|\mathcal{F}_{-m}) \overset{q.m.}{\to} 0$ as $m \to \infty$, we must show that $E(Z_t|\mathcal{F}_{t-m}) \overset{q.m.}{\to} 0$ as $m \to \infty$ in the special case $t = 0$. In fact, we can prove convergence in quadratic mean for *any* t if we can show $E([E(Z_t|\mathcal{F}_{t-m})]^2) \to 0$ as $m \to \infty$ (see White [1984, p. 117]). To derive

$E(Z_t|\mathcal{F}_{t-m})$, we first consider the term $Z_t + \sigma^2 = (X_t - \mu)^2$ as follows:

$$\begin{aligned} X_t - \mu &= \rho(X_{t-1} - \mu) + \epsilon_t \\ &\vdots \\ &= \rho^m(X_{t-m} - \mu) + \sum_{k=0}^{m-1} \rho^k \epsilon_{t-k}. \end{aligned} \tag{13}$$

With $Z_t + \sigma^2 = (X_t - \mu)^2$, it follows from (13) that

$$\begin{aligned} E(Z_t + \sigma^2|\mathcal{F}_{t-m}) &= E\left[\left(\rho^m(X_{t-m} - \mu) + \sum_{k=0}^{m-1} \rho^k \epsilon_{t-k}\right)^2 \middle| \mathcal{F}_{t-m}\right] \\ &= \rho^{2m}(X_{t-m} - \mu)^2 + 0 + \sum_{k=0}^{m-1} \rho^{2k}\sigma_\epsilon^2 \\ &= \rho^{2m}(X_{t-m} - \mu)^2 + \left(\frac{1-\rho^{2m}}{1-\rho^2}\right)\sigma_\epsilon^2 \\ &= \rho^{2m}(X_{t-m} - \mu)^2 + \left(\frac{1-\rho^{2m}}{1-\rho^2}\right)[\sigma^2(1-\rho^2)] \\ &= \rho^{2m}(X_{t-m} - \mu)^2 + \sigma^2(1-\rho^{2m}). \end{aligned} \tag{14}$$

If we now cancel σ^2 from both sides of (14), we find

$$E(Z_t|\mathcal{F}_{t-m}) = \rho^{2m}[(X_{t-m} - \mu)^2 - \sigma^2] = \rho^{2m} Z_{t-m}. \tag{15}$$

It follows that $E([E(Z_t|\mathcal{F}_{t-m})]^2) = E([\rho^{2m} Z_{t-m}]^2) = \rho^{4m} E(Z_{t-m}^2) =$

$\rho^{4m} 2\sigma^4$ (using (12) and stationarity of Z_t). With $|\rho| < 1$, we deduce that $E([E(Z_t|\mathcal{F}_{t-m})]^2) \to 0$ as $m \to \infty$, and, thus, that $E(Z_t|\mathcal{F}_{t-m}) \overset{q.m.}{\to} 0$ as $m \to \infty$ (using White [1984, p. 117]), as required.

Condition 5: Applying (15) to the definition of $\mathcal{R}_{t,j}$ yields

$$\begin{aligned} \mathcal{R}_{t,j} &\equiv E(Z_t|\mathcal{F}_{t-j}) - E(Z_t|\mathcal{F}_{t-j-1}) \\ &= \rho^{2j} Z_{t-j} - \rho^{2(j+1)} Z_{t-(j+1)}. \end{aligned} \tag{16}$$

By definition, $E(Z_t) = 0$, so $E(\mathcal{R}_{t,j}) = 0$, and, thus, $\mathrm{var}(\mathcal{R}_{t,j}) = E(\mathcal{R}_{t,j}^2)$. Manipulating (16), we get

$$\begin{aligned} \mathrm{var}(\mathcal{R}_{t,j}) &= E(\mathcal{R}_{t,j}^2) \\ &= E([\rho^{2j} Z_{t-j} - \rho^{2(j+1)} Z_{t-(j+1)}]^2) \\ &= \left(\rho^{4j} + \rho^{4(j+1)}\right) 2\sigma^4 - 2\rho^{4j+2} E(Z_{t-j} Z_{t-(j+1)}), \\ &= \left(\rho^{4j} + \rho^{4(j+1)}\right) 2\sigma^4 - 2\rho^{4j+2} E(Z_t Z_{t-1}), \end{aligned} \tag{17}$$

where we used (12) and the fact that $E(Z_{t-j}) = E(Z_{t-(j+1)}) = 0$. We also used stationarity of Z_t to rewrite $E(Z_{t-j} Z_{t-(j+1)})$ as $E(Z_t Z_{t-1})$.

The term $E(Z_t Z_{t-1})$ in (17) may be expanded as follows:

$$
\begin{aligned}
E(Z_t Z_{t-1}) &= E[((X_t-\mu)^2-\sigma^2)((X_{t-1}-\mu)^2-\sigma^2)] \\
&= E[(X_t-\mu)^2(X_{t-1}-\mu)^2]-\sigma^4.
\end{aligned}
$$

Plugging this expression for $E(Z_t Z_{t-1})$ into (17) gives

$$
\begin{aligned}
\operatorname{var}(\mathcal{R}_{t,j}) &= \left(\rho^{4j}+\rho^{4(j+1)}\right)2\sigma^4 \\
&\quad -2\rho^{4j+2}\left(E[(X_t-\mu)^2(X_{t-1}-\mu)^2]-\sigma^4\right) \\
&= \left(\rho^{4j}+\rho^{4(j+1)}\right)2\sigma^4-2\rho^{4j+2}\left(E(Y_t^2\,Y_{t-1}^2)-\sigma^4\right), \quad (18)
\end{aligned}
$$

where $Y_t \equiv (X_t-\mu)$. The term $E(Y_t^2\,Y_{t-1}^2)$ is a special case of a more general term $E(Y_t^2\,Y_{t-d}^2)$, which we now evaluate (we need the general term later in the proof). From the definition of the Gaussian AR(1) ((1)) and from (13), we deduce that $Y_t=\rho^d Y_{t-d}+\sum_{k=0}^{d-1}\rho^k\epsilon_{t-k}$ and that Y_t is Gaussian with zero-mean. It follows that

$$
E(Y_t^2\,Y_{t-d}^2) = E\left[\left(\rho^d Y_{t-d}+\sum_{k=0}^{d-1}\rho^k\epsilon_{t-k}\right)^2 Y_{t-d}^2\right]
$$

$$\begin{aligned}
&= \rho^{2d} E(Y_{t-d}^4) + 2\rho^d E\left(\sum_{k=0}^{d-1} \rho^k \epsilon_{t-k}\right)^2 E(Y_{t-d}^3) \\
&\quad + E\left[\left(\sum_{k=0}^{d-1} \rho^k \epsilon_{t-k}\right)^2\right] E(Y_{t-d}^2) \\
&= 3\rho^{2d}\sigma^4 + 0 + \sigma^2 \sum_{k=0}^{d-1} \rho^{2k} \sigma_\epsilon^2 \\
&= 3\rho^{2d}\sigma^4 + \sigma^2 \left(\frac{1-\rho^{2d}}{1-\rho^2}\right) [\sigma^2(1-\rho^2)] \\
&= \sigma^4(1+2\rho^{2d}), \qquad (19)
\end{aligned}$$

where we used independence of Y_{t-d} and ϵ_{t-k} for $k < d$ to separate expectations in the cross-product term. We also used the mean-zero Normality of Y_{t-d} to write $E(Y_{t-d}^3) = 0$, and $E(Y_{t-d}^4) = 3\sigma^4$. If we now set $d = 1$ in (19) and plug this into (18), we obtain

$$\begin{aligned}
\text{var}(\mathcal{R}_{t,j}) &= \left(\rho^{4j} + \rho^{4(j+1)}\right) 2\sigma^4 - 2\rho^{4j+2}\left(\sigma^4(1+2\rho^2) - \sigma^4\right) \\
&= 2\sigma^4(1-\rho^4)\rho^{4j}. \qquad (20)
\end{aligned}$$

Thus, $[\text{var}(\mathcal{R}_{t,j})]^{1/2} = \sqrt{E(\mathcal{R}_{t,j}^2)} = \sqrt{2\sigma^4(1-\rho^4)}\,\rho^{2j}$. It follows that

$$\sum_{j=0}^{\infty} (\text{var}\mathcal{R}_{t,j})^{1/2} = \sqrt{2\sigma^4(1-\rho^4)} \sum_{j=0}^{\infty} \rho^{2j}$$

$$\begin{aligned}
&= \frac{\sqrt{2\sigma^4(1-\rho^4)}}{1-\rho^2} \\
&= \frac{\sqrt{2\sigma^4(1+\rho^2)(1-\rho^2)}}{1-\rho^2} \\
&= \sqrt{\frac{2\sigma^4(1+\rho^2)}{1-\rho^2}} \;<\; \infty.
\end{aligned}$$

This latter result holds in the special case $t = 0$, so the fifth and final prerequisite for applying White's Theorem to Z_t is satisfied.

We must now find $\bar{\sigma}^2 \equiv \lim_{n\to\infty} \bar{\sigma}_n^2 = \lim_{n\to\infty} \text{var}(\sqrt{n}\bar{Z}_n)$. Recall that we have $Z_t = (X_t - \mu)^2 - \sigma^2$, so that $\sqrt{n}\bar{Z}_n = \sqrt{n}(s^2 - \sigma^2)$, where $s^2 \equiv (1/n)\sum_{t=1}^n (X_t - \mu)^2 = (1/n)\sum_{t=1}^n Y_t^2$. With σ^2 a constant, we know that $\text{var}(\sqrt{n}\bar{Z}_n) = \text{var}(\sqrt{n}s^2)$. It is easier to work with $\text{var}(ns^2)$, so we do that and then adjust the result.

$$\begin{aligned}
\text{var}(ns^2) &= \text{var}\left(\sum_{t=1}^n Y_t^2\right) \\
&= E\left[\left(\sum_{t=1}^n Y_t^2\right)^2\right] - \left[E\left(\sum_{t=1}^n Y_t^2\right)\right]^2 \\
&= E\left[\sum_{t=1}^n \sum_{s=1}^n Y_t^2 Y_s^2\right] - (n\sigma^2)^2 \\
&= \left[2\sum_{t=2}^n \sum_{d=1}^{t-1} E\left(Y_t^2 Y_{t-d}^2\right) + nE(Y_t^4)\right] - (n\sigma^2)^2 \\
&= \left[2\sigma^4 \sum_{t=2}^n \sum_{d=1}^{t-1} \left(1 + 2\rho^{2d}\right) + 3n\sigma^4\right] - (n\sigma^2)^2, \qquad (21)
\end{aligned}$$

where we used (19) to replace $E\left(Y_t^2 Y_{t-d}^2\right)$. If we divide (21) by σ^4 and combine the final two terms, we get

$$\begin{aligned}
\frac{\mathrm{var}(ns^2)}{\sigma^4} &= 2\sum_{t=2}^{n}\sum_{d=1}^{t-1}\left(1+2\rho^{2d}\right)+n(3-n) \\
&= 2\sum_{t=2}^{n}\left[(t-1)+2\rho^2\left(\frac{1-\rho^{2(t-1)}}{1-\rho^2}\right)\right]+n(3-n)
\end{aligned}$$

It is easily shown that $2\sum_{t=2}^{n}(t-1) = -n(3-n)+2n$, so we get some cancellation as follows:

$$\begin{aligned}
\frac{\mathrm{var}(ns^2)}{\sigma^4} &= \frac{4\rho^2}{1-\rho^2}\left[(n-1)-\sum_{t=2}^{n}\rho^{2(t-1)}\right]+2n \\
&= \frac{4\rho^2(n-1)+2n(1-\rho^2)}{1-\rho^2}-\frac{4\rho^2}{1-\rho^2}\cdot\rho^2\left(\frac{1-\rho^{2(n-1)}}{1-\rho^2}\right) \\
&= \frac{4\rho^2 n+2n-2n\rho^2}{1-\rho^2}-\frac{4\rho^2}{1-\rho^2}\left[1+\rho^2\left(\frac{1-\rho^{2(n-1)}}{1-\rho^2}\right)\right] \\
&= \frac{2n(1+\rho^2)}{(1-\rho^2)}-\frac{4\rho^2(1-\rho^{2n})}{(1-\rho^2)^2}. \qquad (22)
\end{aligned}$$

It follows immediately that $\mathrm{var}(\sqrt{n}s^2) \to \frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)}$ as $n \to \infty$. Using this result in the last part of White's theorem yields

$$\sqrt{n}(s^2-\sigma^2) \overset{A}{\sim} \mathcal{N}\left(0, \frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)}\right),$$

thus proving (7)—the first of the three parts of Lemma 4.

To demonstrate (8)—the second of the three parts of Lemma 4—we need the probability limit of $[\sqrt{n}\,(s^2-((n-1)/n)\,\hat{\sigma}^2)]$. Direct algebraic manipulation yields

$$\begin{aligned}
\left[\sqrt{n}\left(s^2-\left(\frac{n-1}{n}\right)\hat{\sigma}^2\right)\right] &= \sqrt{n}(\hat{\mu}-\mu)^2 \\
&= \frac{1}{\sqrt{n}}\left[\frac{\sigma^2(1+\rho)}{1-\rho}\right]\cdot\left[\frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{\frac{\sigma^2(1+\rho)}{1-\rho}}}\right]^2 \\
&= \frac{1}{\sqrt{n}}\left[\frac{\sigma^2(1+\rho)}{1-\rho}\right]\mathcal{Z}_n^2, \qquad (23)
\end{aligned}$$

where $\mathcal{Z}_n \equiv \left[\frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{\frac{\sigma^2(1+\rho)}{1-\rho}}}\right]$ is asymptotically standard Normal (a consequence of Theorem 1). We may now apply a result analogous to Slutsky's Theorem for probability limits (see Greene [2008, p. 1045]) to deduce that $\mathcal{Z}_n^2 \overset{A}{\sim} \chi_1^2$ (that is, $\mathcal{Z}_n^2$ is asymptotically chi-square with one degree of freedom). Thus, $\mathcal{Z}_n^2$ is of bounded variance. It follows that one application of Tchebychev's Inequality to (23) produces the result:

$$\text{plim}\left[\sqrt{n}\left(s^2-\left(\frac{n-1}{n}\right)\hat{\sigma}^2\right)\right]=0,$$

thus proving (8)—the second of the three parts of Lemma 4.

To demonstrate (9)—the third and final part of Lemma 4—we need the probability limit of $(\hat{\sigma}^2/\sqrt{n})$. Algebraic manipulation gives

$$\hat{\sigma}^2 = \left(\frac{n}{n-1}\right) s^2 - \left(\frac{n}{n-1}\right) (\hat{\mu} - \mu)^2. \tag{24}$$

The variance of s^2 goes to zero as $n \to \infty$ (a consequence of (22)). The variance of $(\hat{\mu} - \mu)^2$ goes to zero as $n \to \infty$ (a consequence of $\mathcal{Z}_n^2 \overset{A}{\sim} \chi_1^2$, from above). In (24), the coefficients $n/(n-1) \to 1$ as $n \to \infty$. It follows that $\mathrm{var}(\hat{\sigma}^2) \to 0$ with n. An application of Tchebychev's Inequality yields immediately

$$\mathrm{plim}\left[\frac{\hat{\sigma}^2}{\sqrt{n}}\right] = 0,$$

thus proving the third and final part of Lemma 4. $\square$

ENDNOTES

1. If a sequence b_n of random variables converges in distribution to a random variable $\mathcal{Z}$ (often written "$b_n \xrightarrow{d} \mathcal{Z}$"), then b_n is said to be *asymptotically distributed* as $F_{\mathcal{Z}}$, where $F_{\mathcal{Z}}$ is the distribution of $\mathcal{Z}$. This is denoted here by "$b_n \overset{A}{\sim} F_{\mathcal{Z}}$" (as in White [2001, p. 66]).

2. Note that White's "stationarity" is *strict stationarity.* That is, $\{Z_t\}_{t=1}^{\infty}$and $\{Z_{t-k}\}_{t=1}^{\infty}$ have the same joint distribution for every $k > 0$ (see White [2001, p. 43] and Davidson [1997, p. 193]).

REFERENCES

Bajgrowicz, P. and O. Scaillet. "Technical Trading Revisited: Persistence Tests, Transaction Costs, and False Discoveries," *Swiss Finance Institute Research Paper No. 08-05*, (January 1, 2008) Paper available at SSRN: http://ssrn.com/abstract=1095202.

Brockwell, P. J., and R. A. Davis, *Time Series: Theory and Methods* (New York 1991), 2nd Edition, Springer.

Davidson, J. *Stochastic Limit Theory* (New York, 1997), Oxford University Press.

Fama, E. F. "The Behavior of Stock Market Prices," *The Journal of Business* 38 (1965), 34–105.

Feller, W. *An Introduction to Probability Theory and Its Applications* (New York, 1968), Volume I, 3rd Edition, John Wiley and Sons.

Feller, W. *An Introduction to Probability Theory and Its Applications* (New York, 1971), Volume II, 2nd Edition, John Wiley and Sons.

Fuller, W. A. *Introduction to Statistical Time Series* (New York, 1996), 2nd Edition, John Wiley and Sons.

Greene, W. H. *Econometric Analysis* (Upper Saddle River, 2008), 6th Edition, Prentice Hall.

Hamilton, J. D. *Time Series Analysis* (Princeton, 1994), Princeton University Press.

Ibragimov, I. A. and Y. V. Linnik *Independent and Stationary Sequences of Random Variables* (The Netherlands, 1971), ed. by J. F. C. Kingman, Wolters-Noordhoff Publishing Groningen.

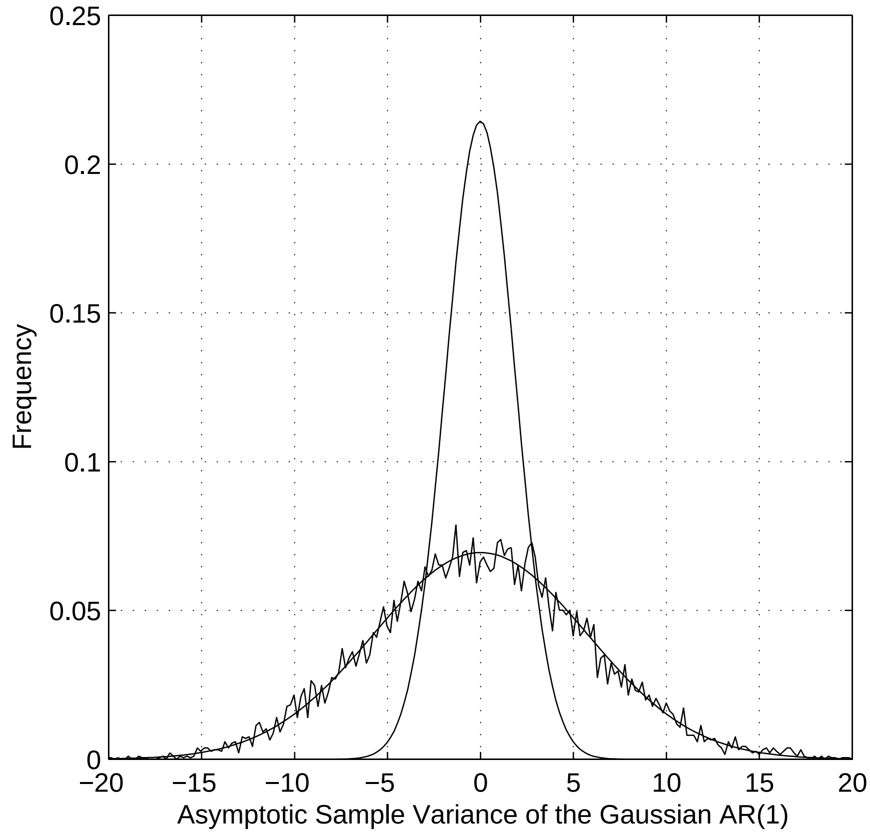
Lo, A. W. and A. C. MacKinlay, "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1 (1988), 41–66.

Rosenblatt, M. "Dependence and Asymptotic Dependence for Random Processes," *Studies in Probability Theory* (Washington, D.C., 1978), Murray Rosenblatt (ed.), Mathematical Association of America, 24–45.

White, H. *Asymptotic Theory for Econometricians* (San Diego, 1984), Academic Press.

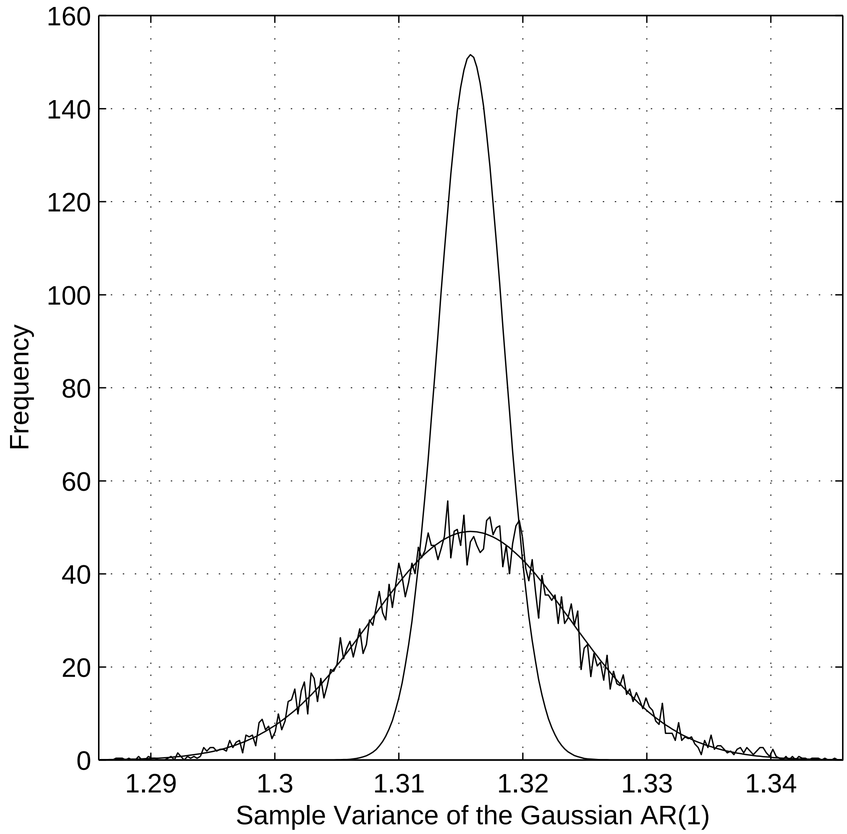
White, H. *Asymptotic Theory for Econometricians* (San Diego, 2001), Revised 2nd Edition, Academic Press.

Figure 1: Histogram of Simulated Empirical pdf of $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$



We use MATLAB to simulate a time series of 500,000 observations of the Gaussian AR(1) using $\rho = 0.90$, $\sigma_\epsilon = 0.50$, and $\mu = 0$. We then record the sample variance $\hat{\sigma}^2$ of the process. We repeat this 10,000 times and plot (the uneven line) the realized density of $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$. We overlay on the plot the correct theoretical density $\mathcal{N}\left(0, \frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)}\right)$ and the most common incorrect student-derived theoretical density $\mathcal{N}\left(0, 2\sigma^4\right)$. The correct density is the one close to the empirical density; the incorrect density is more peaked.

Figure 2: Histogram of Simulated Empirical pdf of $\hat{\sigma}^2$



We use MATLAB to simulate a time series of 500,000 observations of the Gaussian AR(1) using $\rho = 0.90$, $\sigma_\epsilon = 0.50$, and $\mu = 0$. We then record the sample variance $\hat{\sigma}^2$ of the process. We repeat this 10,000 times and plot (the uneven line) the realized density of $\hat{\sigma}^2$. We overlay on the plot the correct theoretical density $\mathcal{N}\left(\sigma^2, \frac{2\sigma^4(1+\rho^2)}{n(1-\rho^2)}\right)$ and the most common incorrect student-derived theoretical density $\mathcal{N}\left(\sigma^2, \frac{2\sigma^4}{n}\right)$. The correct density is the one close to the empirical density; the incorrect density is more peaked.