CHAPTER 10

Least-Squares Autoregression with Near-unit Root

Jan R. MAGNUS and Thomas J. ROTHENBERG[1]

10.1 Introduction

In autoregressive models, assumptions about the initial conditions seem to matter a lot when one of the roots of the characteristic polynomial is near the unit circle. In particular, the fixed start-up model (where the first few observations are treated asymmetrically) behaves quite differently in small samples than the stationary model which assumes that all the observations have the same distribution.

In this chapter, we explore some implications for the sampling distributions of least-squares estimates, test statistics, and forecasts in the leading case of the first-order autoregressive model. This is accomplished by studying the limiting behaviour of the statistics, for fixed sample size, as the autoregressive parameter approaches unity. When regressions are fitted *without* intercept, the stationary and fixed start-up models produce strikingly different sampling distributions for the least-squares estimator as the autoregressive parameter tends to one in absolute value. When regressions are fitted *with* intercept, the two models produce similar sampling properties when the autoregressive parameter tends to plus one, but very different properties when the parameter tends to minus one.

Previous Monte Carlo and numerical integration calculations have demonstrated that least-squares estimates, test statistics and forecasts in autoregressive time-series models exhibit unusual features when one of the roots of the characteristic polynomial lies near the unit circle. In particular, the traditional approximate distribution theory based on normality is often unsatisfactory. Furthermore, small differences in the initial conditions sometimes yield noticeable differences in the sampling distributions even when the number of observations is rather large. See, e.g., Phillips (1977, 1987), Evans and Savin (1981, 1984)

[1]The first version of this paper was completed in July 1988 when both authors were at the London School of Economics.

and Magnus and Pesaran (1989, 1991). In this chapter, we provide further evidence and offer some simple explanations for these phenomena in the leading case of the AR(1) model

$$y_t = \alpha + \beta y_{t-1} + \sigma u_t, \quad t = 1, 2, \ldots, \tag{1}$$

where $\{u_t\}$ is a sequence of independent standard normal random variables.

We shall examine some statistics obtained from a least-squares regression of y_t on its lagged value, using observations $y_0, \ldots, y_T$. Section 2 considers the case where α is known to be zero and the regression is performed without an intercept. Section 3 considers the case where α may be nonzero and the regression contains an intercept. The behaviour of the sampling distributions turns out to be rather different in the two cases when β is near one. The sampling distributions are also sensitive to the specification of the initial observation y_0. We shall assume that y_0 has the representation

$$y_0 = \mu + \delta u_0, \tag{2}$$

where u_0 is standard normal and independent of $\{u_1, u_2, \ldots\}$

Specification (2) covers a number of possibilities that have been discussed in the literature. If, for example, $\alpha = (1 - \beta)\mu$ and $\sigma^2 = (1 - \beta^2)\delta^2$, we have the stationary model; that is, the $\{y_t\}$ are a stationary stochastic process with expectation μ and variance δ^2. Alternatively, if $\delta = 0$ and $\mu = c$, the first observation y_0 is the constant c; Evans and Savin call this the "fixed start-up" model. If $\delta = \sigma, \mu = \alpha + \beta c$, and (1) also holds for $t = 0$, then the value of y preceding y_0 is the constant c; this is another version of the fixed start-up assumption, used by Fuller (1976) and by Magnus and Pesaran (1989, 1991). (Note that Fuller and Magnus and Pesaran assume that the sample consists of observations $y_1, \ldots, y_n$. Hence our T corresponds to their $n - 1$.)

Our approach will be to investigate the limiting distributions of the least-squares estimates, test statistics and forecasts, for fixed sample size, as β tends to one. Although some large-sample approximations will be employed to simplify the formulae, our main results are valid for all sample sizes. Indeed, one of our main purposes is to discover the small-sample properties of least-squares statistics when β is close to one. Asymptotic analysis of dynamic models when the autocorrelation coefficient tends to unity has been employed previously by Chipman (1979) and Jensen (1986). Our results can be viewed as a generalisation of those found by Jensen.

The model represented by (1) and (2) is characterised by the five parameters $\alpha, \beta, \sigma^2, \mu$ and δ. When examining sequences of models as β approaches one, we must specify the behaviour of the other parameters. For example, in the stationary case, α and σ^2 could be held fixed while $\mu = \alpha/(1 - \beta)$ and $\delta = \sigma/\sqrt{1 - \beta^2}$ both tend to infinity. Alternatively, μ and δ could be held fixed while α and σ tend to zero. In the nonstationary case, all sorts of dependency of α, σ^2, μ and δ on β could be examined. Luckily, not all of the possibilities lead to different results. In fact, only a few basic forms of limiting behaviour for the least-squares statistics turn out to be possible.

10.2 Regression without Intercept

When α is zero and the regression has no intercept, the least-squares estimates of β and σ^2 based on $y_0, y_1, \ldots, y_T$ can be written as

$$\hat{\beta} = \frac{\sum_t y_{t-1} y_t}{\sum_t y_{t-1}^2} = \beta + \sigma \frac{\sum_t y_{t-1} u_t}{\sum_t y_{t-1}^2}, \tag{3}$$

$$\hat{\sigma^2} = \frac{1}{T-1} \sum_t (y_t - \hat{\beta} y_{t-1})^2 = \frac{\sigma^2}{T-1} \sum_t u_t^2 - \frac{(\hat{\beta}-\beta)^2}{T-1} \sum_t y_{t-1}^2, \tag{4}$$

where $\sum_t$ represents summation from 1 to T. Also of interest are the t-ratio for β,

$$t_\beta = (\hat{\beta} - \beta) \left[\sum_t y_{t-1}^2 / \hat{\sigma^2} \right]^{1/2} = \frac{\sigma \sum_t y_{t-1} u_t}{\hat{\sigma} \left[\sum_t y_{t-1}^2 \right]^{1/2}}, \tag{5}$$

and the error in the s-period ahead forecast,

$$e_{T+s} = \hat{\beta}^s y_T - y_{T+s} = (\hat{\beta}^s - \beta^s) y_T - \sigma \sum_{i=0}^{s-1} \beta^i u_{T+s-i}. \tag{6}$$

When y_0 is specified as in (2), the difference equation (1) may be solved as

$$y_t = \beta^t (\mu + \delta u_0) + \sigma \sum_{i=0}^{t-1} \beta^i u_{t-i} \qquad (t \geq 0), \tag{7}$$

where the summation term is taken to be zero when $t = 0$. For any β, the behaviour of y_t (and hence that of $\hat{\beta}, \hat{\sigma^2}, t_\beta$ and e_{T+s} depends on the relative magnitudes of the three parameters μ, δ and σ. In fact, when considering the limit as β tends to one, there are three possible cases:

Case 1: As β approaches one, both μ/δ and σ/δ approach zero. (This occurs, for example, when $\mu = 0$ and $\sigma^2 = \delta^2(1-\beta^2)$ so that the y_t are stationary.) Then, as $\beta \to 1$,

$$y_t/\delta \to u_0 \quad \text{for all} \quad t \geq 0.$$

Case 2: As β approaches one, σ/μ approaches zero and δ/μ approaches the constant θ. (This occurs, for example, when μ and δ are nonzero constants and $\sigma^2 = \delta^2(1-\beta^2)$; the y_t are variance stationary but not mean stationary.) Then, as $\beta \to 1$,

$$y_t/\mu \to 1 + \theta u_0 \quad \text{for all} \quad t \geq 0.$$

Case 3: As β approaches one, μ/σ approaches the constant $\bar{\mu}$ and δ/σ approaches the constant $\bar{\delta}$. (This occurs, for example, in either form of the fixed start-up model.) Then, as $\beta \to 1$,

$$y_t/\sigma \to \bar{\mu} + \bar{\delta} u_0 + u_1 + \cdots + u_t.$$

Clearly, the limiting behaviour in case 3 (where y_t behaves like a random walk) is quite different from that in cases 1 and 2 (where y_t does not depend on t). This carries over to the distributions of the least-squares statistics. Taking limits in equations (3)–(6), we find the following extensions of a result given in Jensen (1986, p. 139).

Proposition 1. In case 1, as β tends to one, $\hat{\beta}$ converges in probability to one. Furthermore, for fixed T,

(a) $T^{1/2}(\hat{\beta}-\beta)\dfrac{\delta}{\sigma} \to \dfrac{T^{1/2}\bar{u}}{u_0} \sim$ Cauchy,

(b) $(T-1)\dfrac{\hat{\sigma}^2}{\sigma^2} \to \sum_t (u_t-\bar{u})^2 \sim \chi^2(T-1)$,

(c) $t_\beta \to \sqrt{T(T-1)}\dfrac{\bar{u}}{\sqrt{\sum_t (u_t-\bar{u})^2}} \sim$ Student $(T-1)$,

(d) $\dfrac{e_{T+s}}{\sigma} \to -(u_{T+1}+\cdots+u_{T+s})+s\bar{u} \sim N\Big(0, s+\dfrac{s^2}{T}\Big)$,

where $\bar{u} = T^{-1}\sum_t u_t$.

Proposition 2. In case 2, as β tends to one, $\hat{\beta}$ converges in probability to one. For fixed T,

$$T^{1/2}(\hat{\beta}-\beta)\frac{\mu}{\sigma} \to \frac{T^{1/2}\bar{u}}{1+\theta u_0}.$$

The statistics $(T-1)\hat{\sigma}^2/\sigma^2$, t_β, and e_{T+s}/σ approach the same random variables and therefore have the same limiting distributions as in Proposition 1.

The random walk case is considerably more complicated. Since y_t/σ remains, in the limit, a linear function of all the errors, the least-squares estimator converges to a ratio of two quadratic functions of the errors. The exact distributions of the least-squares statistics have no simple expression, but large sample representations in terms of a continuous time process are available; see Phillips (1987). We have the following result comparable to those above.

Proposition 3. In case 3, as β tends to one, $\hat{\beta}$ converges in probability to a nondegenerate random variable. The limiting distributions (as $\beta \to 1$) of the least-squares statistics have the large sample approximations:

(a) $T(\hat{\beta}-\beta) \to \dfrac{\int B dB}{\int B^2} + \mathcal{O}(T^{-1/2})$,

(b) $(T-1)\dfrac{\hat{\sigma}^2}{\sigma^2} \to \sum_t (u_t-\bar{u})^2 + \mathcal{O}(T^0)$,

(c) $t_\beta \to \dfrac{\int BdB}{\left[\int B^2\right]^{1/2}} + \mathcal{O}(T^{-1/2}),$

(d) $\dfrac{e_{T+s}}{\sigma} \to -(u_{T+1} + \cdots + u_{T+s}) + \dfrac{s}{\sqrt{T}} \dfrac{B(1) \int BdB}{\int B^2} + \mathcal{O}(T^{-1}),$

where $B(r)$ is a Wiener process on the unit interval given by limit (as $T \to \infty$) of $(u_1 + \cdots + u_{[rT]})/\sqrt{T}$ and $[rT]$ denotes the integer part of rT.

The assumption about the initial value y_0 clearly matters a lot in determining the limiting behaviour, as $\beta \to 1$, of the least-squares estimator $\hat{\beta}$. In the stationary case, the standardised statistic $\sqrt{T}(\hat{\beta}-\beta)/\sqrt{1-\beta^2}$ tends to a Cauchy variate; in the fixed start-up case, this statistic's dispersion grows without bound as β tends to one. Or, put another way, under stationarity $\hat{\beta}$ converges rapidly to one for any sample size, whereas, under fixed start-up, $\hat{\beta}$ converges to a random variable with considerable dispersion unless T is large.

At first glance, this finding seems to be at variance with previous Monte Carlo results. Phillips (1977), for example, reports that $\hat{\beta}$ has a highly skewed distribution in the stationary case when β is large. Yet, as $\beta \to 1$, the limiting distribution is symmetric about 1. This suggests that the rate of convergence of $\sqrt{T}(\hat{\beta} - \beta)\delta/\sigma$ in Proposition 1 is very slow. To investigate this possibility, we calculated the percentiles of $\sqrt{T}(\hat{\beta} - \beta)/\sqrt{1-\beta^2}$ for various values of β under the stationary assumption that $\mu = 0$ and $\sigma^2 = \delta^2(1-\beta^2)$. The value for T was set at 14 and the Imhof routine was employed. The results are presented in the top panel of Table 1. The middle panel gives the percentiles of the variable $\hat{\beta} - \beta$. The lower panel gives the percentiles of the distribution of $\hat{\beta} - \beta$ under the fixed start-up assumption that $\mu = 0$ and $\delta = \sigma$. The last two columns give the interquartile range (IQR) and a skewness measure

$$\mathrm{SK} = \frac{x_{.75} - x_{.50}}{x_{.50} - x_{.25}},$$

where x_α is the α percentile point. SK will be one for a symmetric distribution.

It is clear from the top panel that the rate of convergence to Cauchy is extremely slow. Although the standardisation factor δ/σ seems to stabilise the dispersion, the median and skewness are always far from the limiting values. Indeed, the largest departure for the median occurs at $\beta = 0.95$ and the largest departure for the skewness occurs at $\beta = 0.99$. The limiting distribution gives little indication of the shape of the true sampling distribution even for β values as large as 0.999. On the other hand, a comparison of the middle and bottom panels suggests that the distribution of $\hat{\beta}$ in the fixed start-up model is a good approximation to the distribution of $\hat{\beta}$ in the stationary model for $\beta = 0.5$ but is not very good for β greater than 0.9. This conclusion agrees well and, to some extent, explains the exact results on the mean-square forecast error obtained by Hoque, Magnus and Pesaran (1988).

10.3 Regression with Intercept

When the regression contains an intercept, the relevant least-squares statistics based on $y_0, y_1, \ldots, y_T$ can be expressed as

$$\tilde{\beta} = \beta + \sigma \frac{\sum_t (y_{t-1} - \bar{y}) u_t}{\sum_t (y_{t-1} - \bar{y})^2}, \tag{8}$$

$$\tilde{\sigma}^2 = \frac{\sigma^2}{T-2} \sum_t (u_t - \bar{u})^2 - (\tilde{\beta} - \beta)^2 \frac{1}{T-2} \sum_t (y_{t-1} - \bar{y})^2, \tag{9}$$

$$\tilde{t}_\beta = \frac{\sigma \sum_t (y_{t-1} - \bar{y}) u_t}{\tilde{\sigma} \left[\sum_t (y_{t-1} - \bar{y})^2\right]^{1/2}}, \tag{10}$$

$$\begin{aligned} \tilde{e}_{T+s} &= -\sigma \sum_{i=0}^{s-1} \beta^i u_{T+s-i} + \sigma \bar{u} \frac{1-\beta^s}{1-\beta} \\ &\quad + (\tilde{\beta}^s - \beta^s)(y_T - \bar{y}) + \left[\frac{1-\tilde{\beta}^s}{1-\tilde{\beta}} - \frac{1-\beta^s}{1-\beta}\right] \frac{y_T - y_0}{T}, \end{aligned} \tag{11}$$

where $\bar{y} = T^{-1} \sum_t y_{t-1}$. (We shall not examine the least-squares estimator or t-statistic for α, whose behaviour is somewhat different.) The analysis is similar to that conducted in the previous section with $y_t - \bar{y}$ replacing y_t. By successive substitution in (1), we find

$$y_t = \alpha \frac{1-\beta^t}{1-\beta} + \beta^t(\mu + \delta u_0) + \sigma \sum_{i=0}^{t-1} \beta^i u_{t-i} \quad (t \geq 0)$$

and

$$\bar{y} = \frac{\alpha}{1-\beta} + \frac{1}{T} \frac{1-\beta^T}{1-\beta} \left(\mu + \delta u_0 - \frac{\alpha}{1-\beta}\right) + \frac{\sigma}{T} \sum_{j=1}^{T-1} \frac{1-\beta^{T-j}}{1-\beta} u_j,$$

and hence, for $t \geq 0$,

$$\begin{aligned} y_t - \bar{y} &= \frac{\beta^t - T^{-1}(1-\beta^T)/(1-\beta)}{1-\beta} \left[\delta(1-\beta)u_0 - \{\alpha - \mu(1-\beta)\}\right] \\ &\quad + \sigma \sum_{j=1}^{t} \beta^{t-j} u_j - \frac{\sigma}{T} \sum_{j=1}^{T-1} \frac{1-\beta^{T-j}}{1-\beta} u_j. \end{aligned} \tag{12}$$

The behaviour of $y_t - \bar{y}$ is rather different from that of y_t when β is near one. Now everything depends on the relative magnitudes of $\alpha - \mu(1-\beta)$, $\delta(1-\beta)$, and σ. Only one case seems to be of any interest (and covers both the fixed start-up and stationary models).

Case 4: As β approaches one, $\delta(1-\beta)/\sigma$ approaches zero and $[\alpha - \mu(1-\beta)]/\sigma$ approaches the constant λ. Then, as $\beta \rightarrow 1$,

$$\frac{y_t - \bar{y}}{\sigma} \rightarrow \lambda\Big(t - \frac{T-1}{2}\Big) + \sum_{j=1}^{t} u_j - \sum_{j=1}^{T-1}\Big(1 - \frac{j}{T}\Big)u_j. \tag{13}$$

Again, when T is large, the weighted sum of errors can be approximated by a functional on a Wiener process. We have, as β tends to one and $t = rT$,

$$\frac{y_t - \bar{y}}{\sigma\sqrt{T}} \rightarrow \lambda\Big(r - \frac{1}{2}\Big)\sqrt{T} + B(r) - \int B + \mathcal{O}(T^{-1/2}). \tag{14}$$

The sampling distributions of the least-squares statistics depend crucially on λ. If the y_t process is mean stationary (i.e., Ey_t does not depend on t), then λ is necessarily zero. Likewise, if $\alpha = 0$ and μ does not depend on β, then λ will be zero. In these cases, we have

Proposition 4. If, as $\beta \rightarrow 1, \delta(1-\beta)/\sigma$ and $[\alpha - \mu(1-\beta)]/\sigma$ both tend to zero, then $\tilde{\beta}$ converges in probability to a nondegenerate random variable. The limiting distributions (as $\beta \rightarrow 1$) of the least-squares statistics have the large sample approximations

(a) $T(\tilde{\beta} - \beta) \rightarrow \dfrac{\int BdB \; - \; B(1)\int B}{\int B^2 - (\int B)^2} + \mathcal{O}(T^{-1/2}),$

(b) $(T-2)\dfrac{\tilde{\sigma}^2}{\sigma^2} \rightarrow \chi^2(T-2) + \mathcal{O}(T^0),$

(c) $\tilde{t}_\beta \rightarrow \dfrac{\int BdB - B(1)\int B}{\left[\int B^2 - (\int B)^2\right]^{1/2}} + \mathcal{O}(T^{-1/2}),$

(d) $$\frac{\tilde{e}_{T+s}}{\sigma} \rightarrow -(u_{T+1} + \cdots + u_{T+s}) + \frac{s}{\sqrt{T}}B(1) + \frac{s}{\sqrt{T}}\frac{\int BdB - B(1)\int B}{\int B^2 - (\int B)^2}\left\{B(1) - \int B\right\} + \mathcal{O}(T^{-1}).$$

If the $\{y_t\}$ are not mean stationary, then λ will generally be nonzero. In this case y_t behaves like a time trend (plus noise) when β is near one. We have

Proposition 5. If, as $\beta \rightarrow 1$, $\delta(1-\beta)/\sigma$ tends to zero but $[\alpha - \mu(1-\beta)]/\sigma$ tends to the nonzero constant λ, then $\tilde{\beta}$ converges in probability to a nondegenerate random variable. The limiting distributions (as $\beta \rightarrow 1$) of the least-squares statistics have the large sample approximations

(a) $\dfrac{\lambda}{\sqrt{12}}T\sqrt{T}(\tilde{\beta} - \beta) \rightarrow N(0,1) + \mathcal{O}(T^{-1/2}),$

(b) $(T-2)\dfrac{\tilde{\sigma}^2}{\sigma^2} \to \chi^2(T-2) + \mathcal{O}(T^{-1/2})$,

(c) $\tilde{t}_\beta \to$ Student $(T-2) + \mathcal{O}(T^{-1/2})$,

(d) $\dfrac{\tilde{e}_{T+s}}{\sigma} \to N\Big(0, s + \dfrac{4s^2}{T}\Big) + \mathcal{O}(T^{-1})$.

These results indicate that, when β is near one, regression with an intercept is very different from regression without an intercept. In the former case, the limiting behaviour of $\tilde{\beta}$ is quite sensitive to assumptions on μ, but not sensitive to assumptions on δ. In the latter case, the opposite holds. In regression with intercept but $\alpha = 0$, the fixed start-up and the stationary models produce the same limiting distributions. In regression without intercept, the two models produce very different limiting distributions.

Some percentiles for the distribution of $\tilde{\beta} - \beta$ when T=14 are given in Table 2. The top panel is for the stationary model where $\alpha = \mu(1-\beta)$ and $\sigma^2 = \delta^2(1-\beta^2)$. The bottom panel is for the fixed start-up model where $\mu = \alpha = 0$ and $\delta = \sigma$. The two distributions are very similar, although not identical, when β is close but not equal to one. For example, when $\beta = 0.90$, the dispersion and skewness of the two distributions are close, but the medians differ by about 8%. This agrees with the results obtained by Evans and Savin (1984, p. 1265) and Magnus and Pesaran (1989, 1991).

10.4 Negative Unit Root

Although perhaps of less practical importance, the situation where β is near minus one also produces some unusual sampling behaviour. In regression without intercept, a symmetry argument employed by Fuller (1976) can be exploited. With $x_t = (-1)^t y_t, v_t = (-1)^t u_t$ and $\alpha = 0$, the model (1)–S(2) can be written as

$$x_t = \gamma x_{t-1} + \omega v_t, \tag{15}$$

$$x_0 = \mu + \delta v_0, \tag{16}$$

where $\gamma = -\beta$ and $\omega = \sigma$. Since the v_t are independent standard normal variables, the results of Section 2 apply to this new model for γ tending to one. Furthermore, a regression of x_t on x_{t-1} produces statistics $\hat{\gamma} = -\hat{\beta}, t_\gamma = -t_\beta, \hat{\omega}^2 = \hat{\sigma}^2$, and $\hat{x}_{T+s} - x_{T+s} = (-1)^{T+s}(\hat{y}_{T+s} - y_{T+s})$. Thus we have

Proposition 6. The limiting distributions of $-(\hat{\beta}-\beta), -t_\beta, \hat{\sigma}^2$, and $(-1)^{T+s}e_{T+s}/\sigma$ as $\beta \to -1$ are identical to those given in Propositions 1-3 (where the three cases now describe the behaviour of the parameters as $\beta \to -1$ and the u's are replaced with the v's.)

Regression with an intercept, however, does not demonstrate the same symmetries. From (12), the behaviour of $y_t - \bar{y}$ as β tends to minus one depends

Table 2: Distributions for regression with intercept in the model $y_t = \alpha + \beta y_{t-1} + \sigma u_t (t = 1, \ldots, T), y_0 = \delta u_0$ for $T = 14$.

β	2.5%	10%	25%	50%	75%	90%	97.5%	IQR	SK
Percentage points of $\tilde{\beta} - \beta, \alpha = 0, \delta = \sigma/\sqrt{1-\beta^2}$									
0.50	−0.717	−0.513	−0.334	−0.151	0.011	0.136	0.255	0.346	0.879
0.80	−0.809	−0.579	−0.391	−0.213	−0.071	0.033	0.143	0.320	0.796
0.90	−0.844	−0.607	−0.417	−0.242	−0.106	−0.003	0.104	0.311	0.776
0.95	−0.864	−0.624	−0.433	−0.259	−0.125	−0.023	0.082	0.308	0.774
0.99	−0.882	−0.639	−0.447	−0.274	−0.140	−0.040	0.061	0.307	0.774
1.00	−0.887	−0.644	−0.451	−0.278	−0.144	−0.045	0.056	0.307	0.774
Percentage points of $\tilde{\beta} - \beta,\ \alpha = 0, \delta = \sigma$									
0.50	−0.723	−0.519	−0.339	−0.154	0.010	0.135	0.256	0.349	0.880
0.80	−0.831	−0.601	−0.410	−0.228	−0.082	0.024	0.135	0.329	0.801
0.90	−0.867	−0.631	−0.441	−0.263	−0.124	−0.020	0.090	0.316	0.779
0.95	−0.881	−0.642	−0.451	−0.277	−0.142	−0.038	0.070	0.310	0.774
0.99	−0.887	−0.645	−0.453	−0.280	−0.146	−0.045	0.058	0.307	0.775
1.00	−0.887	−0.644	−0.451	−0.278	−0.144	−0.045	0.056	0.307	0.774

on the relative magnitudes of $\alpha - \mu(1-\beta), \delta$, and σ. As noted by Fuller (1976), the results are quite different from the situation where β tends to plus one. In fact, the results are close to those for regression without intercept. Again three cases can be distinguished.

Proposition 7. Suppose both σ/δ and $[\alpha - \mu(1-\beta)]/\delta$ tend to zero as β approaches minus one; e.g., the y_t are stationary. Then, as $\beta \to -1$,

$$\frac{y_t - \bar{y}}{\delta} \to \begin{cases} (-1)^t u_0 & (T \text{ even}) \\ \left\{(-1)^t - \frac{1}{T}\right\} u_0 & (T \text{ odd}) \end{cases}$$

and $\tilde{\beta}$ converges in probability to minus one. Furthermore, for fixed T,

(a) $\varphi_T^{1/2}(\tilde{\beta} - \beta)\frac{\delta}{\sigma} \to$ Cauchy,

(b) $(T-2)\frac{\tilde{\sigma}^2}{\sigma^2} \to \chi^2(T-2)$,

(c) $\tilde{t}_\beta \to$ Student $(T-2)$,

(d) $\frac{\tilde{e}_{T+s}}{\sigma} \to N(0, \omega_{T,s}^2)$,

where

$$\varphi_T = \begin{cases} T & (T \text{ even}) \\ T - \frac{1}{T} & (T \text{ odd}) \end{cases}$$

and

$$\omega_{T,s}^2 = \begin{cases} s + \dfrac{s^2}{T} & (T \text{ even}, s \text{ even}) \\ s + \dfrac{s^2+1}{T} & (T \text{ even}, s \text{ odd}) \\ s + \dfrac{s^2(T+s)^2}{T(T^2-1)} & (T \text{ odd}, s \text{ even}) \\ s + \dfrac{s^2(T+s)^2}{T(T^2-1)} + \dfrac{1}{T} & (T \text{ odd}, s \text{ odd}). \end{cases}$$

Proposition 8. Let $\mu_* = -\frac{1}{2}\left[\alpha - \mu(1-\beta)\right]$ and suppose σ/μ_* tends to zero and δ/μ_* tends to the constant $\bar{\theta}$ as β approaches minus one; e.g. the y_t are variance stationary, but not mean stationary. Then, as $\beta \to -1$,

$$\frac{y_t - \bar{y}}{\mu_*} \to \begin{cases} (-1)^t(1+\bar{\theta}u_0) & (T \text{ even}) \\ \{(-1)^t - T^{-1}\}(1+\bar{\theta}u_0) & (T \text{ odd}), \end{cases}$$

and $\tilde{\beta}$ converges in probability to minus one. Furthermore, for fixed T,

$$\varphi_T^{1/2}(\tilde{\beta} - \beta)\frac{\mu_*}{\sigma} \to \frac{z}{1+\bar{\theta}u_0},$$

where φ_T is defined in Proposition 7 and $z \sim N(0,1)$ is independent of u_0; the remaining statistics behave as in Proposition 7.

Proposition 9. Suppose $\delta/\sigma \to \bar{\delta}$ and $\left[\alpha - \mu(1-\beta)\right]/\sigma \to \lambda$ as β approaches minus one; e.g., the starting value y_0 is fixed. Then, as $\beta \to -1$,

$$\frac{y_t - \bar{y}}{\sigma} \to (-1)^t\Big[\bar{\delta}u_0 - \frac{1}{2}\lambda + \sum_{j=1}^{t} v_j\Big] + \frac{1}{T}S_T,$$

where

$$S_T = \begin{cases} v_1 + v_3 + \cdots + v_{T-1} & (T \text{ even}) \\ -\left[\bar{\delta}v_0 - \frac{1}{2}\lambda + v_2 + v_4 + \cdots + v_{T-1}\right] & (T \text{ odd}). \end{cases}$$

If, for large T, $(v_1 + \cdots + v_t)/\sqrt{T}$ is approximated by the Wiener process $\bar{B}(r)$ where $r = t/T$, then

(a) $T(\tilde{\beta} - \beta) \to -\dfrac{\int \bar{B} d\bar{B}}{\int \bar{B}^2} + \mathcal{O}(T^{-1/2})$,

(b) $(T-2)\dfrac{\tilde{\sigma}^2}{\sigma^2} \to \chi^2(T-2) + \mathcal{O}(T^0)$,

(c) $\tilde{t}_\beta \to -\dfrac{\int \bar{B} d\bar{B}}{\left[\int \bar{B}^2\right]^{1/2}} + \mathcal{O}(T^{-1/2})$,

(d) $\frac{\tilde{e}_{T+s}}{\sigma} \to (-1)^{T+s-1}\left[v_{T+1} + \cdots + v_{T+s}\right]$

$$+ (-1)^{T-1}\frac{s}{\sqrt{T}}\frac{\int \bar{B}d\bar{B}}{\int \bar{B}^2}\bar{B}(1) + \frac{B(1)}{\sqrt{T}} + \mathcal{O}(T^{-1})$$

for s odd; the term $B(1)/\sqrt{T}$ is dropped when s is even.

10.5 Conclusions

Assumptions about the initial observation can matter a lot in the AR(1) model with near-unit root. When regressions are fitted without intercept, the stationary and fixed start-up models produce very different sampling properties when the autoregression parameter is near plus or minus one. When regressions are fitted with intercept, the two models have similar sampling properties when the autoregression parameter is near plus one, but very different properties when the parameter is near minus one.

As noted by Jensen (1986), the least-squares statistics in regression without intercept have simple limiting distributions in the stationary model as the autoregression parameter tends to unity. Unfortunately, those limiting distributions do not seem to provide useful approximations for any given stationary model.

Appendix: Derivations of Propositions

Proposition 1. Using the fact that $y_t/\delta \to u_0$, we find

$$\sum y_{t-1}u_t/\delta \to u_0 \sum u_t \tag{17}$$

and

$$\sum y_{t-1}^2/\delta^2 \to Tu_0^2. \tag{18}$$

This, together with (3) proves (a). Then (b) follows from (4) and the fact that

$$\left[\frac{\hat{\beta}-\beta}{\sigma}\right]^2 \sum y_{t-1}^2 \to T(\bar{u})^2.$$

Next, (c) follows from (5), (17), (18) and (b). To prove (d), we use a slight generalisation of (a), proved by a simple Taylor expansion, namely

$$T^{1/2}\left[\frac{\hat{\beta}^s - \beta^s}{s}\right]\frac{\delta}{\sigma} \to \frac{T^{1/2}\bar{u}}{u_0} \sim \text{ Cauchy.} \tag{19}$$

Using (19), we obtain

$$(\hat{\beta}^s - \beta^s)y_T/\sigma \to s\bar{u} \sim N\left(0, \frac{s^2}{T}\right)$$

and hence (d).

Proposition 2. Similar to the derivation of Proposition 1.

Proposition 3. We have, as $\beta \to 1$,

$$y_t/\sigma \to \bar{\mu} + \bar{\delta} u_0 + T^{1/2} w_t,$$

where

$$w_t = T^{-1/2}(u_1 + u_2 + \cdots + u_t), \qquad w_0 = 0. \tag{20}$$

We shall employ the following approximations for large T; see Phillips (1987, p. 294):

$$T^{-1} \sum w_{t-1}^2 = \int_0^1 B(r)^2 dr + \mathcal{O}(T^{-1}), \tag{21}$$

$$\sum w_{t-1}(w_t - w_{t-1}) = \int_0^1 B(r) dB(r) + \mathcal{O}(T^{-1/2}). \tag{22}$$

In addition, we note that $w_T = B(1) \equiv N(0,1)$. Then, as $\beta \to 1$,

$$\begin{aligned} T^{-1} \sum y_{t-1} u_t/\sigma &\to \sum w_{t-1}(w_t - w_{t-1}) + T^{-1/2}(\bar{\mu} + \bar{\delta} u_0) w_T \\ &= \int B dB + \mathcal{O}(T^{-1/2}) \end{aligned}$$

and

$$\begin{aligned} T^{-2} \sum y_{t-1}^2/\sigma^2 &\to T^{-1} \sum w_{t-1}^2 + 2T^{-1/2}(\bar{\mu} + \bar{\delta} u_0)\Big(T^{-1} \sum w_{t-1}\Big) \\ &\quad + T^{-1}(\bar{\mu} + \bar{\delta} u_0)^2 \\ &= \int B^2 + \mathcal{O}(T^{-\frac{1}{2}}), \end{aligned}$$

and, hence, using (3)–(6), the proposition follows.

Proposition 4. Since $\lambda = 0$, we have from (13), as $\beta \to 1$,

$$\frac{y_t - \bar{y}}{\sigma \sqrt{T}} \to w_t - T^{-1} \sum w_{t-1}$$

where w_t is defined in (20). Hence

$$\begin{aligned} T^{-1} \sum (y_{t-1} - \bar{y}) u_t/\sigma &\to \sum w_{t-1}(w_t - w_{t-1}) - w_T\Big(T^{-1} \sum w_{t-1}\Big) \\ &= \int B dB - B(1) \int B + \mathcal{O}(T^{-1/2}), \\ T^{-2} \sum (y_{t-1} - \bar{y})^2/\sigma^2 &\to T^{-1} \sum w_{t-1}^2 - \Big(T^{-1} \sum w_{t-1}\Big)^2 \\ &= \int B^2 - \Big(\int B\Big)^2 + \mathcal{O}(T^{-1}), \end{aligned}$$

where, in addition to (21) and (22), we have used

$$T^{-1}\sum w_{t-1} = \int_0^1 B(r)dr + \mathcal{O}(T^{-1}). \qquad (23)$$

The results now follow from (8)–(11).

Proposition 5. We define

$$c_t = \left[12T^{-3}\right]^{1/2}\left(t - \frac{T+1}{2}\right) \qquad (t = 1, 2, \ldots, T),$$

so that $\sum c_t = 0$ and $\sum c_t^2 = 1 + \mathcal{O}(T^{-2})$. Then, letting $\gamma = \lambda/\sqrt{12}$, we obtain from (13), as $\beta \to 1$,

$$\frac{y_{t-1} - \bar{y}}{\sigma\sqrt{T}} \to \gamma T c_t + w_{t-1} - \frac{1}{T}\sum w_{t-1}$$

and hence

$$\begin{aligned}
T^{-1}\sum \frac{y_{t-1} - \bar{y}}{\sigma\sqrt{T}} u_t &\to \gamma z_1 + T^{-1/2}Q, \\
T^{-2}\sum \left[\frac{y_{t-1} - \bar{y}}{\sigma\sqrt{T}}\right]^2 &\to \gamma^2 + 2\gamma T^{-1/2} z_2 + \mathcal{O}(T^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
z_1 &= \sum c_t u_t \sim N(0,1) + \mathcal{O}(T^{-2}), \\
z_2 &= T^{-1/2}\sum c_t w_{t-1} \sim N\Big(0, \frac{1}{10}\Big) + \mathcal{O}(T^{-1}), \\
Q &= \sum w_{t-1}(w_t - w_{t-1}) - w_T\Big(T^{-1}\sum w_{t-1}\Big).
\end{aligned}$$

Using these results, we obtain, from (8)–(11),

$$\begin{aligned}
\gamma T^{3/2}(\tilde{\beta} - \beta) &\to z_1 + T^{-1/2}\frac{Q - 2z_1 z_2}{\gamma} + \mathcal{O}(T^{-1}), \\
\frac{(T-2)\tilde{\sigma}^2}{\sigma^2} &\to \sum (u_t - \bar{u})^2 - z_1^2 - 2T^{-1/2} z_1 (Q - z_1 z_2)/\gamma + \mathcal{O}(T^{-1}), \\
\tilde{t}_\beta &\to \left[\frac{\sum (u_t - \bar{u})^2 - z_1^2}{T-2}\right]^{-1/2} z_1 + \mathcal{O}(T^{-1/2}), \\
\frac{\tilde{e}_{T+s}}{\sigma} &\to -(u_{T+1} + \cdots + u_{T+s}) + \frac{s}{\sqrt{T}}(w_T + z_1\sqrt{3}) \\
&+ \frac{s}{\gamma T}\left[\sqrt{3}(Q - 2z_1 z_2) + z_1\big(w_T - T^{-1}\sum w_{t-1}\big)\right] + \mathcal{O}(T^{-3/2}).
\end{aligned}$$

The results of Proposition 5 then follow. In fact, we have proved a bit more than Proposition 5. For example, noting that

$$EQ = -\frac{1}{2} + \mathcal{O}(T^{-1}), \qquad Ez_1 z_2 = \mathcal{O}(T^{-1}),$$

$$Ez_1 w_T = 0, \qquad Ez_1(T^{-1}\sum w_{t-1}) = -\frac{1}{6}\sqrt{3} + \mathcal{O}(T^{-2}),$$

we obtain the following expectations:

$$T^2 E(\tilde{\beta} - \beta) \rightarrow -\frac{6}{\lambda^2} + \mathcal{O}(T^{-1/2}), \tag{24}$$

$$E\frac{\tilde{\sigma}^2}{\sigma^2} \rightarrow 1 + \mathcal{O}(T^{-2}), \tag{25}$$

$$TE\frac{\tilde{e}_{T+s}}{\sigma} \rightarrow \frac{-2s}{\lambda} + \mathcal{O}(T^{-1/2}). \tag{26}$$

Proposition 6. See text.

Proposition 7. From (12), we have

$$(y_t - \bar{y})/\delta \rightarrow \bar{k}_t u_0,$$

where

$$\bar{k}_t = \begin{cases} (-1)^t & (T \text{ even}) \\ (-1)^t - 1/T & (T \text{ odd}). \end{cases}$$

Since $\sum \bar{k}_{t-1}^2 = \varphi_T$ and defining $k_t = \varphi_T^{-1/2}\bar{k}_t$, we obtain

$$\sum (y_{t-1} - \bar{y})u_t/\delta \rightarrow u_0 \varphi_T^{1/2} \sum k_{t-1} u_t$$

and

$$\sum (y_{t-1} - \bar{y})^2/\delta^2 \rightarrow u_0^2 \varphi_T.$$

Using (8), we then find

$$\varphi_T^{1/2}(\tilde{\beta} - \beta)\delta/\sigma \rightarrow \frac{\sum k_{t-1}u_t}{u_0} \sim \text{ Cauchy}.$$

Next, since

$$(\tilde{\beta} - \beta)^2 \sum (y_{t-1} - \bar{y})^2/\sigma^2 \rightarrow \left(\sum k_{t-1}u_t\right)^2 \sim \chi^2(1),$$

and defining

$$M = I_T - ii'/T, \qquad i = (1, 1, \ldots, 1)',$$

$$k = (k_0, k_1, \ldots, k_{T-1})' \qquad u = (u_1, u_2, \ldots, u_T)',$$

we obtain, using (9),

$$(T-2)\tilde{\sigma}^2/\sigma^2 \to \sum(u_t-\bar{u})^2 - \left(\sum k_{t-1}u_t\right)^2 = u'(M-kk')u \sim \chi^2(T-2),$$

since $Mk = k$. Also, using (10),

$$\tilde{t}_\beta \to \left[\frac{u'(M-kk')u}{T-2}\right]^{-1/2} \sum k_{t-1}u_t \sim \text{ Student } (T-2),$$

since $(M-kk')k = 0$. Finally, using (11),

$$\tilde{e}_{T+s}/\sigma \to z_1 + z_2 + z_3,$$

where z_1, z_2 and z_3 are independent random variables defined by

$$z_1 = -\sum_{i=0}^{s-1}(-1)^i u_{T+s-i} \sim N(0,s),$$

$$z_2 = \begin{cases} 0 & (s \text{ even}) \\ \bar{u} \sim N(0,1/T) & (s \text{ odd}), \end{cases}$$

$$z_3 = \begin{cases} T^{-1/2}s\sum k_{t-1}u_t \sim N(0,s^2/T) & (T \text{ even}) \\ -[T(T^2-1)]^{-1/2}s(s+T)\sum k_{t-1}u_t \sim N\left[0, \dfrac{s^2(s+T)^2}{T(T^2-1)}\right] & (T \text{ odd}). \end{cases}$$

The result follows.

Proposition 8. Similar to the derivation of Proposition 7.

Proposition 9. We have, as $\beta \to 1$, from (12):

$$(y_t-\bar{y})/\sigma \to (-1)^t T^{1/2}\bar{w}_t + (-1)^t(\bar{\delta}u_0 - \frac{1}{2}\lambda) + T^{-1/2}(T^{-1/2}S_T),$$

where

$$\bar{w}_t = T^{-1/2}(v_1 + v_2 + \cdots + v_t).$$

Hence,

$$T^{-1}\sum(y_{t-1}-\bar{y})u_t/\sigma \to -\int \bar{B}d\bar{B} + \mathcal{O}(T^{-1/2})$$

and

$$T^{-2}\sum(y_{t-1}-\bar{y})^2/\sigma^2 \to \int \bar{B}^2 + \mathcal{O}(T^{-1/2}).$$

The results then follow using (8)–(11).

References

[1] Chipman, J.S., (1979), Efficiency of Least-squares Estimation of Linear Trend when Residuals are Autocorrelated, *Econometrica,* 47, 115–128.

[2] Evans, G.B.A., and N.E. Savin, (1981), Testing for Unit Roots: 1, *Econometrica,* 49, 753–779.

[3] Evans, G.B.A., and N.E. Savin, (1984), Testing for Unit Roots: 2, *Econometrica,* 52, 1241–1269.

[4] Fuller, W.A., (1976), *Introduction to Statistical Time Series*, John Wiley and Sons, New York.

[5] Hoque, A., J.R. Magnus, and B. Pesaran, (1988), The Exact Multi-period Mean-square Forecast Error for the First-order Autoregressive Model, *Journal of Econometrics,* 39, 327–346.

[6] Jensen, J.L., (1986), *Asymptotic Expansions for Sums of Dependent Variables*, Memoirs No. 10, Department of Theoretical Statistics, University of Aarhus, Denmark.

[7] Magnus, J.R., and B. Pesaran, (1989), The Exact Multiperiod Mean-square Forecast Error for the First-order Autoregressive Model with an Intercept, *Journal of Econometrics*, 42, 157–179.

[8] Magnus, J.R., and B. Pesaran, (1991), The Bias of Forecasts from a First-order Autoregression, *Econometric Theory,* 7, 222–235.

[9] Phillips, P.C.B., (1977), Approximations to Some Finite Sample Distributions Associated with a First-order Stochastic Difference Equation, *Econometrica*, 45, 463–485.

[10] Phillips, P.C.B., (1987), Time Series Regression with a Unit Root, *Econometrica,* 55, 277–301.