The Unit-Root Revolution Revisited: Where Do Non-Standard Sampling Distributions and Related Statistical Conundrums Originate In?

Aris Spanos* <aris@vt.edu>

Department of Economics, Virginia Tech

October 2024

Abstract

The primary objective of the paper is twofold. *First*, to answer the question posed in the title by arguing that the conundrums: [C1] the non-standard sampling distributions, [C2] the low power of unit-root tests, [C3] the two competing parametrizations of the AR(1) models, (B) $Y_t{=}\alpha_0{+}\alpha_1 Y_{t-1}{+}\varepsilon_t$, (C) $Y_t{=}\alpha_0{+}\gamma t{+}\alpha_1 Y_{t-1}{+}\varepsilon_t$, and [C4] the incompleteness and non-testability of the error term ε_t assumptions, originate in viewing such models as stochastic difference equations where one aPriori Postulates (aPP): $\langle\mathbf{a}\rangle$ their functional form, $\langle\mathbf{b}\rangle$ the range of values of their unknown parameters, and $\langle\mathbf{c}\rangle$ their error term ε_t assumptions. The key culprit behind [C1] and [C2] is $\alpha_1{\in}[-1,1]$ in $\langle\mathbf{b}\rangle$ entailing $\langle\mathbf{d}\rangle$ the AR(1) nests the Unit Root [UR(1)] model for $\alpha_1{=}1$, which belies Kolmogorov's existence theorem. *Second*, to bring out explicitly the well-defined (existence theorem holds) statistical models corresponding to the AR(1): (B)-(C) models using R.A. Fisher's statistical perspective. This perspective reveals that there is no well-defined stochastic process $\{Y_t,\ t{\in}\mathbb{N}{:=}(1,2,...)\}$ underlying both the AR(1) and the related UR(1) model. Instead, the statistical AR(1) and UR(1) models are (i) grounded on two distinct processes $\{Y_t,\ t{\in}\mathbb{N}\}$, based on (ii) different probabilistic assumptions and (iii) statistical parametrizations, (iv) rendering them *non-nested*, and (v) their respective likelihood-based inferential components are free from conundrums [C1]-[C4]. These 'outlandish' claims are affirmed by analytical derivations, simulations, as well as proposing a non-stationary AR(1) model that nests the related UR(1) model, where testing $\alpha_1{=}1$ relies on likelihood-based tests that are free from conundrums [C1]-[C2].

KEYWORDS: Autoregressive models; stochastic difference equations; Dickey-Fuller unit root testing; non-standard sampling distributions; Brownian motion process; cointegration; statistical vs. apriori postulated (aPP) perspective/model; nested vs. non-nested models; statistical adequacy.

JEL codes: C12, C18, C22, C32, C51, C52.

*Thanks are due to Anya McGuirk for invaluable help with the simulations and numerous discussions on these issues, and to Nikitas Pittis and Rick Ashley for many helpful comments, as well as Bent Nielsen for discussions that helped me fathom the counterpoint on the statistical perspective.

1 Introduction

1.1 Statistical analysis of time series in the 20th century

The statistical analysis of time series data has undergone several important paradigm shifts during the 20th century, changing greatly how such data are modeled and used for inference purposes; see Mills (2011).

The first quarter of the 20th century was dominated by descriptive statistics in the form of summarizing key features of the data such as means, variances, correlations, autocorrelations, histograms, periodograms as well as attempts to quantify or eliminate trends and cycles exhibited by data; see Yule and Kendall (1950).

The first model-based approach, beyond descriptive statistics, was initiated by Yule (1927) putting forward the AutoRegressive of order p [AR(p)] model:

$$\text{AR(p)}:\quad Y_t=\alpha_0+\sum_{i=1}^{p}\alpha_i Y_{t-i}+\varepsilon_t,\ \varepsilon_t\backsim\mathsf{NIID}(0,\sigma_\varepsilon^2),\ t\in\mathbb{N}:=(1,2,...,n,...),\tag{1}$$

and Slutsky (1927) proposing the Moving Average of order q [MA(q)] model:

$$\text{MA(q)}:\quad Y_t=\beta_0+\sum_{i=1}^{q}\beta_i\varepsilon_{t-i}+\varepsilon_t,\ \varepsilon_t\backsim\mathsf{NIID}(0,\sigma_\varepsilon^2),\ t\in\mathbb{N},\tag{2}$$

where '$\mathsf{NIID}(0,\sigma_\varepsilon^2)$' stands for 'Normal, Independent and Identically Distributed' with mean 0 and variance σ_ε^2. The specifications in (1) and (2) are viewed as **aPriori Postulated (aPP)** models that could have given rise to data $\mathbf{y}_0:=(y_1,y_2,...,y_n)$.

Wold (1938) combined the AR(p) and MA(q) models into the ARMA(p,q) model:

$$\text{ARMA(p,q)}:\quad Y_t=\alpha_0+\sum_{i=1}^{p}\alpha_i Y_{t-i}+\sum_{i=1}^{q}\beta_i\varepsilon_{t-i}+\varepsilon_t,\ \varepsilon_t\backsim\mathsf{NIID}(0,\sigma_\varepsilon^2),\ t\in\mathbb{N}.\tag{3}$$

This was based on his 'decomposition theorem' that views a stationary process $\{Y_t,\ t\in\mathbb{N}\}$ as an element in a Hilbert space of square-integrable functions $(E(|Y_t|^2)<\infty)$. This is primarily a mathematical approximation result whose inferential component was developed subsequently; see Anderson (1971).

Box and Jenkins (1970) extended the ARMA(p,q) family of models to allow for a very special form of nonstationarity for the stochastic processes $\{Y_t,\ t\in\mathbb{N}\}$: "Models which describe such homogeneous nonstationarity behavior can be obtained by supposing some suitable *difference* of the process is stationary." (p. 85). That is the non-stationary process $\{X_t,\ t\in\mathbb{N}\}$ is transformed into a stationary process using d-differencing, $\{Y_t=\Delta^d X_t,\ t\in\mathbb{N}\}$, for $d\geq 1$, $\Delta=(1-L)$, $L^k X_t=X_{t-k}$. This gave rise to the ARIMA(p,d,q) family of models ('I' stands for 'Integrated') based on $\{Y_t,\ t\in\mathbb{N}\}$, which imposes unit roots on the AR(p) component.

The relevance and role of the unit roots imposed by $(1-L)^d$ onto $\{X_t,\ t\in\mathbb{N}\}$ were not fully appreciated until the late 1970s, when Dickey and Fuller (1979) proposed a framework for unit-root testing by extending White's (1958) results for the simplest AR(1) model: $Y_t=\alpha_1 Y_{t-1}+\varepsilon_t,\ \alpha_1\in[-1,1],\ \varepsilon_t\backsim\mathsf{NIID}(0,\sigma_\varepsilon^2),\ t\in\mathbb{N}$. This brought about radical and rapid changes to time series modeling, known as the 'unit-root revolution', which involved a novel inferential component for Unit-Root (UR) processes $\{Y_t,\ t\in\mathbb{N}\}$ based on non-standard asymptotic sampling distributions, framed in terms of functionals of the Brownian motion process; see Phillips (1987).

The unit-root revolution also expanded into multivariate models, including the Vector AR(p) (Hendry, 1995, Lutkepohl, 2005, Juselius, 2006):

$$\text{VAR(p)}: \quad \mathbf{Z}_t = \mathbf{a}_0 + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{Z}_{t-i} + \boldsymbol{\varepsilon}_t, \ \boldsymbol{\varepsilon}_t \backsim \mathsf{NIID}(\mathbf{0}, \boldsymbol{\Sigma}), \ t \in \mathbb{N}, \tag{4}$$

and the related cointegration analysis (Engle and Granger, 1987; Johansen, 1991). The new unit-root procedures were also extended to panel data; see Pesaran (2015).

1.2 Dickey-Fuller (D-F) unit-root testing

It should be noted at the outset that focusing on the AR(1) and Unit Root [UR(1)] models does not restrict the generality of the discussion since the AR(p) model can be reparametrized to isolate a potential unit root $\alpha_1 = \sum_{i=1}^{p} \beta_i$ (Said and Dickey, 1984):

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \varepsilon_t \rightarrow \quad Y_t = \beta_0 + \alpha_1 Y_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta Y_{t-i} + \varepsilon_t. \tag{5}$$

Dickey and Fuller (1979) pioneered the modern era of unit root testing based on:

$$H_0: \ \alpha_1 = 1 \text{ vs. } H_1: \ |\alpha_1| < 1, \tag{6}$$

by extending White's (1958) results to AR(1) models (B)-(C) in Table 1.

Table 1: Dickey-Fuller (D-F) AR(1) models

(A)	$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t, \ \varepsilon_t \backsim \mathsf{NIID}(0, \sigma_\varepsilon^2), \ t \in \mathbb{N},$	$\alpha_1 \in [-1, 1], \ \sigma_\varepsilon^2 > 0,$
(B)	$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t, \ \varepsilon_t \backsim \mathsf{NIID}(0, \sigma_\varepsilon^2), \ t \in \mathbb{N},$	$\alpha_1 \in [-1, 1], \alpha_0 \in \mathbb{R} := (-\infty, \infty),$
(C)	$Y_t = \alpha_0 + \delta t + \alpha_1 Y_{t-1} + \varepsilon_t, \ \varepsilon_t \backsim \mathsf{NIID}(0, \sigma_\varepsilon^2), \ t \in \mathbb{N},$	$\alpha_1 \in [-1, 1], (\alpha_0, \delta) \in \mathbb{R}^2.$

White (1958) considered the AR(1)-(A) model, pioneering the asymptotic *non-standard* sampling distribution of the OLS estimator $\widehat{\alpha}_1 = [\sum_{t=2}^{n} (Y_t Y_{t-1}) / \sum_{t=2}^{n} Y_{t-1}^2]$ in three steps. *First*, he pondered over the different choices of the initial Y_0:

(i) $Y_0 = c$ (constant), (ii) $Y_0 \backsim \mathsf{N}(0, \sigma_\varepsilon^2 / (1 - \alpha_1^2))$, (iii) $Y_0 = Y_n$,

selecting $Y_0 = c$, and thus omitting $f(y_0; \boldsymbol{\varphi}_0)$ from the joint distribution of $\mathbf{Y}$:

$$f(\mathbf{y}; \boldsymbol{\varphi}) = \prod_{t=1}^{n} f(y_t | y_{t-1}; \boldsymbol{\varphi}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} (y_t - \alpha_1 y_{t-1})^2\right\}, \ \forall \mathbf{y} \in \mathbb{R}^n. \tag{7}$$

Second, he derived the proper scaling $g(n)$ for $\widehat{\alpha}_1$ by using Fisher's information based on (7), noting his invoking 'independence' but explaining that "This approach does, however, give an heuristic method for finding a function $g(n)$ such that $g(n)(\widehat{\alpha}_1 - \alpha_1)$ has a non-degenerate limiting distribution." *Third*, White deduced the asymptotic distribution of $\widehat{\alpha}_1$ for $\alpha_1 = 1$ by invoking Donsker's (1951) invariance principle (p. 1196):

$$\text{AR(1)-(A)}\mathbf{:} \ n(\widehat{\alpha}_1 - 1) \underset{n \to \infty}{\overset{\alpha_1 = 1}{\backsim}} [\int_0^1 \mathsf{B}^2(r) dr]^{-1} \int_0^1 \mathsf{B}(r) d\mathsf{B}(r)], \tag{8}$$

where $\{\mathsf{B}(r), \ r \in [0, 1]\}$ denotes the Brownian motion process; see Billingsley (1995).

Using the aPP perspective, Dickey and Fuller (1979), and the subsequent unit-root literature extended (8) to models (B)-(C) in Table 1 (Phillips and Perron, 1988):

$$\text{AR(1)-(B)}\mathbf{:} \ n(\widehat{\alpha}_1 - 1) \underset{n \to \infty}{\overset{\alpha_1 = 1}{\backsim}} [\int_0^1 \overline{\mathsf{B}}^2(r) dr]^{-1} \int_0^1 \overline{\mathsf{B}}(r) d\overline{\mathsf{B}}(r), \quad \overline{\mathsf{B}}(r) = \mathsf{B}(r) - \int_0^1 \mathsf{B}(r) dr, \tag{9}$$

$$\begin{gathered} \text{AR(1)-(C)}\mathbf{:} \ n(\widehat{\alpha}_1 - 1) \underset{n \to \infty}{\overset{\alpha_1 = 1}{\backsim}} [\int_0^1 \widetilde{\mathsf{B}}^2(r)(r) dr]^{-1} \int_0^1 \widetilde{\mathsf{B}}(r) d\widetilde{\mathsf{B}}(r), \\ \widetilde{\mathsf{B}}(r) = \mathsf{B}(r) - 4[\int_0^1 \mathsf{B}(r) dr - \tfrac{3}{2} \int_0^1 r \mathsf{B}(r) dr] - 6r[\int_0^1 \mathsf{B}(r) dr - 2 \int_0^1 r \mathsf{B}(r) dr]. \end{gathered} \tag{10}$$

These results have been extended further by assigning more elaborate probabilistic structures, such as ARMA(p,q), to $\{\varepsilon_t, \ t \in \mathbb{N}\}$; see Choi (2015), Hatanaka (1996).

1.3 The aPP vs. the statistical perspective: a preview

Broadly speaking, empirical modeling and inference are guided by two perspectives that serve interrelated but complementary objectives: the traditional *aPP perspective*, relying primarily on substantive (theoretical) subject matter information, and the *statistical perspective*, relying exclusively on the statistical systematic information mirrored by the chance regularity patterns exhibited by data $\mathbf{y}_0$; see Spanos (2010).

The aPP perspective, dominating econometrics since the 1930s, views the AR(1): (A)-(C) models in Table 1 as stochastic difference equations, framed by prespecifying:

$\langle\mathbf{a}\rangle$ the functional form of the equations,
$\langle\mathbf{b}\rangle$ the range of values of their unknown parameters, and
$\langle\mathbf{c}\rangle$ the probabilistic assumptions of their error term ε_t; see Mann and Wald (1943).

The aPP perspective is appropriate for evaluating the behavior of a given *substantive* model, say $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$, $\boldsymbol{\varphi}\in\Phi$, under different shocks $\{\varepsilon_t,\ t\in\mathbb{N}\}$, with a view to evaluate its stability as a *given* system (Hespanha, 2018), and fine-tune its performance for forecasting, policy, and control purposes; see Box and Jenkins (1970).

It turns out, however, that not every aPP model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$, $\boldsymbol{\varphi}\in\Phi$, is well-grounded for inference with data $\mathbf{y}_0$ since imposing $\langle\mathbf{a}\rangle$-$\langle\mathbf{c}\rangle$ on $\mathbf{y}_0$ would often give rise to unreliable inferences. Why? The probabilistic structure that matters for inference purposes is not that of the error term, but the one indirectly assigned by $\{\varepsilon_t,\ t\in\mathbb{N}\}$ to the observable process $\{Y_t,\ t\in\mathbb{N}\}$ underlying data $\mathbf{y}_0$. The letter is unveiled using the **statistical perspective** that provides the missing link between an aPP model, $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$, $\boldsymbol{\varphi}\in\Phi$, and the real-world mechanism that generated $\mathbf{y}_0$, in the form of the implicit statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$, $\boldsymbol{\theta}\in\Theta$, framed entirely in terms of probabilistic assumptions assigned to the process $\{Y_t,\ t\in\mathbb{N}\}$ (section 2).

The primary role of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ is (a) to ensure that $\{Y_t,\ t\in\mathbb{N}\}$ is *well-defined* in both a probabilistic sense (Kolmogorov's existence theorem), and (b) statistical sense (its probabilistic assumptions are *complete* and *testable,* rendering feasible the task of establishing its *statistical adequacy* for data $\mathbf{y}_0$). (a)-(b) frame the preconditions for $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$ to be *well-grounded* apropos of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$:

[S1] the aPP $\langle\mathbf{a}\rangle$-$\langle\mathbf{c}\rangle$ are affirmed by probabilistic assumptions assigned to $\{Y_t,\ t\in\mathbb{N}\}$.
[S2] $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$ does not belie the statistical systematic information in data $\mathbf{y}_0$.

When [S1] is false there is no well-defined process $\{Y_t,\ t\in\mathbb{N}\}$ (in Kolmogorov's sense) underlying $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$, and when [S2] is false, either (i) $\boldsymbol{\varphi}\in\Phi$ is not uniquely defined (identified) with respect to $\boldsymbol{\theta}\in\Theta$, or (ii) $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$ is statistically misspecified for $\mathbf{y}_0$.

Kolmogorov's (1933) existence/extension theorem specifies when a stochastic process $\{Y_t,\ t\in\mathbb{N}\}$ is 'well-defined' in terms of its joint distributions:

$$F_n(y_1,...,y_n;\boldsymbol{\psi})=\mathbb{P}(s\colon\ Y_1(s)\leq y_1,\cdots,Y_n(s)\leq y_n),\ \ s\in S,\ \ \mathbf{y}\in\mathbb{R}^n_Y,\ \text{ for } n=1,2,3,..., \tag{11}$$

satisfying the *symmetry* (often trivially valid) and the *consistency* condition:

$$\lim_{y_n\rightarrow\infty}F_n(y_1,...,y_n;\boldsymbol{\psi})=F_{n-1}(y_1,...,y_{n-1};\boldsymbol{\psi}),\ \ \mathbf{y}\in\mathbb{R}^n_Y,\ \text{ for } n=2,3,... \tag{12}$$

When these conditions hold for (11), it can be inferred that:

(i) existence: $\{Y_t,\ t\in\mathbb{N}\}$ is a well-defined stochastic process, with distribution $F_n(y_1, ..., y_n; \boldsymbol{\psi})$, $\mathbf{y}\in\mathbb{R}^n_Y$, defined on a probability space $(S, \Im, \mathbb{P}(.))$, and

(ii) extension: $F_n(y_1, ..., y_n; \boldsymbol{\psi})$ can be extended asymptotically as $n \to \infty$.

As Billingsley (1995) points out: "For many purposes the underlying probability space is irrelevant, the joint distribution of the variables in the process being all that matters." p. 486. Interestingly, (12) ensures that $F_n(y_t|y_{t-1}, ..., y_1; \boldsymbol{\phi}_t)$, t=2, 3, ..., n, is properly defined to frame the dependence in $\{Y_t,\ t\in\mathbb{N}\}$, providing the cornerstone for the theory of dependent stochastic processes; see Renyi (1970), Shiryaev (1996).

The primary objective of the paper is twofold. *First*, to argue that the AR(1) models in Table 1 are not well-grounded since stipulations [S1]-[S2] do not hold, but the traditional unit-root literature ignores that by taking presumptions $\langle\mathbf{a}\rangle$-$\langle\mathbf{c}\rangle$ at face value, where $\langle\mathbf{b}\rangle$ specifies $\alpha_1\in[-1, 1]$ entailing:

$\langle\mathbf{d}\rangle$ the AR(1) *nests* the Unit Root [UR(1)] model when α_1=1,

which is the primary source of several statistical conundrums, including:

[C1] The asymptotic non-standard sampling distributions (Phillips, 1987).

[C2] The low power of unit-root tests for α_1 near 1. "One general conclusion to emerge is that, ..., the discriminatory power in all of the tests between models with a root at unity and a root close to unity is generally low." (Phillips and Xiao,1998, p. 22).

Attempts to meliorate [C2], include re-framing the unit-root hypotheses (Leybourne, 1995), reshaping the initial conditions (Müller and Elliott, 2003), adding extraneous variables to the AR(1) model (Hansen, 1995), as well as:

[C3] replacing the D-F in Table 1 with the Error-Autocorrelation (E-A) parametrization the Error-Autocorrelation (E-A); see Dickey (1984), Bhargava (1986).

[C4] The incompleteness and non-testability of the error term assumptions undermining the statistical adequacy of the estimated AR(p) models.

Second, to bring out explicitly the statistical models corresponding to the AR(1): (A)-(C) models in Table 1 using Fisher's statistical perspective to frame the stipulations [S1]-[S2] in terms of probabilistic assumptions assigned to $\{Y_t,\ t\in\mathbb{N}\}$. It is argued that presumption $\langle\mathbf{d}\rangle$ belies Kolmogorov's existence theorem and thus there is no stochastic process $\{Y_t,\ t\in\mathbb{N}\}$ underlying both the AR(1) and UR(1) models. Instead, the statistical AR(1) and UR(1) models (i) are based on two distinct processes $\{Y_t,\ t\in\mathbb{N}\}$, with different (ii) probabilistic assumptions and (iii) statistical parametrizations, (iv) rendering them *non-nested* (Hatanaka, 1996), and (v) their respective likelihood-based inferential components are free from the conundrums [C1]-[C4]. The claims (i)-(iv) are affirmed by the recursive 'solution' of the AR(1) models when Y_0 is treated as a typical element of $\{Y_t,\ t\in\mathbb{N}\}$ (Appendix A). The Error-Autocorrelation (E-A) framing of the AR(1): (B)-(C) models in Table 1 also confirms the statistical parametrizations of $(\alpha_0, \beta_0, \beta_1)$, but ignores those of the key parameters α_1 and σ^2_ε at its peril. Simulations based on the statistical AR(1) and UR(1) models affirm the claims (i)-(v) but indicate that (vi) the two models can be observationally equivalent for α_1 near 1, suggesting that the choice between them should be based on statistical adequacy grounds. To affirm the key role of well-defined models, a family of non-

-stationary AR(1) models that nests the corresponding UR(1) models is proposed, evidencing by analytical derivations and simulations that neither of the conundrums [C1]-[C2] arises when testing for $\alpha_1{=}1$ using likelihood-based tests

2 Empirical modeling and inference in practice

2.1 The aPP (substantive) vs. the statistical model

Adopting a broader view of empirical modeling, we distinguish between the aPP substantive (theoretical/structural) and the statistical information in data $\mathbf{y}_0$.

The aPP perspective gives rise to a 'substantive model', denoted by:

$$\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})=\{f(\mathbf{y};\boldsymbol{\varphi}),\ \boldsymbol{\varphi}\in\Phi\subset\mathbb{R}^p\},\ \text{for all } (\forall)\ \mathbf{y}\in\mathbb{R}^n_Y,\ p<n, \tag{13}$$

that revolves around its joint distribution $f(\mathbf{y};\boldsymbol{\varphi})$, $\boldsymbol{\varphi}\in\Phi$, $\forall\mathbf{y}\in\mathbb{R}^n_Y$, indirectly specified via the error term assumptions. Implicit in every substantive model, there is a 'statistical model', whose generic form is (Spanos, 2010):

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})=\{f(\mathbf{y};\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta\subset\mathbb{R}^m\},\ \forall\mathbf{y}\in\mathbb{R}^n_Y,\ p\leq m<n, \tag{14}$$

and (a) comprises the probabilistic assumptions imposed (often implicitly) on the observable process $\{Y_t,\ t\in\mathbb{N}\}$ underlying data $\mathbf{y}_0$, determining $f(\mathbf{y};\boldsymbol{\theta})$, $\boldsymbol{\theta}\in\Theta$, $\forall\mathbf{y}\in\mathbb{R}^n_Y$, and (b) provides the crucial link between $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$ and the real-world mechanism that generated data $\mathbf{y}_0$. Hence, $\boldsymbol{\theta}\in\Theta$ is framed purposefully to relate the two sets of parameters via restrictions, say $\mathbf{g}(\boldsymbol{\varphi},\boldsymbol{\theta})=\mathbf{0}$. The (substantive) parameters $\boldsymbol{\varphi}\in\Phi$ are 'identified' only when $\mathbf{g}(\boldsymbol{\varphi},\boldsymbol{\theta})=\mathbf{0}$ defines $\boldsymbol{\varphi}$ uniquely in terms of $\boldsymbol{\theta}$; see Spanos (1990b).

2.2 Fisher's model-based statistical perspective

The concept of a 'statistical model' plays a crucial role in both frequentist and Bayesian inference, but it has not been adequately delineated for empirical modeling and inference purposes. McCullagh (2002) discourses this issue by posing and answering several key questions, including: "What is a statistical model?" and "What is a parameter?" The notions that a model must "make sense," and that a parameter must "have a well-defined meaning" are deeply ingrained in applied statistical work, ..., but absent from most formal theories of modelling and inference. " (p. 1225). He proceeds to criticize the statistics literature for ignoring such questions and oppugning postulating statistical models in 'ad hoc and idiosyncratic ways', including attaching generic error terms to mathematical equations.

McCullagh's (2002) questions and concerns can be addressed in the context of R. A. Fisher's (1922) statistical model-based perspective, grounded on a parametric statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ in (14) being viewed as an 'approximate and idealized' description of the statistical mechanism that could have given rise to $\mathbf{y}_0$. The specification of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ revolves around selecting: (i) the appropriate probabilistic assumptions from all three broad categories, Distribution (D), Dependence (M), and Heterogeneity (H), assigned to $\{Y_t,\ t\in\mathbb{N}\}$ to render data $\mathbf{y}_0$ a 'typical realization' thereof, and (ii) the particular parametrization $\boldsymbol{\theta}\in\Theta$ to probe the questions of interest; see Spanos (2019).

The cornerstone of frequentist inference is the sampling distribution, $f(z_n;\boldsymbol{\theta})$, of a statistic $Z_n{=}g(Y_1, Y_2, ..., Y_n)$ (estimator, test, predictor), derived via:

$$F_n(z){=}\mathbb{P}(Z_n{\leq}z){=}\underbrace{\int\int\cdots\int}_{\{\mathbf{y}:\ g(\mathbf{y})\leq z\}} f(\mathbf{y};\boldsymbol{\theta})d\mathbf{y},\ \forall z{\in}\mathbb{R}_Z. \tag{15}$$

with the derivation of $f(z_n;\boldsymbol{\theta}){=}dF_n(z)/dz$ in (15) presuming that $f(\mathbf{y};\boldsymbol{\theta}),\ \boldsymbol{\theta}{\in}\Theta$, $\forall\mathbf{y}{\in}\mathbb{R}^n_Y$, (a) is well-defined, and (b) its probabilistic assumptions are valid for $\mathbf{y}_0$.

The statistical perspective comprises two facets, the *modeling* (specification, Mis-Specification (M-S) testing, respecification) and the *inference*. The modeling aims to reach a statistically adequate $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ with data $\mathbf{y}_0$, which can only be secured when $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ comprises a complete set of testable probabilistic assumptions (Table 2), in conjunction with M-S testing to probe their validity; see Spanos (2018).

The **statistical adequacy** (approximate validity) of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ with data $\mathbf{y}_0$ is crucially important for the reliability of the invoked inference procedures since the sampling distributions are derived via (15) presuming the validity of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$, i.e. $f(\mathbf{y};\boldsymbol{\theta}),\ \mathbf{y}{\in}\mathbb{R}^n_Y$. The latter underwrites the effectiveness (optimality) and reliability of statistical inference, by ensuring that the *actual* error probabilities approximate closely the *nominal* ones – derived by presuming the validity of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$. That is the 'approximate validity' of $f(\mathbf{y};\boldsymbol{\theta}),\ \mathbf{y}{\in}\mathbb{R}^n_Y$, is what matters for inference purposes.

When $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ is *statistically misspecified*, (i) $f(\mathbf{y};\boldsymbol{\theta}),\ \mathbf{y}{\in}\mathbb{R}^n_Y$, and the likelihood function $L(\boldsymbol{\theta};\mathbf{y}_0){\propto}f(\mathbf{y}_0;\boldsymbol{\theta}),\ \boldsymbol{\theta}{\in}\Theta$, are erroneous, (ii) distorting the sampling distribution $f(z_n;\boldsymbol{\theta})$ derived via (15), (iii) giving rise to 'non-optimal' inference procedures, and (iv) inducing sizeable discrepancies between the *actual* and *nominal* error probabilities; see Spanos (2019). Regrettably, the aPP perspective pays little to no attention to the statistical adequacy of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$, largely ignoring statistical misspecification. Hence, $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$ provides reliable inferential underpinnings for $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$ only when it is well-grounded in both probabilistic (Kolmogorov) and statistical sense (adequate).

2.3 The aPP perspective: key questions and conundrums

As a preview of the issues raised and discoursed, consider the following example.

Example 1. The AR(1)-(B) model is specified using presumptions $\langle\mathbf{a}\rangle$-$\langle\mathbf{d}\rangle$:

$$\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y}):\ \ Y_t{=}\alpha_0{+}\alpha_1 Y_{t-1}{+}\varepsilon_t,\ \boldsymbol{\varphi}{:=}(\alpha_0,\alpha_1,\sigma^2_\varepsilon){\in}\Phi,\ \varepsilon_t{\backsim}\mathsf{NIID}(0,\sigma^2_\varepsilon),\ t{\in}\mathbb{N}, \tag{16}$$

where $\Phi{:=}[\mathbb{R}{\times}[-1,1]{\times}\mathbb{R}_+$. An aPP practitioner is likely to use the transfer:

$$(\varepsilon_t|Y_{t-1})\backsim\mathsf{NIID}(0,\sigma^2_\varepsilon)\ \longrightarrow\ (Y_t|Y_{t-1})\backsim\mathsf{N}\left(\alpha_0{+}\alpha_1 Y_{t-1},\sigma^2_\varepsilon\right),\ t{\in}\mathbb{N}, \tag{17}$$

to define the distribution of the sample $\mathbf{Y}{:=}(Y_1, Y_2, ..., Y_n)$ for the AR(1)-(B) by:

$$f(\mathbf{y};\boldsymbol{\varphi}){=}f(y_0;\boldsymbol{\phi}_0)\textstyle\prod_{t=1}^n f(y_t|y_{t-1};\boldsymbol{\varphi}),\ \boldsymbol{\varphi}{:=}(\alpha_0,\alpha_1,\sigma^2_\varepsilon){\in}\boldsymbol{\Phi},\ \forall\mathbf{y}{\in}\mathbb{R}^n. \tag{18}$$

The initial Y_0 is handled by choosing one from the different choices for $f(y_0;\boldsymbol{\varphi}_0)$:

$\langle\mathbf{e}\rangle$ (i) $Y_0{=}c$, (ii) $Y_0{\backsim}\mathsf{N}(0,\frac{\sigma^2_\varepsilon}{(1-\alpha_1^2)})$, (iii) $Y_0{\backsim}\mathsf{D}(0,\frac{\sigma^2_\varepsilon}{(1-\alpha_1^2)})$, (iv) $Y_0{=}Y_n$, etc.

Invoking, say $\langle\mathbf{e}\rangle$-(i), the practitioner would define the likelihood function by:

$$L(\boldsymbol{\varphi};\mathbf{y}){\propto}\textstyle\prod_{t=1}^n f(y_t|y_{t-1};\boldsymbol{\varphi}){=}(2\pi\sigma^2)^{-\frac{n}{2}}\exp[\text{-}\frac{1}{2\sigma^2}\sum_{t=1}^n(y_t{-}\alpha_0{-}\alpha_1 y_{t-1})^2],\ \forall\boldsymbol{\varphi}{\in}\boldsymbol{\Phi}. \tag{19}$$

To evaluate whether the AR(1)-(B) model in (16) is well-grounded, one is required to evaluate the validity of stipulations [S1]-[S2] (section 1.3) for the joint distribution in (18). This calls for unveiling its implicit statistical model, $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$, $\boldsymbol{\theta}\in\Theta$, framed in terms of a complete set of probabilistic assumptions assigned to $\{Y_t,\ t\in\mathbb{N}\}$, specifying its joint distribution $f(\mathbf{y};\boldsymbol{\psi})$, $\boldsymbol{\psi}\in\Psi$, to be juxtaposed with $f(\mathbf{y};\boldsymbol{\varphi}), \boldsymbol{\varphi}\in\boldsymbol{\Phi}$, $\forall\mathbf{y}\in\mathbb{R}^n$, in (18). The transfer in (17), however, is inadequate to determine $f(\mathbf{y};\boldsymbol{\psi})$, $\forall\mathbf{y}\in\mathbb{R}^n$, since: [i] Normality: $(Y_t|Y_{t-1})\backsim\mathsf{N}(.,.)$, [ii] Linearity: $E(Y_t|Y_{t-1})=\alpha_0+\alpha_1 Y_{t-1}$, [iii] Homoskedasticity: $Var(Y_t|Y_{t-1})=\sigma_\varepsilon^2$, [iv] Markov: $f(y_t|y_{t-1},...y_1;\boldsymbol{\varphi})=f(y_t|y_{t-1};\boldsymbol{\varphi})$, are incomplete because there is an additional (implicit) assumption:

$$\boxed{\text{[v] the parameters } \boldsymbol{\varphi}:=(\alpha_0,\alpha_1,\sigma_0^2) \text{ are constant over } t\in\mathbb{N},} \tag{20}$$

which renders the transfer $\{(Y_t|Y_{t-1})\,,\ t\in\mathbb{N}\}\longrightarrow\{Y_t,\ t\in\mathbb{N}\}$ more challenging.

It can be shown that the assumptions [i]-[iv] imply that $\{Y_t,\ t\in\mathbb{N}\}$ is Markov (M) ([iv]), and Normal (N) since [i]-[iv] imply that $f(y_t,y_{t-1};\boldsymbol{\varphi})$ is Normal, stemming from the two-way linearity, [ii] and $E(Y_{t-1}|Y_t)=a_0+a_1 Y_t$ (due to [iv]), combined with [iii], which extends to $f(\mathbf{y};\boldsymbol{\psi})$, $\boldsymbol{\psi}\in\Psi$, of $\{Y_t,\ t\in\mathbb{N}\}$ due to [iv]; see Spanos (1995). There is, however, a missing Heterogeneity (H) assumption for $\{Y_t,\ t\in\mathbb{N}\}$ needed to affirm assumption [v] in (20). This is because assuming $\{Y_t,\ t\in\mathbb{N}\}$ is M and N, yields:

$$f(\mathbf{y};\boldsymbol{\psi})\overset{\mathsf{M}}{=}f_1(y_1;\boldsymbol{\varphi}_0)\prod_{t=2}^{n} f_t(y_t|y_{t-1};\boldsymbol{\varphi}(t)),\ \forall\mathbf{y}\in\mathbb{R}^n, \tag{21}$$

with $\mathsf{N}\rightarrow(Y_t|Y_{t-1})\backsim\mathsf{N}\left(\alpha_0(t)+\alpha_1(t)Y_{t-1},\sigma_\varepsilon^2(t)\right),\ \boldsymbol{\varphi}(t):=\left(\alpha_0(t),\alpha_1(t),\sigma_0^2(t)\right),\ t\in\mathbb{N}$. This result shows that assumptions [i]-[iv] above are affirmed, but N and M in (21) give rise to an AR(1) model with t-varying parameters $\boldsymbol{\varphi}(t)$, belying [v]. Note: it is important to distinguish between heterogeneity: $Var\left(Y_t|Y_{t-1}\right)=g_1(t),\ t\in\mathbb{N}$, and heteroskedasticity: $Var\left(Y_t|Y_{t-1}\right)=g_2(Y_{t-1})$; see Spanos (1986), p. 473.

This brings out the broader potential conflict between the aPP perspective, based on presumptions $\langle\mathbf{a}\rangle$-$\langle\mathbf{e}\rangle$, and the statistical perspective on the AR(1)-(B) model since for a well-defined (as per Kolmogorov's existence theorem) process $\{Y_t,\ t\in\mathbb{N}\}$ underlying the statistical AR(1) model, its joint distribution $f(\mathbf{y};\boldsymbol{\psi})$, $\boldsymbol{\psi}\in\Psi$, determines:

$\langle\mathbf{a}\rangle^*$ the functional form of $E\left(Y_t|Y_{t-1}\right)=h(Y_{t-1};\boldsymbol{\theta}_1)$ and $Var\left(Y_t|Y_{t-1}\right)=g(Y_{t-1};\boldsymbol{\theta}_2)$,
$\langle\mathbf{b}\rangle^*$ their statistical prametrizations $\boldsymbol{\theta}:=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\in\Theta\subset\mathbb{R}^m$, $m<n$,
$\langle\mathbf{c}\rangle^*$ the probabilistic structure of the (statistical) error term $u_t=Y_t-E\left(Y_t|Y_{t-1}\right)$,
$\langle\mathbf{d}\rangle^*$ whether or not the AR(1) nests the UR(1) model for $\alpha_1=1$ stems from $\langle\mathbf{b}\rangle^*$,
$\langle\mathbf{e}\rangle^*$ Y_1 has the same probabilistic structure as every other element of $\{Y_t,\ t\in\mathbb{N}\}$.

The stipulations $\langle\mathbf{a}\rangle^*$-$\langle\mathbf{e}\rangle^*$ are internally consistent and interrelated via the well-defined $f(\mathbf{y};\boldsymbol{\psi})$, $\boldsymbol{\psi}\in\Psi$, but the presumptions $\langle\mathbf{a}\rangle$-$\langle\mathbf{e}\rangle$ are not necessarily so unless $f(\mathbf{y};\boldsymbol{\varphi})$, $\boldsymbol{\varphi}\in\boldsymbol{\Phi}$, $\forall\mathbf{y}\in\mathbb{R}^n$, in (18) is also well-defined. It is argued that this is not the case, since $\langle\mathbf{b}\rangle^*$-$\langle\mathbf{e}\rangle^*$ oppugn $\langle\mathbf{b}\rangle$-$\langle\mathbf{e}\rangle$ for being internally inconsistent.

Why has the traditional literature ignored these potential conflicts?

The choices in (17)-(19) appear right-minded to a traditional practitioner since $\langle\mathbf{a}\rangle$-$\langle\mathbf{e}\rangle$ are, indeed, pertinent for evaluating the behavior of the aPP AR(1)-(B) model under different shocks $\{\varepsilon_t,\ t\in\mathbb{N}\}$. Hence, to be mindful of any problems when the aPP is used to guide statistical modeling and inference requires one to be cognizant of the statistical perspective with its well-grounded models and statistical parametrizations, which seems gratuitously laborious and needless to such a practitioner.

2.4 Well-defined statistical AR(1) and UR(1) models

Key question: does there exist a well-defined process $\{Y_t,\ t\in\mathbb{N}\}$ as per Kolmogorov's existence theorem, grounded on $\langle\mathbf{a}\rangle$-$\langle\mathbf{e}\rangle$ with a joint distribution given in (18) that nests parametrically the corresponding UR(1) model when $\alpha_1{=}1$?

To answer this question necessitates a complete set of probabilistic assumptions for $\{Y_t,\ t\in\mathbb{N}\}$ to affirm, not only assumptions [i]-[iv], but also [v] in (20). This calls for adding a Heterogeneity (H) assumption to N and M in (21), likely to be 'stationarity'.

Example 2. Assuming $\{Y_t,\ t\in\mathbb{N}\}$ is a Markov (M), Stationary (S) process yields:

$$f(\mathbf{y};\boldsymbol{\psi})\overset{\text{M}}{=}f_1(y_1;\boldsymbol{\phi}_0)\prod_{t=2}^{n} f_t(y_t|y_{t-1};\boldsymbol{\phi}_{1t})\overset{\text{M\&S}}{=}f(y_1;\boldsymbol{\phi}_0)\prod_{t=2}^{n} f(y_t|y_{t-1};\boldsymbol{\phi}_1),\ \ \forall\mathbf{y}\in\mathbb{R}^n. \tag{22}$$

Adding the Normality (N) assumption yields $f(y_t,y_{t-1};\boldsymbol{\phi})$ (Spanos, 1990a):

$$\begin{pmatrix} Y_t \\ Y_{t-1}\end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_0 \\ \mu_0\end{pmatrix},\begin{pmatrix} \sigma(0) & \sigma(1) \\ \sigma(1) & \sigma(0)\end{pmatrix}\right), \tag{23}$$

where $E(Y_t){=}\mu,\ Var(Y_t){=}\sigma(0),\ Cov(Y_t,Y_{t-1}){=}\sigma(1),\ t\in\mathbb{N}$. This gives rise to the statistical AR(1) model in Table 2, based on the probabilistic reduction:

$$f(y_t,y_{t-1};\boldsymbol{\phi}){=}f(y_t|y_{t-1};\boldsymbol{\phi}_1)\cdot f(y_{t-1};\boldsymbol{\phi}_2),\ \forall\,(y_t,y_{t-1})\in\mathbb{R}^2, \tag{24}$$

$$\text{where }\ (Y_t|Y_{t-1})\backsim\mathsf{N}\left(\alpha_0+\alpha_1 Y_{t-1},\ \sigma_0^2\right),\ Y_{t-1}\backsim\mathsf{N}\left(\mu_0,\sigma(0)\right),\ t\in\mathbb{N}. \tag{25}$$

Table 2: Statistical AR(1) model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$

Statistical GM:		$Y_t{=}\alpha_0+\alpha_1 Y_{t-1}+u_{0t},\ t\in\mathbb{N}.$	
[1]	Normality:	$(Y_t,Y_{t-1})\backsim\mathsf{N}(.,.),\ (y_t,y_{t-1})\in\mathbb{R}^2,$	$t\in\mathbb{N}.$
[2]	Linearity:	$E\left(Y_t|\sigma(Y_{t-1})\right){=}\alpha_0+\alpha_1 Y_{t-1},$	$t\in\mathbb{N}.$
[3]	Homoskedasticity:	$Var\left(Y_t|\sigma(Y_{t-1})\right){=}\sigma_0^2,$	$t\in\mathbb{N}.$
[4]	Markov:	$\{Y_t,\ t\in\mathbb{N}\}$ is a Markov process,	$t\in\mathbb{N}.$
[5]	t-invariance:	$\boldsymbol{\theta}_{AR}{:=}(\alpha_0,\alpha_1,\sigma_0^2)$ are *not* changing with t,	$t\in\mathbb{N}.$
Parametrization:		$\alpha_0{=}\mu_0(1{-}\alpha_1)\in\mathbb{R},\ \alpha_1{=}\frac{\sigma(1)}{\sigma(0)}\in(-1,1),\ \sigma_0^2{=}\sigma(0)(1{-}\alpha_1^2)\in\mathbb{R}_+.$	

In contrast to the aPP AR(1)-(B) model in (16), its implicit statistical AR(1) model in Table 2 is specified in terms of the bivariate distribution $f(y_t,y_{t-1};\boldsymbol{\phi})$ in (24) because Y_{t-1} is *not* weakly exogenous for $\boldsymbol{\theta}_{AR}{=}\boldsymbol{\phi}_1$ since $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are *not* variation-free, e.g. $\mu{=}0\rightarrow\alpha_0{=}0$ (Engle et al., 1983). Hence, the statistical AR(1) model in Table 2 retains $f(y_{t-1};\boldsymbol{\phi}_2)$ by conditioning on $\sigma(Y_{t-1})$**,** the σ-field generated by Y_{t-1}; see Billingsley (1995). That brings out the fundamental role of $f(y_1;\boldsymbol{\phi}_0)$ in defining $f(\mathbf{y};\boldsymbol{\psi})$ by treating Y_1 as a typical element of $\{Y_t,\ t\in\mathbb{N}\}$. It is important to mention that for $\mu_0{=}0$, Table 2 would yield the statistical AR(1)-(A) model.

The probabilistic reduction of $f(\mathbf{y};\boldsymbol{\psi})$ in (22)-(25) based on the assumptions N, M, and S, ensures that the statistical AR(1) model in Table 2 "makes probabilistic sense" (McCullagh, 2002) in light of the following key features.

(i) The stochastic process $\{Y_t,\ t\in\mathbb{N}\}$ underlying the AR(1) model in Table 2 is well-define since it satisfies Kolmogorov's existence theorem.

(ii) The model assumptions [1]-[5] are equivalent (necessary and sufficient) to the reduction assumptions, i.e. (N, M, S) $\rightleftarrows$[1]-[5]; see Spanos (1995). This confirms that stationarity (S) is the missing H assumption for $\{Y_t,\ t\in\mathbb{N}\}$, and N is in-built, rendering ill-conceived any attempt to relax $\varepsilon_t\backsim\mathsf{N}(.,.)$ and arrogate generality.

(iii) Table 2 comprises a complete, internally consistent, and testable set of probabilistic assumptions [1]-[5] specifying the statistical AR(1) model. In contrast, the AR(1) models (Table 1) are often specified in terms of an incomplete and non-testable set of assumptions using a generic error process $\{\varepsilon_t,\ t\in\mathbb{N}\}$ (Phillips, 1987):

{i} $E(\varepsilon_t)=0$, for all $t\in\mathbb{N}$,	{ii} $\sup_t E\lvert\varepsilon_t\rvert^{\delta+\epsilon}<0$ for $\delta>2,\ \epsilon>0$,
{iii} $\lim\limits_{n\rightarrow\infty} E(\frac{1}{n}(\sum_{t=1}^{n}\varepsilon_t)^2=\sigma_\infty^2>0$,	{iv} $\{\varepsilon_t,\ t\in\mathbb{N}\}$ is strongly mixing with coefficient $\alpha_m \underset{m\rightarrow\infty}{\rightarrow} 0$ such that $\sum_{m=1}^{\infty}\alpha_m^{1-\delta/2}<\infty, \delta>2$.

Although the unit-root literature views {i}-{iv} as 'weak' assumptions expedient for the generality of the asymptotic non-standard sampling distributions in (8)-(10), what matters for the reliability of the ensuing inferences is whether (a) $\{Y_t,\ t\in\mathbb{N}\}$ is well-defined, and (b) the model assumptions [1]-[5] are valid for data $\mathbf{y}_0$.

(iv) Assumptions [1]-[5] assign to the statistical error term, $u_{0t}=[Y_t-E(Y_t|\sigma(Y_{t-1}))]$, "a well-defined meaning" as a model-specific Martingale difference (Md) process: $(u_{0t}|\sigma(Y_{t-1}))\backsim\mathsf{NMd}\,(0,\sigma_0^2)\,,\ t\in\mathbb{N}$; see section 3 for more details.

(v) The statistical Generating Mechanism (GM) in Table 2 designates 'how one could simulate artificial replicas of data $\mathbf{y}_0$'; see Cox (1990), p. 172.

The log-likelihood function of the AR(1) model in Table 2 takes the form:

$$\ln L(\boldsymbol{\theta}_{AR};\mathbf{y}_0)=\ln f(y_1;\boldsymbol{\phi}_0)+\sum_{t=2}^{n}\ln f(y_t|y_{t-1};\boldsymbol{\theta}_{AR}),\ \ \boldsymbol{\theta}_{AR}\in[\mathbb{R}\times(-1,1)\times\mathbb{R}_+], \qquad (26)$$

where $Y_1\backsim\mathsf{N}\,(\mu_0,\sigma(0))$ would be reparametrized in terms of $\boldsymbol{\theta}_{AR}$, yielding:

$$f(y_1;\boldsymbol{\theta}_{AR})=(2\pi(\sigma_0^2(1-\alpha_1^2))^{-\frac{1}{2}}\exp[-\tfrac{1}{2\sigma^2}(y_t-[\alpha_0/(1-\alpha_1)])^2],\ y_1\in\mathbb{R}. \qquad (27)$$

This indicates that invoking $\langle\mathbf{e}\rangle^*$ $Y_0\backsim\mathsf{N}\,(\mu_0,\sigma(0))$ would render the aPP joint distribution in (18) and the related likelihood function in (19) incongruous for $\alpha_1=1$. It is intriguing to note that White (1958) would have stumbled upon this problem if he were to select either (ii) $Y_0\backsim\mathsf{N}(0,\sigma_\varepsilon^2/(1-\alpha_1^2))$ or (iii) $Y_0=Y_n$, instead of (i) $Y_0=c$.

The problem with $\alpha_1=1$ could be affirmed, however, via the limits in probability of the OLS estimators of $\boldsymbol{\theta}_{AR}$ (Table 2): $\widehat{\alpha}_1=[\sum_{t=2}^{n}(Y_t-\overline{Y})(Y_{t-1}-\overline{Y})/\sum_{t=2}^{n}(Y_{t-1}-\overline{Y})^2]$, $\widehat{\alpha}_0=\overline{Y}-\widehat{\alpha}_1\overline{Y}_{-1}$, $\overline{Y}=\frac{1}{n}\sum_{t=1}^{n}Y_t$, $\widehat{\sigma}_\varepsilon^2=\frac{1}{n-2}\sum_{t=1}^{n}\widehat{\varepsilon}_t^2$, $\widehat{\varepsilon}_t=Y_t-\widehat{\alpha}_0-\widehat{\alpha}_1 Y_{t-1}$, using the first two sample moments of $\{Y_t,\ t\in\mathbb{N}\}$ converging in probability ($\stackrel{\mathbb{P}}{\longrightarrow}$) to the following limits: $\overline{Y}\stackrel{\mathbb{P}}{\longrightarrow}\mu_0,\ \frac{1}{n}\sum_{t=2}^{n}(Y_t-\overline{Y})(Y_{t-1}-\overline{Y})]\stackrel{\mathbb{P}}{\longrightarrow}\sigma(1),\ \frac{1}{n}\sum_{t=2}^{n}(Y_{t-1}-\overline{Y})^2]\stackrel{\mathbb{P}}{\longrightarrow}\sigma(0)$ entails:

$$\boxed{\widehat{\alpha}_1\stackrel{\mathbb{P}}{\longrightarrow}\tfrac{\sigma(1)}{\sigma(0)},\ \widehat{\alpha}_0\stackrel{\mathbb{P}}{\longrightarrow}(1-\alpha_1)\mu_0,\ \widehat{\sigma}_\varepsilon^2\stackrel{\mathbb{P}}{\longrightarrow}\sigma(0)(1-\alpha_1^2).} \qquad (28)$$

These results confirm the parametrizations in Table 2, implying that for $|\alpha_1|=1$ i.e. $(|\sigma(1)|=\sigma(0))$, $\widehat{\alpha}_0 \overset{\mathbb{P}}{\longrightarrow} 0$, $\widehat{\sigma}_\varepsilon^2 \overset{\mathbb{P}}{\longrightarrow} 0$, oppugning the nestedness presumption $\langle\mathbf{d}\rangle$, and affirming that $\alpha_1=1$ renders the covariance matrix in (23) singular. This result belies Kolmogorov's existence theorem rendering the distribution in (18) incongruous.

The results in (26)-(27) also pertain to the derivation of the non-standard sampling distributions since the distribution $f(\mathbf{y};\boldsymbol{\theta}_{AR})$, $\forall\mathbf{y}\in\mathbb{R}^n$, of the AR(1) model in Table 2 provides the cornerstone for deriving all relevant sampling distributions for inference via (15). Following White (1958), the unit-root literature sidesteps this touchstone procedure by using an offhand derivation based on the OLS estimator $\widehat{\alpha}_1$ whose scaled difference $g(n)(\widehat{\alpha}_1-\alpha_1)$ is inferred to converge in distribution (as $n\rightarrow\infty$) to (9), after (i) imposing $\alpha_1=1$ to transform $\{Y_t,\ t\in\mathbb{N}\}$ into a partial sum process $\{[Y_t-E(Y_t)],\ t\in\mathbb{N}\}$, where $E(Y_t)=\mu_0$, and (ii) invoking the functional central limit and the continuous function theorems (Hall and Heyde, 1980). Note that for $E(Y_t)=0$, and $E(Y_t)=\mu_0+\mu_1 t$, the same procedure would yield (8) and (10); see Choi (2015), pp. 18-19. That is, the derivation of these non-standard distributions is viewed as a mathematical deduction problem with presumptions $\langle\mathbf{a}\rangle$-$\langle\mathbf{e}\rangle$ the invoked premises. As described more broadly by Rao (2004): "... statistics as a deductive discipline of deriving consequences from given premises. The need for examining the premises $[\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})]$, which is important for practical applications of results of data analysis, is seldom emphasized."

Worse, the convergence in distribution of $n(\widehat{\alpha}_1-1)$ to (8)-(10) contradicts the convergence in probability results in (28) which obviate $\widehat{\alpha}_1 \overset{\mathbb{P}}{\rightarrow} 1$. The traditional approach reconciles this conflict by hypothesizing a 'quantum leap' between $\alpha_1=(1-\epsilon)$, for a tiny $\epsilon>0$, and $\alpha_1=1$ that transforms a stationary process $\{Y_t,\ t\in\mathbb{N}\}$ into a non-stationary one, fusing the AR(1) and UR(1) models, and transmuting the standard into non-standard distributions. A less embrangled account advocated next is that the two models are grounded on two distinct stochastic processes $\{Y_t,\ t\in\mathbb{N}\}$.

Potential objection from traditional unit-root testing

Due to unfamiliarity with the statistical perspective, an advocate of the D-F unit-root testing is likely to dispute the above direct derivation of the statistical model in Table 2, based on assuming that $\{Y_t,\ t\in\mathbb{N}\}$ is Normal (N), Markov (M) and Stationary (S) as a fabrication, despite the case made above (section 2.3) that these assumptions are in-built via [i]-[v] in conjunction with the equivalence (N, M, S) $\rightleftarrows$ [1]-[5].

A more familiar way to convince this advocate might be to use a traditional procedure to affirm the statistical AR(1) in Table 2 and lead us to the related UR(1) model. Such a procedure is the recursive substitution 'solution' of the AR(1)-(B) model, when viewed as a stochastic difference equation (Enders, 2004):

$$Y_t=\alpha_1^t Y_0+\alpha_0\left(\textstyle\sum_{i=0}^{t-1}\alpha_1^i\right)+\textstyle\sum_{i=0}^{t-1}\alpha_1^i\varepsilon_{t-i}. \tag{29}$$

Invoking presumption $\langle\mathbf{f}\rangle$ 'as $t\rightarrow\infty$', or more appositely, using $\langle\mathbf{e}\rangle^*$ Y_0 is a typical element of $\{Y_t,\ t\in\mathbb{N}\}$, (29) yields the 'asymptotic rendering' of $\boldsymbol{\varphi}$, comprising two non-nested and separate (Hatanaka, 1996) parametrizations (Appendix A, $\mu_1=\mu_2=0$):

AR(1): $\alpha_1<1$, $\boldsymbol{\varphi}_0:=[\alpha_0=\mu_0(1-\alpha_1),\ \alpha_1=\frac{\sigma(1)}{\sigma(0)},\ \sigma_0^2=\sigma(0)(1-\alpha_1^2)]$ (Table 2), and

UR(1): $\alpha_1=1$, $\boldsymbol{\varphi}_1:=(\mu_1,\ \sigma_1^2=\sigma(0))$; see Appendix A with $\mu_2=0$.

The bivariate distribution in (23) and the parametrization for (16)(29) can also be unveiled using (29); see Appendix A for the AR(1)-(C) model with $\mu_1=\mu_2=0$.

The probabilistic structure and the ensuing parametrizations implicitly imposed on $\{Y_t,\ t\in\mathbb{N}\}$ for the UR(1) model can also be unveiled by 'solving' (29) for $\alpha_1=1$:

$$Y_t=Y_0+\alpha_0 t+\sum_{i=0}^{t-1}\varepsilon_{t-i},$$

which, after invoking $\langle\mathbf{e}\rangle^*$ Y_0 is a typical element of $\{Y_t,\ t\in\mathbb{N}\}$, i.e. $E(Y_0)=\mu_0$, $Var(Y_0)=0$, gives rise to (Appendix A with $\mu_2=0$):

$$E(Y_t)=\mu_0+\mu_1 t,\ Cov(Y_t,Y_{t-i})=\sigma(0)(t-i),\ i=0,1, \tag{30}$$

where the notation $\alpha_0:=\mu_1$ and $\sigma_\varepsilon^2:=\sigma(0)$ is used to avoid conflating model parameters and moments of $\{Y_t,\ t\in\mathbb{N}\}$. (30) reveals the bivariate distribution $f(y_t,y_{t-1};\boldsymbol{\psi})$:

$$\begin{pmatrix} Y_t \\ Y_{t-1}\end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix}\mu_0+\mu_1 t \\ \mu_0+\mu_1(t-1)\end{pmatrix},\begin{pmatrix}\sigma(0)\cdot t & \sigma(0)\cdot(t-1) \\ \sigma(0)\cdot(t-1) & \sigma(0)\cdot(t-1)\end{pmatrix}\right), \tag{31}$$

$$(Y_t|Y_{t-1})\backsim\mathsf{N}\left(\mu_1+Y_{t-1},\sigma(0)\right),\ Y_{t-1}\backsim\mathsf{N}\left(\mu_0+\mu_1(t-1),\sigma(0)\right).\ t\in\mathbb{N}, \tag{32}$$

which gives rise to the UR(1) model in Table 3; see Spanos (1990a).

Table 3: Statistical Unit Root [UR(1)] model

Statistical GM:		$Y_t=\mu_1+Y_{t-1}+u_{1t},\ t\in\mathbb{N}.$	
[1]	Normality:	$(Y_t,Y_{t-1})\backsim\mathsf{N}(.,.),\ (y_t,y_{t-1})\in\mathbb{R}^2,$	$t\in\mathbb{N}.$
[2]	Linearity:	$E\left(Y_t|\sigma(Y_{t-1})\right)=\mu_1+Y_{t-1},$	
[3]	Homoskedasticity:	$Var\left(Y_t|\sigma(Y_{t-1})\right)=\sigma_1^2,$	
[4]	Markov:	$\{Y_t,\ t\in\mathbb{N}\}$ is a Markov process,	
[5]	t-invariance:	$\boldsymbol{\theta}_{UR}:=(\mu_1,\sigma_1^2)$ are *not* changing with t,	
Parametrization:		$\mu_1\in\mathbb{R},\ \sigma_1^2=\sigma(0)\in\mathbb{R}_+.$	

$f(y_t,y_{t-1};\boldsymbol{\psi})$ in (31) addresses the singularity problem with the AR(1) model in Table 2 for $\alpha_1=1$ since its covariance matrix is non-singular for all $t>1$, confirming the two distinct processes underlying the AR(1) and UR(1) models conjectured above.

A cursory glance at Tables 2 and 3 by an advocate of D-F unit-root testing might confer the impression that the AR(1) nests the UR(1) model when $\alpha_1=1$, due to the similitude of the two conditional distributions, (25) and (32). The similitude, however, is more apparent than real since Y_{t-1} is not weakly exogenous for $\boldsymbol{\theta}_{AR}$ or $\boldsymbol{\theta}_{UR}$, and thus the relevant probabilistic structure is that of their joint distributions in (23) and (31). A more heedful comparison reveals: (i) covariance stationarity [AR(1)] vs. partial sum heterogeneity [UR(1)], and (ii) their parametrizations, $\boldsymbol{\theta}_{AR}:=(\alpha_0,\alpha_1,\sigma_0^2)$ vs. $\boldsymbol{\theta}_{UR}:=(\mu_1,\sigma_1^2)$ with $\mu_1\neq\alpha_0,\ \sigma_1^2\neq\sigma_0^2$, (iii) rendering the two models non-nested.

The above derivations and arguments challenge the aPP presumptions $\langle\mathbf{b}\rangle$-$\langle\mathbf{f}\rangle$ since they run afoul the statistical stipulations $\langle\mathbf{b}\rangle^*$-$\langle\mathbf{e}\rangle^*$ based on the joint distribution of $\{Y_t,\ t\in\mathbb{N}\}$. In particular, $\alpha_1\in[-1,1]$ in $\langle\mathbf{b}\rangle$ belies the statistical $\alpha_1\in(-1,1)$ in $\langle\mathbf{b}\rangle^*$, which in turn renders $\langle\mathbf{d}\rangle$ incongruous since $\alpha_1=1$ runs afoul Kolmogorov's existence theorem. Also, presumption $\langle\mathbf{c}\rangle$ belies $\langle\mathbf{c}\rangle^*$ since the statistical error term is a model-specific Md process, and not a generic process for all AR(1) models in Table 1. Presumption $\langle\mathbf{e}\rangle$ belies $\langle\mathbf{e}\rangle^*$ since treating the initial Y_1 (or Y_0) unlike all other elements of $\{Y_t,\ t\in\mathbb{N}\}$ makes sense from the aPP perspective, but is unfounded on probability theory grounds. Also, $\langle\mathbf{e}\rangle^*$ renders the presumption $\langle\mathbf{f}\rangle$ 'as $t\rightarrow\infty$' uncalled-for since it obscures the fact that the t-varying problem of the recursive 'solution' in (29) is an artifact created by the aPP perspective; see Appendix A.

Key question answered. There is no stochastic process $\{Y_t,\ t\in\mathbb{N}\}$ satisfying Kolmogorov's existence theorem underlying the AR(1) in (16) that nests the UR(1) model when α_1=1. Instead, there are two different processes, underlying two non-nested but potentially partly overlapping models; see Vuong (1989), Hatanaka (1996).

3 The statistical AR(1) and UR(1) models

3.1 The statistical perspective on the AR(1)-(C) model

Example 3. For the statistical AR(1) model with a trend, $\{Y_t,\ t\in\mathbb{N}\}$ is assumed to be Normal (N), Markov (M), and covariance stationary, but mean-heterogeneous (1st order polynomial in t) (S(t)). When imposed on $\{Y_t,\ t\in\mathbb{N}\}$, these assumptions yield:

$$f(y_1,...,y_n;\boldsymbol{\psi})\overset{\text{M\&S}(t)}{=}f(y_1;\boldsymbol{\phi}_0)\prod_{t=2}^{n}f(y_t|y_{t-1};\boldsymbol{\phi}_1),\ \mathbf{y}\in\mathbb{R}^n, \tag{33}$$

where the statistical parameters $\boldsymbol{\phi}_1$ are derived from those of $f(y_t,y_{t-1};\boldsymbol{\phi})$ (Table 4).

Table 4: Bivariate Normal distribution of the AR(1) model

$$\begin{pmatrix} Y_t \\ Y_{t-1}\end{pmatrix}\backsim\mathsf{N}\left(\begin{pmatrix}\mu_0+\mu_1 t \\ \mu_0+\mu_1(t-1)\end{pmatrix},\begin{pmatrix}\sigma(0) & \sigma(1)\\ \sigma(1) & \sigma(0)\end{pmatrix}\right),$$

where $E(Y_t)=\mu_0+\mu_1 t,\ Var(Y_t)=\sigma(0),\ Cov(Y_t,Y_{t-1})=\sigma(1)$

Table 5: Statistical AR(1) model with a trend

Statistical GM:	$Y_t=\beta_0+\beta_1 t+\alpha_1 Y_{t-1}+u_{0t},\ t\in\mathbb{N}.$		
[1] Normality:	$(Y_t,Y_{t-1})\backsim\mathsf{N}(.,.),\ (y_t,y_{t-1})\in\mathbb{R}^2,$		
[2] Linearity:	$E\left(Y_t\|\sigma(Y_{t-1})\right)=\beta_0+\beta_1 t+\alpha_1 Y_{t-1},$		
[3] Homoskedasticity:	$Var\left(Y_t\|\sigma(Y_{t-1})\right)=\sigma_0^2,$	$t\in\mathbb{N}.$	
[4] Markov:	$\{Y_t,\ t\in\mathbb{N}\}$ is a Markov process,		
[5] t-invariance:	$\boldsymbol{\theta}_{AR}:=(\beta_0,\beta_1,\alpha_1,\sigma_0^2)$ are *not* changing with t,		

$\beta_0=(1-\alpha_1)\mu_0+\alpha_1\mu_1\in\mathbb{R},\ \beta_1=(1-\alpha_1)\mu_1\in\mathbb{R},\ \alpha_1=\frac{\sigma(1)}{\sigma(0)}\in(-1,1),\ \sigma_0^2=\sigma(0)(1-\alpha_1^2)\in\mathbb{R}_+.$

Table 5 specifies the statistical AR(1) model in terms of a complete, internally consistent, and testable set of probabilistic assumptions [1]-[5], including the statistical parameterization of $\boldsymbol{\theta}_{AR}:=(\beta_0,\beta_1,\alpha_1,\sigma_0^2)=\boldsymbol{\phi}_1$. As in Table 2, Y_{t-1} is not weakly exogenous for $\boldsymbol{\theta}_{AR}$, since $\boldsymbol{\phi}_1:=(\beta_0,\beta_1,\alpha_1,\sigma_0^2)$ and $\boldsymbol{\phi}_2:=(\mu_0,\mu_1,\sigma(0))$ are *not* variation-free (Engle et al., 1983), and thus the statistical AR(1) model is specified in terms of $f(y_t,y_{t-1};\boldsymbol{\phi})$. In direct analogy to Tables 2-3, the AR(1) model assumptions [1]-[5] in Table 5 are equivalent to the reduction assumptions (N,M,$\mathsf{S}(t)$).

The AR(1) model in Table 5 does not nest the UR(1) model since the value $\alpha_1 = Corr(Y_t, Y_{t-1}) = 1$ is inadmissible; it renders the covariance matrix in Table 4 singular. Equivalently, the aPP parameters $\boldsymbol{\varphi}$ in Table 1 for AR(1)-(C) model are not identified relative to the statistical parameters $\boldsymbol{\theta}_{AR}$ in Table 5; see Spanos (1990b). Also, the generic error $\varepsilon_t \backsim \mathsf{NIID}(0, \sigma_\varepsilon^2)$, $t \in \mathbb{N}$, becomes a model-specific statistical error term $\{(u_{0t}|\sigma(Y_{t-1}))\,,\ t \in \mathbb{N}\}$ since $u_{0t} = (Y_t - E(Y_t|\sigma(Y_{t-1})))$ implies that:

$$(u_{0t}|\sigma(Y_{t-1})) \backsim \mathsf{NMd}\,(0, \sigma_0^2)\,,\ t \in \mathbb{N}, \tag{34}$$

where 'NMd' stands for 'Normal Martingale difference'.

The log-likelihood function of the AR(1) model relating to (33) takes the form:

$$\ln L(\boldsymbol{\theta}_{AR}; \mathbf{y}_0) = \ln f(y_1; \boldsymbol{\varphi}_0) + \textstyle\sum_{t=2}^{n} \ln f(y_t|y_{t-1}; \boldsymbol{\theta}_{AR}),\ \boldsymbol{\theta}_{AR} \in [\mathbb{R}^2 \times (-1,1) \times \mathbb{R}_+], \tag{35}$$

which calls for reparametrizing $Y_1 \backsim \mathsf{N}(\mu_0 + \mu_1, \sigma(0))$ in terms of $\boldsymbol{\theta}_{AR}$ (Table 5):

$$f(y_1; \boldsymbol{\theta}_{AR}) = \left(\tfrac{(1-\alpha_1^2)}{2\pi\sigma_0^2}\right)^{\frac{1}{2}} \exp\left\{-\tfrac{(1-\alpha_1^2)[(1-\alpha_1)^2 y_1 - \beta_0(1-\alpha_1) + \beta_1(2\alpha_1 - 1)]^2}{2\sigma_0^2(1-\alpha_1)^4}\right\},\ y_1 \in \mathbb{R}. \tag{36}$$

Note the crucial importance of the parametrizations in $\boldsymbol{\theta}_{AR}$ in Table 5 and (36) which provide constraints, i.e. $|\alpha_1| < 1$, for the ML estimators based on (35).

Example 4. The statistical UR(1) model corresponding to the AR(1) in Table 5 arises from imposing Normal, Markov, partial-sum heterogeneity, and mean-heterogeneity (2nd order polynomial in t) on $\{Y_t,\ t \in \mathbb{N}\}$, which is confirmed in Appendix A using the recursive 'solution'. The underlying bivariate distribution, $f(y_t, y_{t-1}; \boldsymbol{\psi})$, in Table 6 (Spanos, 1990a), gives rise to the probabilistic reduction:

$$f(y_t, y_{t-1}; \boldsymbol{\psi}) = f(y_t|y_{t-1}; \boldsymbol{\psi}_1) \cdot f(y_{t-1}; \boldsymbol{\psi}_2),\ \forall\,(y_t, y_{t-1}) \in \mathbb{R}^2, \tag{37}$$

underlying the parametrization of the probabilistic UR(1) model in Table 7.

Table 6: Bivariate Normal distribution of the UR(1) model with a trend

$$\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_0 + \mu_1 t + \mu_2 t^2 \\ \mu_0 + \mu_1(t-1) + \mu_2(t-1)^2 \end{pmatrix}, \begin{pmatrix} \sigma(0) \cdot t & \sigma(0) \cdot (t-1) \\ \sigma(0) \cdot (t-1) & \sigma(0) \cdot (t-1) \end{pmatrix}\right),$$

where $E(Y_t) = \mu_0 + \mu_1 t + \mu_2 t^2$, $Var(Y_t) = \sigma(0) \cdot t$, $Cov(Y_t, Y_{t-1}) = \sigma(0) \cdot (t-1)$

Table 7: Statistical UR(1) model with a trend

Statistical GM:	$Y_t = \gamma_0 + \gamma_1 t + Y_{t-1} + u_{1t},\ t \in \mathbb{N}$.	
[1] Normality:	$(Y_t, Y_{t-1}) \backsim \mathsf{N}(.,.),\ (y_t, y_{t-1}) \in \mathbb{R}^2$,	$t \in \mathbb{N}$.
[2] Linearity:	$E\,(Y_t|\sigma(Y_{t-1})) = \gamma_0 + \gamma_1 t + Y_{t-1}$,	
[3] Homoskedasticity:	$Var\,(Y_t|\sigma(Y_{t-1})) = \sigma_1^2$,	
[4] Markov:	$\{Y_t,\ t \in \mathbb{N}\}$ is a Markov process,	
[5] t-invariance:	$\boldsymbol{\theta}_{UR} := (\gamma_0, \gamma_1, \sigma_1^2)$ are *not* changing with t,	
Parametrization:	$\gamma_0 = (\mu_1 - \mu_2) \in \mathbb{R},\ \gamma_1 = (2\mu_2) \in \mathbb{R},\ \sigma_1^2 = \sigma(0) \in \mathbb{R}_+$.	

In direct analogy to the AR(1) models in Tables 2 and 5, the reduction assumptions are equivalent to the model assumptions of the UR(1) in Table 7, and in view of its parametrization $\boldsymbol{\phi}_1 := (\gamma_0, \gamma_1, \sigma_1^2)$ and $\boldsymbol{\phi}_2 := (\mu_0, \mu_1, \mu_2, \sigma(0))$, Y_{t-1} is also not weakly exogenous for $\boldsymbol{\theta}_{UR}$=$\boldsymbol{\phi}_1$, since μ_1=μ_2=0 → γ_0=γ_1=0. Hence, the UR(1) model is specified in terms of the joint distribution $f(y_t, y_{t-1}; \boldsymbol{\phi})$. The entailed distribution of the statistical error u_{1t}= $(Y_t - E(Y_t|\sigma(Y_{t-1})))$ is: $(u_{1t}|\sigma(Y_{t-1})) \backsim \mathsf{NMd}\,(0, \sigma_1^2)\,,\ t\in\mathbb{N}$.

The log-likelihood function of the UR(1) model relating to (33) takes the form:

$$\ln L(\boldsymbol{\theta}_{UR}; \mathbf{y}_0) = \ln f(y_1; \boldsymbol{\theta}_{UR}) + \textstyle\sum_{t=2}^{n} \ln f(y_t|y_{t-1}; \boldsymbol{\theta}_{UR}),\ \ \boldsymbol{\theta}_{UR}\in[\mathbb{R}^2\times\mathbb{R}_+],$$

$$f(y_t|y_{t-1}; \boldsymbol{\theta}_{UR}) = (1/2\pi\sigma(0)) \exp\left\{-[(y_t - \gamma_0 - \gamma_1 t - y_{t-1})^2/2\sigma(0)\right\},$$

calling for reparametrizing $Y_1 \backsim \mathsf{N}(\mu_0+\mu_1+\mu_2,\ \sigma(0))$ in terms of $\boldsymbol{\theta}_{UR}$ (Table 7):

$$f(y_1; \boldsymbol{\theta}_{UR}) = (2\pi\sigma(0))^{-\frac{1}{2}} \exp\left\{-(y_1 - \gamma_0 - \gamma_1)^2/2\sigma(0)\right\},\ \ y_1\in\mathbb{R}.$$

3.2 Revisiting the Dickey-Fuller AR(1): (A)-(C) models

The statistical AR(1) and UR(1) models (with their parametrizations) relating to the AR(1): (A)-(C) in Table 1 are shown in Table 8, where [A] and [B] are special cases of [C] in Tables 5 and 7, arising by imposing zero restrictions on μ_0, μ_1, μ_2.

Table 8: Statistical AR(1) vs. UR(1) models relating to Table 1

[A]	$Y_t = \alpha_1 Y_{t-1} + u_{0t},\ u_{0t} \backsim \mathsf{NMd}(0, \sigma_0^2)$ $\alpha_1 = \frac{\sigma(1)}{\sigma(0)} \in (-1, 1),\ \sigma_0^2 = \sigma(0)(1-\alpha_1^2) \in \mathbb{R}_+,$	$\Delta Y_t = u_{1t},\ u_{1t} \backsim \mathsf{NMd}(0, \sigma_1^2)$ $\sigma_1^2 = \sigma(0) \in \mathbb{R}_+,$
[B]	$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + u_{0t}, u_{0t} \backsim \mathsf{NMd}(0, \sigma_0^2)$ $\alpha_0 = \mu_0(1-\alpha_1) \in \mathbb{R},\ \alpha_1 = \frac{\sigma(1)}{\sigma(0)} \in (-1, 1),$ $\sigma_0^2 = \sigma(0)(1-\alpha_1^2) \in \mathbb{R}_+,$	$\Delta Y_t = \mu_1 + u_{1t},\ u_{1t} \backsim \mathsf{NMd}(0, \sigma_1^2)$ $\mu_1 \in \mathbb{R},\ \sigma_1^2 = \sigma(0) \in \mathbb{R}_+,$
[C]	$Y_t = \beta_0 + \beta_1 t + \alpha_1 Y_{t-1} + u_{0t},\ u_{0t} \backsim \mathsf{NMd}(0, \sigma_0^2)$ $\beta_0 = (1-\alpha_1)\mu_0 + \alpha_1\mu_1 \in \mathbb{R},\ \beta_1 = (1-\alpha_1)\mu_1 \in \mathbb{R},$ $\alpha_1 = \frac{\sigma(1)}{\sigma(0)} \in (-1, 1),\ \sigma_0^2 = \sigma(0)(1-\alpha_1^2) \in \mathbb{R}_+,$	$\Delta Y_t = \gamma_0 + \gamma_1 t + u_{1t}, u_{1t} \backsim \mathsf{NMd}(0, \sigma_1^2)$ $\gamma_0 = (\mu_1 - \mu_2) \in \mathbb{R},\ \gamma_1 = (2\mu_2) \in \mathbb{R},$ $\sigma_1^2 = \sigma(0) \in \mathbb{R}_+.$

The statistical AR(1) and UR(1) models in Table 8 (i) are two distinct families of models, with different (ii) probabilistic assumptions, and (iii) parametrizations, (iv) rendering them non-nested, (v) whose respective likelihood-based inferential components are based on regular sampling distributions, including 'near' unit-root testing in the context of the AR(1) models in Table 8 based on the hypotheses:

$$H_0\!:\ \alpha_1 = (1-\epsilon),\ \text{vs.}\ H_1\!:\ |\alpha_1| < (1-\epsilon),\ \text{for}\ \epsilon > 0, \tag{38}$$

as evidenced by simulation in section 3.4. No additional D-F type restrictions on the constant and trend coefficients (Table 10) are needed in testing (38). These evidence suggests and the choice between the AR(1) and UR(1) models should be based primarily on statistical adequacy grounds, i.e. the validity their respective assumptions [1]-[5]. The key role of nestedness is demonstrated in section 4 where a well-defined family of non-stationary AR(1) models that actually nest the related UR(1) models is proposed in the context of which α_1=1 can be tested using likelihood-based tests grounded on standard sampling distributions.

Taken together, these results affirm that conundrums [C1]-[C2] stem primarily from the aPP AR(1) models (Table 1) being incongruous in a probabilistic sense (belying Kolmogorov's existence theorem), as well as statistically misspecified since the validity of the probabilistic assumptions [1]-[5] relating to $\{Y_t,\ t\in\mathbb{N}\}$ is invariably ignored due to [C4]. As argued above, failures that distort/belie the joint distribution $f(\mathbf{y};\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta$, of $\{Y_t,\ t\in\mathbb{N}\}$, would give rise to non-optimal inference procedures, and induce sizeable discrepancies between the actual and nominal error probabilities, undermining the reliability of any inferences; see Andreou and Spanos (2003).

3.3 Revisiting the Error–Autocorrelation (E-A) formulation

To bring out the crucial importance of the statistical parametrizations, let us revisit the Error-Autocorrelation (E-A) formulation of the AR(1) models in Table 1 (Fuller, 1976; Dickey, 1984; Bhargava, 1986). The E-A is framed in terms of $Y_t=E(Y_t)+u_t$ in conjunction with $\{u_t,\ t\in\mathbb{N}\}$ being a Markov process, known as the "components representation of a time series." (Schmidt and Phillips, 1992, p. 258):

$$\begin{aligned}&\text{E-A: AR(1)-(B):}\quad Y_t=\mu_0+u_t,\ u_t=\alpha_1 u_{t-1}+\varepsilon_t,\ \varepsilon_t\backsim\mathsf{NIID}(0,\sigma_\varepsilon^2),\ t\in\mathbb{N},\\&\text{E-A: AR(1)-(C):}\quad Y_t=\mu_0+\mu_1 t+u_t,\ u_t=\alpha_1 u_{t-1}+\varepsilon_t,\ \varepsilon_t\backsim\mathsf{NIID}(0,\sigma_\varepsilon^2),\ t\in\mathbb{N}.\end{aligned}\tag{39}$$

(39) yields the model-specific E-A parametrizations of $(\alpha_0,\beta_0,\beta_1)$ in Table 9, which coincide with the statistical parametrizations in Tables 2 and 5!

Table 9: Error-Autocorrelation (E-A) induced parametrizations	
AR(1)-(B)*:	$Y_t=[\overbrace{\mu_0(1-\alpha_1)}^{\alpha_0}]+\alpha_1 Y_{t-1}+\varepsilon_t,$
AR(1)-(C)*:	$Y_t=[\overbrace{\mu_0(1-\alpha_1)+\alpha_1\mu_1}^{\beta_0}]+[\overbrace{\mu_1(1-\alpha_1)}^{\beta_1}]t+\alpha_1 Y_{t-1}+\varepsilon_t.$

The E-A parametrizations in Table 9 raised questions about the pertinence of the Dickey and Fuller (1981), Table X, (p. 1070), on joint and conditional hypotheses in Table 10. Given that the E-A parametrization AR(1)-(B)*, $\alpha_1=1\rightarrow\alpha_0=0$, [T1], [T4]-[T5] are incongruous, and the same holds for [T2]-[T3] and [T6] since $\alpha_1=1\rightarrow\beta_1=0$.

Table 10: Dickey-Fuller Joint and Conditional tests	
[T1]	H_0: $\alpha_1=1$ & $\alpha_0=0$, vs. H_1: $\lvert\alpha_1\rvert<1$ or $\alpha_0\neq0$,
[T2]	H_0: $\alpha_1=1$ & $\beta_1=0$, vs. H_1: $\lvert\alpha_1\rvert<1$ or $\beta_1\neq0$,
[T3]	H_0: $\alpha_1=1$ & $\beta_0=\beta_1=0$, vs. H_1: $\lvert\alpha_1\rvert<1$, or $\beta_0\neq0$ or $\beta_1\neq0$,
[T4]	H_0: $\alpha_1=1$ & $\alpha_0=0$ vs. H_1: $\alpha_0\neq0$ given $\alpha_1\neq1$,
[T5]	H_0: $\alpha_1=1$ & $\alpha_0=0$, vs. H_1: $\alpha_0\neq0$ given $\lvert\alpha_1\rvert<1$,
[T6]	H_0: $\alpha_1=1$ & $\beta_1=0$ vs. H_1: $\beta_1\neq0$ given $\alpha_1\neq1$.

To avoid these incongruities, the E-A formulations of the AR(1) in Table 9 have been widely adopted by the unit-root testing literature; see Banerjee et al. (1993), Maddala and Kim, 1998, Choi (2015), Pesaran (2015) inter alia. As Schmidt and Phillips (1992) argue: "... the important attraction of this parameterization is that the meaning of the nuisance parameters μ_0 and μ_1 does not depend on whether the unit root hypothesis is true." (p. 259). Unfortunately, the E-A parametrizations (Table 9) relate only to the parameters $(\alpha_0, \beta_0, \beta_1)$ in Tables 2 and 5, but (a) ignoring the key parameters $(\alpha_1, \sigma_\varepsilon^2)$ and (b) retaining $\alpha_1 \in [-1, 1]$, renders the E-A parametrizations internally inconsistent, undermining the ensuing unit-root testing.

To unveil this inconsistency, consider first the case $\alpha_1 < 1$ (Anderson, 1971):

$$u_t = \alpha_1 u_{t-1} + \varepsilon_t \rightarrow \quad E(u_t^2) := \sigma(0) = \tfrac{\sigma_\varepsilon^2}{(1-\alpha_1^2)}, \; Cov(u_t, u_{t-1}) := \sigma(1) = \tfrac{\alpha_1 \sigma_\varepsilon^2}{(1-\alpha_1^2)}, \; t \in \mathbb{N}, \tag{40}$$

yielding $\sigma_\varepsilon^2 = \sigma(0)(1-\alpha_1^2) := \sigma_0^2$, $\alpha_1 = [\sigma(1)/\sigma(0)]$, affirming the parametrizations in Table 5 and oppugning nestedness, transforming the generic $\varepsilon_t \backsim \mathsf{NIID}(0, \sigma_\varepsilon^2)$ (Table 1) into a model-specific Martingale difference (Md) process $\{u_{0t} = (u_t - \alpha_1 u_{t-1}), \; t \in \mathbb{N}\}$:

$$(u_{0t} | \sigma(Y_{t-1})) \backsim \mathsf{NMd}\left(0, \sigma_0^2\right), \; t \in \mathbb{N}.$$

To derive the parametrization of σ_ε^2 for $\alpha_1 = 1$, the 'solution' $u_t = u_0 + \sum_{i=1}^t \varepsilon_t$ yields:

$$E(u_t | \sigma(u_{t-1}, ..., u_1)) = u_{t-1}, \; t \in \mathbb{N}, \text{ with } E(u_t) = 0, \; E(u_t u_{t-i}) = \sigma(i)(t-i), \; i = 0, 1,$$

where $\sigma_\varepsilon^2 := \sigma(0)$ is used to avoid confusion, $E(u_0) = E(u_0^2) = 0$ (Appendix A), i.e. $\{(u_t, \sigma(u_t | \sigma(u_{t-1}, ..., u_0)), \; t \in \mathbb{N}\}$ is a second-order Martingale process. Hence, $\{Y_t, \; t \in \mathbb{N}\}$ is a partial sum process (Table 6), with $Var(Y_t | \sigma(Y_{t-1})) = \sigma(0)$ (Table 7), and generic error $\varepsilon_t \backsim \mathsf{NIID}(0, \sigma_\varepsilon^2)$ is transformed into a model-specific Md process $u_{1t} = (u_t - u_{t-1})$:

$$(u_{1t} | \sigma(Y_{t-1})) \backsim \mathsf{NMd}\left(0, \sigma_1^2\right), \; t \in \mathbb{N}.$$

However, to retain the functional form of the autoregression with $\alpha_1 = 1$ requires that $E(Y_t)$ takes the form $E(Y_t) = \mu_0 + \mu_1 t + \mu_2 t^2$ (Table 6) since:

$$E(Y_t | \sigma(Y_{t-1})) = E(Y_t) - E(Y_{t-1}) + Y_{t-1} = [(\mu_1 - \mu_2)] + [2\mu_2] \, t = \gamma_0 + \gamma_1 t + Y_{t-1},$$

affirming $\boldsymbol{\theta}_{UR} := (\gamma_0, \gamma_1, \sigma_1^2)$ in Table 7. It indicates, however, a problem in the traditional pairing of the AR(1) and UR(1) models in Table 1 since the natural pairing is AR(1)-[B] & UR(1)-[A] and AR(1)-[C] & UR(1)-[B].

In summary, the E-A parametrizations in Table 9 are inconsistent since they ignore the (implicit) statistical parametrizations for $(\alpha_1, \sigma_\varepsilon^2)$ shown in Table 8.

3.4 Simulating the statistical AR(1) and UR(1) models

The main conclusions from the above discussion is that the statistical AR(1) and UR(1) models are two different models whose inferential components, based on their likelihood functions, revolve around regular sampling distributions and none of the conundrums [C2]-[C4] arises. A pragmatic approach to confirm these results is to use Monte Carlo simulation beginning with exploring the estimation of the statistical AR(1) and UR(1) models to confirm their statistical parametrizations $\boldsymbol{\theta}_{AR}$ and $\boldsymbol{\theta}_{UR}$, in Tables 5 and 7, respectively, and then delve into their testing components.

Given the non-weak exogeneity of Y_{t-1} for both models, the simulated data are generated via $f(y_t, y_{t-1}; \boldsymbol{\phi})$ in Tables 4 and 6, respectively, using pseudo-random numbers for $\mathbf{u}_t := (u_t, u_{t-1})^\top \backsim \mathsf{N}(\mathbf{0}, \mathbf{I})$, $t = 1, ..., n$, and applying the transformation $(\boldsymbol{\Lambda} \boldsymbol{u}_t + \boldsymbol{\mu}(t))$ where $\mathbf{V} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$ and $\boldsymbol{\Lambda}$ is a lower triangular matrix to generate:

$$\mathbf{Z}_t := (Y_t, Y_{t-1})^\top = (\boldsymbol{\Lambda} \mathbf{U}_t + \boldsymbol{\mu}(t)) \backsim \mathsf{N}(\boldsymbol{\mu}(t), \mathbf{V}), \; t = 1, 2, ..., n, \tag{41}$$

for N replications used to explore their estimation and testing inference procedures.

The values for the primary and model parameters in Table 11 are chosen to ensure that the 'true' AR(1) and UR(1) models are as similar as possible, but potentially distinguishable via the signs and magnitudes of β_1, γ_1 and $\sigma_1^2 > \sigma_0^2$.

Table 11: AR(1) vs. UR(1) statistical models (Tables 5 and 7)

	Primary parameters: $f(y_t, y_{t-1}; \boldsymbol{\phi})$					**Model param.:** $f(y_t \mid y_{t-1}; \boldsymbol{\phi}_1)$			
AR(1):	$\mu_0=2$	$\mu_1=.8$	$\mu_2=0$	$\sigma(0)=1.5$	$\sigma(1)=1.425$	$\alpha_1=1.0$	$\beta_0=.86$	$\beta_1=.04$	$\sigma_1^2=1.463$
UR(1):	$\mu_0=2$	$\mu_1=.8$	$\mu_2=-.04$	$\sigma(0)=1.5$	$\sigma(1)=1.5$	$\alpha_1=.95$	$\gamma_0=.84$	$\gamma_1=-.08$	$\sigma_0^2=.1.5$

Tables 12 and 13 show the simulation results for the statistical AR(1) and UR(1) models, respectively, based on N=10,000 replications of samples of size n=100. These results confirm the parametrizations $\boldsymbol{\theta}_{AR}$ and $\boldsymbol{\theta}_{UR}$ in Tables 5 and 7 since the overall means, $\frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}_k(\mathbf{y}_i)$, k=1, 2, 3, 4, of the approximate sampling distributions of the OLS estimators are close enough to the true values in Table 11, accounting for standard deviation and the finite sample biases; see Kiviet and Phillips (1992).

Table 12: OLS estimates of the AR(1)

True model: Y_t=.86+.04t+.95Y_{t-1}+u_{0t}, $u_{0t}\backsim$N(0, .1463)

Estim.	mean	std. dev.	min	max	skew.	kurtosis
$\widehat{\beta}_0$	.914	.149	.151	1.216	.607	4.617
$\widehat{\beta}_1$	.067	.046	.019	0.174	.916	4.548
$\widehat{\alpha}_1$	.899	.059	.594	1.008	−.914	4.587
$\widehat{\sigma}_0^2$	.142	.021	.072	0.193	.476	3.079

Table 13: OLS estimates of the UR(1)

True model: Y_t=.84−.08t+Y_{t-1}+u_{1t}, $u_{1t}\backsim$ N(0, 1.5)

Estim.	mean	std dev.	min	max	skew.	kurtosis
$\widehat{\gamma}_0$	.910	.367	−.444	1.636	0.085	3.032
$\widehat{\gamma}_1$	−.083	.014	−.145	−.033	0.003	3.107
$\widehat{\alpha}_1$	.999	.004	0.983	1.015	−.007	3.044
$\widehat{\sigma}_1^2$	1.456	.210	0.771	2.362	0.281	3.111

The improved accuracy of the UR(1) estimates (Table 13) stems from its higher rate of convergence in distribution $g(n)$ stemming from the presence of the unit root; see Hamilton (1994). Although the Maximum Likelihood estimates of the AR(1) parameters enjoy better accuracy than the OLS due to the built-in restrictions $(1-\alpha_1)$ and $(1-\alpha_1^2)>0$ in (36), the OLS results are reported to render the comparisons with the traditional unit-root literature more relevant.

Table 14 summarizes the approximate sampling distributions of the t-tests $\tau(\mathbf{y};\theta)=[(\widehat{\theta}-\theta^*)/SD(\widehat{\theta})]$, for θ^*-the true value, using the data in Table 11, and confirms the above comments. As expected, they are all approximately distributed as N(0, 1). The power $\mathcal{P}(\alpha_1)$ for $\alpha_1\in[.9, .99]$ of $\tau(\mathbf{y};\alpha_1)$, α=.05, reminds one of a typical t-test (Table 15) and not a traditional t-type unit root test, where "...(the) power is usually less than 30% for $\alpha_1 \in [0.90, 1.0)$ and $n=100$." Phillips and Xiao (1998), p. 22.

Table 14: $Y_t = .84 - .08t + Y_{t-1} + u_{1t},\ u_{1t} \backsim \mathsf{N}(0, 1.5)$**-simulated t-tests**						
Test stat.	mean	std dev.	min	max	skewness	kurtosis
$\tau(\mathbf{y};\gamma_0)$	$-.4013$	1.038	-4.124	3.742	.057	3.056
$\tau(\mathbf{y};\gamma_1)$	$-.0035$	.999	-3.680	4.170	$-.010$	3.053
$\tau(\mathbf{y};\alpha_1)$	$-.0008$	1.011	-4.118	4.065	.001	3.095

The results of the t-tests in Table 14 indicate that the AR(1) model affirms the unit root imposed by the UR(1) model and suggest that the standard t-test for:

$$H_0\colon\ \alpha_1 = (1-\epsilon),\ \text{vs.}\ H_1\colon\ \alpha_1 < (1-\epsilon),\ \text{for}\ \epsilon > 0, \tag{42}$$

is valid, and for the simulated data does *not* reject H_0 for $\epsilon \leq .001$ at traditional significance levels α. Table 15 shows that the power of the t-test is typical for any discrepancy in the interval $\alpha_1 \in [.9, 1)$.

Table 15: Power of $\tau(\mathbf{y};\alpha_1)$ for $\alpha=.05$							
α_1	.90	.92	.93	.95	.96	.97	.99
$\mathcal{P}(\alpha_1)$	.981	.962	.944	.889	.858	.821	.767

The above simulation results also suggest that testing 'near' unit–roots (Chan and Wei, 1987, Phillips, 2023) makes sense only for the one-sided hypotheses in (42) in the context of the AR(1) model, using standard t-type tests with sufficient power as shown in Tables 14-15. To affirm that, consider generating simulated data using the UR(1) in Table 13 and estimating the AR(1) $Y_t = \beta_0 + \beta_1 t + \alpha_1 Y_{t-1} + u_{0t}$, ensuring that the latter is statistically adequate, in the sense that thorough M-S testing indicates no significant departures from assumptions [1]-[4]; see Andreou and Spanos (2003). The results in Table 16 indicate that the estimated parameters of the AR(1) and UR(1) models are approximately identical (allowing for small n biases), with the AR(1) model affirming the unit root for any $\epsilon \leq .001$.

Table 16: $\Delta Y_t = .84 - .08t + u_{1t},\ u_{1t} \backsim \mathsf{N}(0, 1.5)$, simul. data	
UR(1):	$\Delta Y_t = \underset{(.250)}{.839} - \underset{(.004)}{.081}t + \widehat{u}_{1t},\ \widehat{\sigma}_1^2 = 1.472,$
AR(1):	$Y_t = \underset{(.388)}{.910} - \underset{(.012)}{.083}t + \underset{(.004)}{.999}Y_{t-1} + \widehat{u}_{0t},\ \widehat{\sigma}_0^2 = 1.456.$

3.5 AR(1) and UR(1) can be observationally equivalent

A closer look at the results in Table 16 provides strong hints that for α_1 close enough to unity, the AR(1) and the UR(1) models, although non-nested, can be (approximately) **observationally equivalent** in the following sense:

(i) $\alpha_1 = (1-\epsilon)$, for a small $\epsilon > 0$, (ii) the AR(1) and UR(1) models are non-nested, (iii) both models are statistical adequate for the same data $\mathbf{y}_0$, i.e. they account 'equally well' for the heterogeneity in data $\mathbf{y}_0$.

To explore that, let us simulate $\Delta Y_t = \mu_1 + u_{1t}$, , and estimate $Y_t = \beta_0 + \beta_1 t + \alpha_1 Y_{t-1} + u_{0t}$, ensuring that the latter is statistically adequate. The results in Table 17 suggest that the AR(1) and UR(1) models are (approximately) 'observationally equivalent' in the sense that they account for the heterogeneity in data $\mathbf{y}_0$ equally well.

Table 17: $\Delta Y_t = .800 + u_{1t},\ u_{1t} \backsim \mathsf{N}(0, 1.5)$, simul. data	
UR(1):	$\Delta Y_t = \underset{(.122)}{.799} + \underset{(.004)}{.000}t + \widehat{u}_{1t},\ \widehat{\sigma}_1^2 = 1.472,$
AR(1):	$Y_t = \underset{(.405)}{.836} + \underset{(.038)}{.080}t + \underset{(.056)}{.900}Y_{t-1} + \widehat{u}_{0t},\ \widehat{\sigma}_0^2 = 1.396.$

Surprisingly, the observational equivalence is retained even when the estimated AR(1) model is without a trend, affirming the unit root of the UR(1) model:

$$\text{UR(1): } \Delta Y_t = \underset{(.122)}{.799} + \underset{(1.213)}{\widehat{u}_{1t}}, \quad \text{AR(1): } Y_t = \underset{(.251)}{.876} + \underset{(.006)}{.999}Y_{t-1} + \underset{(1.213)}{\widehat{u}_{0t}}.$$

More surprising is the case where the simulated data are generated by the AR(1) in Table 12, with the OLS estimates of the AR(1) and the UR(1) models in Table 18 indicating an (approximate) observational equivalence since $\gamma_1 \simeq 0$ in the UR(1) model. This suggests that in such cases of (approximate) observational equivalence, the choice between the AR(1) and the appropriate UR(1) model should be based on other grounds, including goodness-of-fit/prediction, invariance to policy simulations.

Table 18: $Y_t = .860 + .04t + .95Y_{t-1} + u_{0t},\ u_{0t} \backsim \mathsf{N}(0, .1463)$	
AR(1):	$Y_t = \underset{(.181)}{.910} + \underset{(.042)}{.087}t + \underset{(.055)}{.897}Y_{t-1} + \widehat{u}_{0t},\ \widehat{\sigma}_0^2 = .1396,$
UR(1):	$\Delta Y_t = \underset{(.053)}{.801} + \underset{(.001)}{.000}t + \widehat{u}_{1t},\ \widehat{\sigma}_1^2 = .1491.$

It is important to emphasize that there is no incongruity between conditions (ii) and (iii) since (ii) the non-nestedness relates to the probabilistic structure of $\{Y_t,\ t \in \mathbb{N}\}$, and (iii) relates to the statistical adequacy of the AR(1) and UR(1) models for the same data $\mathbf{y}_0$. In light of that, the observational equivalence defined by (i)-(iii) is very different from the traditional 'near observational equivalence' since neither of the conditions (ii) are (iii) play any role in the latter; see Phillips and Perron (1988), Cochrane (1991), Blough (1992), Faust (1996) inter alia.

The above analytical and simulation results call into question the pertinence of D-F unit-root testing and suggest that the choice between the AR(1) and UR(1) models should be based on statistical adequacy grounds after thorough M-S testing does *not* detect any departures from their respective assumptions [1]-[5] for data $\mathbf{y}_0$; see Spanos (2018). When both models are statistically adequate for the same $\mathbf{y}_0$, one could choose between them on other grounds, including goodness of fit/prediction, substantive adequacy, and robustness to policy changes, or use non-nested testing; see Vuong (1989). Then again, one could test the unit root hypothesis α_1 in the context of a non-stationary AR(1) model that (actually) nests the related UR(1) model, proposed next, using a traditional likelihood-ratio and a score tests based on standard sampling distributions, free from conundrums [1]-[2].

4 A non-stationary AR(1) nesting a UR(1) model

To affirm the key role of well-grounded models, let $\{Y_t,\ t\in\mathbb{N}\}$ be Normal (N), Markov (M), mean heterogeneous ($\mu_0+\mu_1 t$) combined with separably heterogeneous covariance, giving rise to the bivariate distribution $f(y_t, y_{t-1}; \boldsymbol{\phi})$:

$$\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix} \backsim \mathsf{N}\left(\begin{bmatrix} \mu_0+\mu_1 t \\ \mu_0+\mu_1(t-1) \end{bmatrix}, \begin{bmatrix} \sigma(0)\cdot t & \sigma(1)\cdot(t-1) \\ \sigma(1)\cdot(t-1) & \sigma(0)\cdot(t-1) \end{bmatrix}\right),\ t\in\mathbb{N}, \tag{43}$$

yielding the non-stationary AR(1) model with a Trending Variance (TV) in Table 19.

This model differs from the AR(1) in Table 5 in three crucial respects: (i) $\alpha_1\in[0,1]$, (ii) it is non-stationary since $\sigma^2(t)$ is heterogenous, a convex combination of t, with weight $\alpha_1^2\in[0,1]$, and (iii) for $\alpha_1=1 \rightarrow a_0=\mu_1,\ \delta=0,\ \sigma^2(t)=\sigma(0)$, nesting:

$$\text{UR(1)}:\ \ Y_t=\mu_1+Y_{t-1}+u_{1t},\ (u_{1t}|\sigma(Y_{t-1}))\backsim\mathsf{NMd}(0,\sigma(0)),\ t\in\mathbb{N}. \tag{44}$$

Several special cases of the AR(1)-TV model (Table 19) follow by imposing zero restrictions on the primary parameters (μ_0, μ_1).

Table 19: Non-stationary AR(1)-TV model

Statistical GM:		$Y_t=\delta_0+\delta_1 t+\alpha_1 Y_{t-1}+u_{0t},\ t\in\mathbb{N}.$	
[1]	Normality:	$(Y_t, Y_{t-1}))\ \backsim\mathsf{N}(.,.),\ (y_t, y_{t-1})\in\mathbb{R}^2,$	$t\in\mathbb{N}.$
[2]	Linearity:	$E\left(Y_t\vert\sigma(Y_{t-1})\right)=a_0+\delta t+\alpha_1 Y_{t-1},$	
[3]	Heterogeneity:	$Var\left(Y_t\vert\sigma(Y_{t-1})\right)=\sigma(0)\left[\alpha_1^2+(1-\alpha_1^2)t\right]:=\sigma^2(t),$	
[4]	Markov:	$\{Y_t,\ t\in\mathbb{N}\}$ is a Markov process,	
[5]	t-invariance:	$\boldsymbol{\theta}:=(\delta_0, \delta_1, \alpha_1, \sigma(0))$ are *not* changing with t,	
	where $\delta_0=\mu_0(1-\alpha_1)+\alpha_1\mu_1\in\mathbb{R},\ \delta_1=(1-\alpha_1)\mu_1\in\mathbb{R},$ $\alpha_1=\frac{\sigma(1)}{\sigma(0)}\in[-1,1],\ \sigma^2(t)=\sigma(0)\left[\alpha_1^2+(1-\alpha_1^2)t\right]\in\mathbb{R}_+.$		

Model 1. Consider the special case of the AR(1)-TV in Table 19 when $\mu_0=\mu_1=0$:

$$Y_t=\alpha_1 Y_{t-1}+u_t,\ (u_t|\sigma(Y_{t-1}))\backsim\mathsf{NMd}(0,\sigma^2(t)),\ t\in\mathbb{N},$$
$$\alpha_1=\tfrac{\sigma(1)}{\sigma(0)}\in[-1,1],\ \sigma^2(t)=\sigma(0)\left[\alpha_1^2+(1-\alpha_1^2)t\right]\in\mathbb{R}_+.$$

Model 2. Consider the special case of the AR(1)-TV in Table 19 when $\mu_1=0$:

$$Y_t=\beta_0+\alpha_1 Y_{t-1}+u_t,\ (u_t|\sigma(Y_{t-1}))\backsim\mathsf{NMd}(0,\sigma^2(t)),\ t\in\mathbb{N},$$
$$\beta_0=(1-\alpha_1)\mu_0\in\mathbb{R},\ \alpha_1=\tfrac{\sigma(1)}{\sigma(0)}\in[-1,1],\ \sigma^2(t)=\sigma(0)\left[\alpha_1^2+(1-\alpha_1^2)t\right]\in\mathbb{R}_+.$$

Model 3. Consider the special case of the AR(1)-TV in Table 19 with $\mu_0=0$:

$$\begin{gathered} Y_t=\alpha_0+\delta t+\alpha_1 Y_{t-1}+u_t,\ (u_t|\sigma(Y_{t-1}))\backsim\mathsf{NMd}(0,\sigma^2(t)),\ t\in\mathbb{N}, \\ \alpha_0=\alpha_1\mu_1\in\mathbb{R},\ \delta=(1-\alpha_1)\mu_1\in\mathbb{R},\ \alpha_1=\tfrac{\sigma(1)}{\sigma(0)}\in[-1,1],\ \sigma^2(t)=\sigma(0)\left[\alpha_1^2+(1-\alpha_1^2)t\right]\in\mathbb{R}_+. \end{gathered} \tag{45}$$

A simple extension of the AR(1)-TV model in Table 19 is to increase the degree of the trend polynomial to any $q>1$, and the number of lags to $p>1$ using the *reparametrized* AR(p)-TV formulation, to isolate the (potential) unit root:

$$Y_t=\delta_0+\textstyle\sum_{k=1}^{q}\delta_k t^k+\alpha_1 Y_{t-1}+\sum_{i=1}^{p-1}\beta_i\Delta Y_{t-i}+u_t,\ \left(u_t|\sigma(\mathbf{Y}_{t-p}^0)\right)\backsim\mathsf{N}(0,\sigma^2(t)),\ t\in\mathbb{N},$$

where $\mathbf{Y}^0_{t-p}{:=}(Y_1,...,Y_p)$. It should be stressed that q and p are chosen on statistical adequacy grounds; the validity of the assumptions analogous to [1]-[5] (Table 19).

Although the primary reason for the statistical AR(1)-TV family of models is to shed light on the statistical conundrums [C1]-[C4] (section 1.3), the obvious question is whether these models could potentially be used to account for the chance regularities exhibited by real data. The t-plot of simulated data from the UR(1)-TV model with a trend (table 18, $\mu_0{=}2,\ \mu_1{=}.5$) in Figure A1 (Appendix C), when compared with the t-plot of the $\ln\mathrm{GDP}_t$ (US quarterly data, 1947-2007) in Figure A2, seems to exhibit very similar chance regularity patterns. The simulated data for the AR(1)-TV model 3 in (45) for $\alpha_1{\simeq}1$ in Figures A3-A4 indicate that the t-plots are very similar to fig. A1 but a bit more ragged. They become even more familiar for $\delta{=}0$, where the raggedness is more visible. Figures A5-A6 show the t-plot of simulated data based on $\mu_0{=}2,\ \mu_1{=}.5,\ \sigma(0){=}1$, with $\sigma(1){=}.90$ and $\sigma(1){=}.85$.

4.1 Estimation: AR(1)-TV and UR(1) models

The estimation of the family of AR(1)-TV and the corresponding UR(1) models is straightforward. Consider the problem of estimating AR(1)-TV and the corresponding UR(1) models analogous to the D-F (A)-(C) formulations under H_0: $\alpha_1{=}1$ and H_1: $\alpha_1{<}1$. The unrestricted (H_1) AR(1)-TV model 1 is:

$$\mathcal{M}_1\text{: } Y_t{=}\alpha_1 Y_{t-1}{+}u_t,\ u_t\backsim\mathsf{NMd}(0,\sigma^2(t)),\ \sigma^2(t){=}\sigma(0)\left[\alpha_1^2{+}(1{-}\alpha_1^2)t\right],\ t{\in}\mathbb{N}, \tag{46}$$

whose matrix formulation of (46), using the $\mathbf{y}{=}\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$ notation for $t{=}1,2,...,n$, is:

$$\begin{gathered}\mathbf{y}{=}\mathbf{y}_{-1}\alpha_1{+}\mathbf{u},\ \mathbf{y}{:=}(Y_1,Y_2,...,Y_n)^\top,\ \mathbf{X}{:=}(Y_0,Y_1,...,Y_{n-1})^\top\\ Cov(\mathbf{y}|\mathbf{X}){=}\boldsymbol{\Omega}(\alpha_1,\sigma(0)){=}\sigma(0){\cdot}\mathbf{V}(\alpha_1)\neq\sigma(0){\cdot}\mathbf{I}_n,\\ \mathbf{V}(\alpha_1){=}\mathrm{diag}\left[[\alpha_1^2{+}(1{-}\alpha_1^2)],[\alpha_1^2{+}2(1{-}\alpha_1^2)],...,[\alpha_1^2{+}n(1{-}\alpha_1^2)]\right].\end{gathered} \tag{47}$$

Given that $\boldsymbol{\Omega}(\alpha_1,\sigma(0)){=}\sigma(0){\cdot}\mathbf{V}(\alpha_1)$ the model in (46) is non-linear in the parameter α_1, and thus the Maximum Likelihood (ML) of estimation takes the form of Feasible Generalized Least Squares (FGLS)-type estimator:

$$\widehat{\alpha}_1{=}\left(\mathbf{X}^\top\mathbf{V}^{-1}(\widehat{\alpha}_1)\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{V}^{-1}(\widehat{\alpha}_1)\mathbf{y},\ \widehat{\sigma}(0){=}\tfrac{1}{n}\widehat{\mathbf{u}}^\top\mathbf{V}^{-1}(\widehat{\alpha}_1)\widehat{\mathbf{u}},\ \widehat{\mathbf{u}}{=}\mathbf{y}{-}\mathbf{X}\widehat{\alpha}_1,$$

derived by numerical optimization. The restricted (H_0) AR(1)-TV model 1 is:

$$\mathcal{M}_0\text{: } Y_t{=}Y_{t-1}{+}u_t,\ u_t\backsim\mathsf{NMd}(0,\sigma(0)),\ t{\in}\mathbb{N},$$

and thus the ML estimator of $\sigma(0)$ will be: $\widetilde{\sigma}(0){=}\frac{1}{n}\sum_{t=1}^n(\Delta Y_t)^2$.

The unrestricted (H_1) AR(1)-TV model 2 is:

$$\mathcal{M}_1\text{: } Y_t{=}\alpha_0{+}\alpha_1 Y_{t-1}{+}u_t,\ u_t\backsim\mathsf{NMd}(0,\sigma^2(t)),\ t{\in}\mathbb{N},$$

and thus, $\mathbf{X}{:=}(\mathbf{1},\mathbf{y}_{-1}),\ \mathbf{1}{:=}(1,1,...,1)^\top,\ \boldsymbol{\beta}{:=}(\alpha_0,\alpha_1)^\top$, using the $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}{+}\mathbf{u}$ framing.

The restricted (H_0) AR(1)-TV model 2 coincides with that of model 1. The unrestricted (H_1) AR(1)-TV model 3 is:

$$\mathcal{M}_1\text{: } Y_t{=}\beta_0{+}\delta t{+}\alpha_1 Y_{t-1}{+}u_t,\ u_t\backsim\mathsf{NMd}(0,\sigma^2(t)),\ t{\in}\mathbb{N},$$

and thus, $\mathbf{X}{:=}(\mathbf{1},\mathbf{t},\mathbf{y}_{-1}),\ \boldsymbol{\beta}{:=}(\beta_0,\delta,\alpha_1)$, where $\mathbf{t}{:=}(1,2,...,n)^\top$.

The restricted (H_0) AR(1)-TV model 3 is:

$$\mathcal{M}_0\text{: } Y_t{=}\mu_1{+}Y_{t-1}{+}u_t,\ u_t \backsim \mathsf{NMd}(0,\sigma(0)),\ t{\in}\mathbb{N},$$

and thus the ML estimators will be: $\widetilde{\mu}_1{=}\frac{1}{n}\sum_{t=1}^{n}\Delta Y_t$ and $\widetilde{\sigma}(0){=}\frac{1}{n}\sum_{t=1}^{n}(\Delta Y_t{-}\widetilde{\mu}_1)^2$.

It is important to emphasize that the appropriate model for unit-root testing is chosen on statistical adequacy grounds, i.e. its assumptions, analogous to [1]-[5] (Table 19), are valid for the particular data $\mathbf{y}_0$.

4.2 Unit-root testing in the context of an AR(1)-TV model

The AR(1)-TV model 3 in (45) with $\boldsymbol{\theta}{:=}(\mu_1,\alpha_1,\sigma(0))$, is used to derive a Likelihood Ratio (LR) and a Lagrange Multiplier (LM) (Score) test for the hypotheses:

$$H_0\text{: } \alpha_1{=}1 \text{ vs. } H_1\text{: } \alpha_1{<}1, \tag{48}$$

since it balances generality with fewer primary parameters rendering the Monte Carlo simulations that follow more informative.

4.2.1 A unit root Likelihood Ratio (LR) test

The LR unit root test of (48) in the context of the AR(1)-TV model 3 in (45) is:

$$\boxed{-2\ln\lambda(\mathbf{y}){=}n(\ln\widetilde{\sigma}^2{-}\ln\widehat{\sigma}^2(\widehat{\mu}_1,\widehat{\alpha}_1)){-}\textstyle\sum_{t=1}^{n}\ln(\widehat{\alpha}_1^2(1{-}t){+}t),\quad C_1{=}\{\mathbf{y}{:}{-}2\ln\lambda(\mathbf{y}){>}c_\alpha\};}$$

see Appendix B for the derivations. Its asymptotic distribution is $-2\ln\lambda(\mathbf{y}) \overset{H_0}{\underset{a}{\backsim}} \boldsymbol{\chi}^2(2)$ where $\alpha{=}\int_{c_\alpha}^{\infty}\psi(z)dz$, and $\psi(z)$ denotes the chi-square density function, under certain regularity conditions; see Gourieroux and Monfort (1996).

4.2.2 A unit root Lagrange Multiplier (LM) or Score test

For the AR(1)-TV model 3 in (45) consider a Lagrange Multiplier (Score) test for the hypotheses in (48) takes the form (Appendix B):

$$\boxed{LM(\mathbf{y}){=}\frac{\left(\frac{1}{\widetilde{\sigma}^2}\sum_{t=1}^{n}\widetilde{\psi}_t\widetilde{u}_t{+}\frac{1}{\widetilde{\sigma}^2}\sum_{t=1}^{n}\xi_t\left[\widetilde{u}_t^2{-}\widetilde{\sigma}^2\right]\right)^2}{\left[2\sum_{t=1}^{n}(\xi_t{-}\overline{\xi})^2{+}\left(\frac{1}{\widetilde{\sigma}^2}\right)\sum_{t=1}^{n}(\widetilde{\psi}_t{-}\overline{\psi})^2\right]},\quad C_1{=}\{\mathbf{y}\text{: } LM(\mathbf{y}){>}c_\alpha\},} \tag{49}$$

where $\psi_t{:=}[\mu_1(1{-}t){+}Y_{t-1}]$, $\xi_t{:=}(1{-}t)$, and $LM(\mathbf{y}) \overset{H_0}{\underset{a}{\backsim}} \boldsymbol{\chi}^2(1)$ under certain regularity conditions; see Gourieroux and Monfort (1996). A related form of the LM test, based on test statistic $\sqrt{LM(\mathbf{y})} \overset{H_0}{\underset{a}{\backsim}} \mathsf{N}(0,1)$, is often preferable for small samples since its asymptotic distribution is more accurate; see Kiviet and Phillips (1992).

The test $\sqrt{LM(\mathbf{y})}$ in (49) can be viewed as a combination of two significance tests. The first relates to testing H_0: $(\alpha_1{-}1){=}0$, vs. H_1: $(\alpha_1{-}1){\neq}0$, using:

$$(\Delta Y_t{-}\widetilde{\mu}_1)\ \ {=}(\alpha_1{-}1)\widetilde{\psi}_t{+}v_{1t},\ \widetilde{\psi}_t{:=}\left[\widetilde{\mu}_1(1{-}t){+}Y_{t-1}\right], \tag{50}$$

where the estimated coefficient of $\widetilde{\psi}_t$ is: $(\widehat{\alpha_1}{-}1){=}\frac{\sum_{t=1}^{n}\widetilde{\psi}_t\widetilde{u}_t}{\sum_{t=1}^{n}(\widetilde{\psi}_t{-}\overline{\psi})^2}{=}\frac{\sum_{t=1}^{n}(\widetilde{\psi}_t{-}\overline{\psi})(\Delta Y_t{-}\widetilde{\mu}_1)}{\sum_{t=1}^{n}(\widetilde{\psi}_t{-}\overline{\psi})^2}$, which coincides with the test statistic proposed by Schmidt and Phillips (1992).

The second component of $\sqrt{LM(\mathbf{y})}$ in (49) relates to another t-type test of the form: H_0: $c_1{=}0$, vs. H_1: $c_1{\neq}0$, based on the auxiliary regression:

$$(\Delta Y_t)^2{=}c_0{+}c_1(1{-}t){+}v_{2t},\ t{\in}\mathbb{N}, \tag{51}$$

which comprises the contribution relating to $\sigma^2(t)$, a key component of $\sqrt{LM(\mathbf{y})}$.

4.3 The power of the LR and LM tests

The AR(1)-TV model 3 in (45) differs from the traditional D-F AR(1) model in so far as it nests parametrically the corresponding UR(1) model. To bring out the direct connection between the power of the above LM and LR tests and the nestedness of the AR(1) and UR(1) models, we evaluate the actual type I error and power of these tests using Monte Carlo simulations.

The simulations revolve around $f(y_t, y_{t-1}; \boldsymbol{\phi})$ in (43) and are generated using a procedure analogous to that in (41) to engender:

$$\mathbf{Z}_t{:=}(Y_t, Y_{t-1})^{\top} \backsim \mathsf{N}\left(\boldsymbol{\mu}(t), \mathbf{V}(t)\right), \ t{\in}\mathbb{N}. \tag{52}$$

For comparison purposes, the size and power of the D-F tests are also evaluated as an 'indicator' of the more elaborate variations/extensions of the traditional unit-root tests; see Banerjee et al. (1993), Choi (2015), Pesaran (2015). It should be emphasized that the extensions assigning an ARMA(p,q) type structure to the error term are beside the point since a crucial precondition for the validity of any test is the statistical adequacy of the invoked statistical model for the particular data $\mathbf{y}_0$, which renders such conjectural extensions irrelevant for the simulations the follow.

To bring out the crucial importance of using ML estimation, the histogram of the MLE estimator of $\alpha_1{=}1$ is given in figure A9 (Appendix C) and contrasted with that of the OLS estimator in Figure A10, the latter being considerably more dispersed. The smoothed histograms of the LR (figure A11) and $\sqrt{\text{LM}}$ (figure A12, Appendix C) statistics indicate good approximations of their respective asymptotic sampling distributions, ensuring reliable approximations for the relevant error probabilities.

Table 20 summarizes the size and power properties of the LR, the $\sqrt{\text{LM}}$ and the Dickey-Fuller (D-F) unit-root tests. Note that the D-F unit root tests in Table 20 are based on the aPP AR(1) model with a trend. In contrast, the proposed LR and LM unit root tests are based on the AR(1)-TV model 3 in (45).

$Y_t{=}a_0{+}\delta t{+}\alpha_1 Y_{t-1}{+}u_t, \ u_t{\backsim}\mathsf{NMd}(0, \sigma^2(t)), \ t{\in}\mathbb{N}.$

Table 20: Power of LR, $\sqrt{\text{LM}}$ and D-F tests*

	LR test				LM test				D-F
$\alpha_1 \diagdown \alpha$	.01	.025	.05	.10	.01	.025	.05	.10	.05
$\alpha_1{=}1$	.011	.024	.046	.09	.009	.024	.051	.103	.052
$\alpha_1{=}.99$	.407	.537	.642	.745	.343	.503	.638	.769	.041
$\alpha_1{=}.98$	.733	.828	.886	.935	.624	.767	.861	.931	.046
$\alpha_1{=}.95$	.957	.978	.987	.994	.863	.934	.968	.988	.078
$\alpha_1{=}.90$	.992	.997	.999	.999	.927	.971	.989	.996	.189
$\alpha_1{=}.85$	.998	.999	.999	1.00	.944	.978	.992	.997	.381
$\alpha_1{=}.80$	1.0	1.0	1.0	1.0	.948	.980	.994	.998	.621

$^*\mu_1{=}.4, \ n{=}100, \ N{=}20000$

(i) All three tests have good size (observed α) properties since the nominal and actual error probabilities are close.

(ii) The LR unit-root test shows marginally better power properties than the $\sqrt{\text{LM}}$ test primarily because the LR is constructed using information relating to both the null H_0: α_1=1 and the alternative H_1: α_1<1 hypotheses.

(iii) Both the LR and $\sqrt{\text{LM}}$ unit-root tests exhibit the power of chi-square and Normal-type tests under NIID, which is far superior to that of the D-F test.

(iv) The second component $\sum_{t=1}^{n}\xi_t(\widetilde{u}_t^2-\widetilde{\sigma}^2)$ of the $\sqrt{\text{LM}}$ test in (49) can be used to explain why the power of the D-F unit-root test is similar to the D-F (B) in Table 1, suggesting that the presence of $\sigma^2(t)=\sigma(0)\,[\alpha_1^2+(1-\alpha_1^2)t]$ has no sizeable effect on the power properties of the D-F test, as expected.

(v) To forfend any suspicions that the high power of the LR and $\sqrt{\text{LM}}$ stems from the presence of the trend, the size and power properties of the LR and D-F unit-root tests based on the AR(1)-TV model 2 (no trend) are compared in Tables 21-22.

$Y_t=\alpha_0+\alpha_1 Y_{t-1}+u_t,\ u_t\backsim \mathsf{NMd}(0,\sigma^2(t)),\ t\in\mathbb{N},$

Table 21: Power of the LR test*

$\alpha_1\diagdown\alpha$	.01	.025	.05	.10
α_1=1.0	.012	.027	.049	.097
α_1=.99	.415	.542	.643	.744
α_1=.98	.736	.829	.888	.934
α_1=.97	.875	.927	.956	.976
α_1=.96	.933	.963	.979	.990
α_1=.95	.960	.979	.989	.994
α_1=.9	.993	.996	.998	.999
α_1=.85	.998	.999	1.00	1.00

*N=40000, n=100; μ_1=2;
$E(Y_0)=\mu=2;\ Var(Y_0)=0;\ \sigma(0)=1.$

Table 22: Power of the D-F test*

$\alpha_1\diagdown\alpha$	.01	.025	.05	.10
α_1=1.	.010	.024	.051	.098
α_1=.99	.011	.027	.056	.113
α_1=.98	.013	.034	.066	.132
α_1=.95	.026	.063	.125	.232
α_1=.90	.090	.196	.330	.523
α_1=.80	.525	.746	.882	.966
α_1=.70	.935	.986	.997	.999
α_1=.60	.998	1.00	1.00	1.00

*N=40000, n=100; μ=2;
$\alpha_0=(1-\alpha_1)\mu;\ \sigma(0)=1.$

In conclusion, it should be noted that the above simulation results depend, to a certain extent, on the design and the parameter values chosen. However, after varying the parameter values within reasonable limits, the above conclusions turned out to be sufficiently robust to be considered typical.

5 Summary and conclusions

A case has been made above that using the aPP perspective to guide modeling inference with data $\mathbf{y}_0$ for the AR(1): (A)-(C) models in Table 1 has given rise to several conundrums, including [C1] non-standard distributions and [C2] the low power of D-F type unit root tests for α_1 near 1. The above analytical and simulation results suggest that these conundrums stem primarily from the aPP AR(1) models in Table 1 being probabilistically incongruous (belying Kolmogorov's existence theorem), as well as statistically misspecified since the validity of the probabilistic assumptions [1]-[5] of the statistical AR(1) and UR(1) models in Table 8 is invariably ignored. These flaws distort/belie the joint distribution $f(\mathbf{y};\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta$, of $\{Y_t,\ t\in\mathbb{N}\}$, giving rise to non-optimal inference procedures and inducing sizeable discrepancies between the actual and nominal error probabilities, undermining the reliability of inference.

The statistical perspective, providing the missing link between the AR(1) models in Table 1 and the real-world mechanism that generated data $\mathbf{y}_0$, reveals that failure to satisfy the existence theorem implies that no single processes $\{Y_t,\ t\in\mathbb{N}\}$ that underlies both the AR(1) and UR(1) models in Tables 1 exists. Instead, there are two different processes underlying each of the AR(1): (A)-(C) models, giving rise to two distinct and non-nested statistical AR(1) and UR(1) models. Their likelihood-based inferential components are grounded on standard sampling distributions, and none of the conundrums [C2]-[C4] arises. This suggests that the choice between the AR(1) and UR(1) models should be based on statistical adequacy grounds. It turns out, however, that the two models can be observationally equivalent for α_1 near 1, when both models are statistically adequate for the same data $\mathbf{y}_0$, in which case the choice could be made on other grounds, including goodness-of-fit/prediction and substantive adequacy. When one of the proposed non-stationary AR(1) models that nest the corresponding UR(1) models turns out to be statistically adequate for data $\mathbf{y}_0$, α_1=1 can be tested using likelihood-based tests free from conundrums [C1]-[C2].

The above results also call for a re-evaluation of the empirical modeling and inference, not only of the AR(p) model but also that of the statistical VAR(p) model (Spanos, 1990a) as it relates to stipulations [S1]-[S2], as well as a re-assessment of the cointegration literature using the statistical perspective. Such re-evaluations are likely to be strongly resisted because the call for supplementing the aPP with the statistical perspective on modeling and inference with data $\mathbf{y}_0$ appears superfluous from their aPP standpoint. However, the statistical perspective is crucial when any aPP model, $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{y})$, $\boldsymbol{\varphi}\in\Phi$, is used for inference, since what matters for that is: (a) the process $\{Y_t,\ t\in\mathbb{N}\}$ underlying $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y})$, $\boldsymbol{\theta}\in\Theta$, is well-defined, and (b) its probabilistic assumptions are valid for data $\mathbf{y}_0$, which are sidestepped by the aPP perspective.

References

[1] Anderson, T.W. (1971) *The Statistical Analysis of Time Series*, Wiley, NY.

[2] Andreou, E. and A. Spanos (2003) "Statistical adequacy and the testing of trend versus difference stationarity", *Econometric Reviews*, 22(3): 217-237.

[3] Banerjee, A., J. Dolado, J.W. Galbraith, and D.F. Hendry (1993) *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press, Oxford.

[4] Bhargava, A. (1986) "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 52: 369-384.

[5] Blough, S.R. (1992) "The relationship between power and level for generic unit root tests in finite samples", *Journal of Applied Econometrics* 7: 295-308.

[6] Billingsley, P. (1995) *Probability and Measure*, 3rd edition, Wiley, NY.

[7] Box G. and G. Jenkins (1970) *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA.

[8] Chan, N.H., and C.Z. Wei (1987) "Asymptotic inference for nearly nonstationary AR (1) processes", *The Annals of Statistics*, 15: 1050-1063.

[9] Choi, I. (2015) *Almost All About Unit Roots: Foundations, Developments, and Applications*, Cambridge University Press, Cambridge.

[10] Cochrane, J.H. (1991) "A critique of the application of unit root tests", *Journal of Economic Dynamics and Control,* 15, 275-284.

[11] Cox, D.R. (1990) "Role of Models in Statistical Analysis," *Statistical Science*, 5: 169-174.

[12] Dickey, D.A. (1984) "Powers of unit root tests", Proceedings of the *Business and Economics Sectio*n, *American Statistical Association*, 79: 489–93.

[13] Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74: 427-431.

[14] Dickey, D.A. and W.A. Fuller (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49: 1057-1072.

[15] Diebold, F.X. and M. Nerlove (1990) "Unit Roots in Economics Time Series: a selective survey", *Advances in Econometrics*, 8: 3-69.

[16] Donsker, M. (1951) "An invariance principle for certain probability limit theorems", *Memoirs of the American Mathematical Society*, 6: 1-12.

[17] Enders, W. (2004) Applied Econometric Time Series. Wiley, NJ.

[18] Engle, R.F. and C.W.J. Granger (1987) " Cointegration and error correction: representation, estimation and testing," *Econometrica*, 55: 251-276.

[19] Engle, R.F., D.F. Hendry and J.-F. Richard (1983) "Exogeneity", *Econometrica*, 51: 277-304.

[20] Faust, J. (1996) "Near Observational Equivalence and Theoretical Size Problems with Unit Root Tests", *Econometric Theory*, 12(4): 724-731.

[21] Fisher, R.A. (1922) "On the mathematical foundations of theoretical statistics", *Philosophical Transactions of the Royal Society A*, 222: 309-368.

[22] Fuller, W.A. (1976) *An Introduction to Statistical Time Series*, Wiley, NY.

[23] Gourieroux, C. and A. Monfort (1996) *Statistics and Econometric Models*, Volume 2, Cambridge University Press, Cambridge.

[24] Hall, P. and C.C. Heyde, 1980) *Martingale Limit Theory and Application*, Academic Press, NY.

[25] Hamilton, J.D. (1994) *Time Series Analysis*, Princeton University Press, NJ.

[26] Hansen, B.E. (1995) "Rethinking the univariate approach to unit root tests: how to use covariates to increase power", *Econometric Theory,* 11: 1148-1171.

[27] Hatanaka, M. (1996) *Time-series-based Econometrics: Unit Roots and Cointegration,* Oxford Uninversity Press, Oxford.

[28] Hendry, D.F. (1995) *Dynamic Econometrics*, Oxford Uninversity Press, Oxford.

[29] Hespanha, J.P. (2018) *Linear Systems Theory*, 2nd ed., Princeton University Press, Princeton, NJ.

[30] Juselius, K. (2006) *The Cointegrated VAR Model: methodology and applications*, Oxford University Press, Oxford.

[31] Johansen, S. (1991) "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models", *Econometrica*, 59: 1551-1580.

[32] Kiviet, J. and G.D.A. Phillips (1992) "Exact Similar Tests for Unit Roots and Co-integration," *Oxford Bulletin of Economics and Statistics*, 54: 349-367.

[33] Kolmogorov, A.N. (1933) *Foundations of the theory of Probability*, 2nd English edition, Chelsea Publishing Co. NY.

[34] Leybourne, S.J. (1995), "Testing for unit roots using forward and reverse Dickey-Fuller regressions," *Oxford Bulletin of Economics and Statistics*, 57: 559-571.

[35] Lutkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis*, Springer, NY.

[36] Maddala, G.S. and I-M. Kim (1998) *Unit roots, Cointegration, and Structural Change*, Cambridge University Press, Cambridge.

[37] Mann, H.B., and A. Wald (1943) "On the statistical treatment of linear stochastic difference equations", *Econometrica*, 9: 173-220.

[38] Martin, V., S. Hurn, and D. Harris (2013) *Econometric modelling with time series: specification, estimation and testing*, Cambridge University Press, Cambridge.

[39] McCullagh, P. (2002) "What is a statistical model?" *Annals of Statistics*, 30: 1225-1267.

[40] Mills, T.C. (2011) *The Foundations of Modern Time Series Analysis*, Palgrave MacMillan, London.

[41] Müller, U.K. and G. Elliott (2003) "Tests for Unit Roots and the Initial Condition", *Econometrica*, 71(4): 989–1305.

[42] Pesaran, M.H. (2015) *Time Series and Panel Data Econometrics*, Oxford University Press, Oxford.

[43] Phillips, P.C.B. (1987) "Time series regressions with a unit root," *Econometrica*, 55: 227-301.

[44] Phillips, P.C.B. (2023) "Estimation and inference with near unit roots," *Econometric Theory*, 39(2): 221-263.

[45] Phillips, P.C.B. and P. Perron (1988) "Testing for a unit root in time series regression", Biometrika, 75(2): 335-346.

[46] Phillips, P.C.B. and Z. Xiao (1998) "A Primer on Unit Root Testing", *Journal of Economic Surveys*, 12: 423-470.

[47] Rao, C.R. (2004) "Statistics: Reflections on the Past and Visions for the Future," *Amstat News*, 327: 2-3.

[48] Renyi, A. (1970) *Foundations of Probability*, Holden-Day, San Frascisco, CA.

[49] Said, S.E. and D.A. Dickey (1984) "Testing for unit roots in autoregressive-moving average models of unknown order", *Biometrika*, 71(3): 599-607.

[50] Schmidt, P. and P.C.B. Phillips (1992) "LM test for a unit root in the presence of deterministic trends," *Oxford Bulletin of Economics and Statistics*, 54: 257-287.
[51] Shiryaev, A.N. (1996) *Probability,* 2nd edition, Springer, NY.
[52] Slutzky, E. (1927/1937) "The summation of random causes as the source of cyclic process", English translation of the 1927 paper, *Econometrica*, 1937, 5: 105-146.
[53] Spanos, A. (1986) *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge.
[54] Spanos A. (1990a) "Unit Roots and their Dependence on the Implicit Conditioning Information Set", *Advances in Econometrics*, 8: 271-292.
[55] Spanos, A. (1990b) "The Simultaneous Equations Model revisited: statistical adequacy and identification", *Journal of Econometrics*, 44: 87-108.
[56] Spanos, A. (1995) "On Normality and the Linear Regression model", *Econometric Reviews*, 14: 195-203.
[57] Spanos, A. (2006) "Where Do Statistical Models Come From? Revisiting the Problem of Specification", pp. 98-119 in *Optimality: The Second Erich L. Lehmann Symposium*, edited by J. Rojo, Lecture Notes-Monograph Series, vol. 49, Institute of Mathematical Statistics.
[58] Spanos, A. (2010) "Statistical Adequacy and the Trustworthiness of Empirical Evidence: Statistical vs. Substantive Information", *Economic Modelling*, 27: 1436–1452.
[59] Spanos, A. (2018) "Mis-Specification Testing in Retrospect", *Journal of Economic Surveys*, 32(2): 541–577.
[60] Spanos, A. (2019) *Introduction to Probability Theory and Statistical Inference: Empirical Modeling with Observational Data*, 2nd edition, Cambridge University Press, Cambridge.
[61] Vuong, Q. H. (1989) "Likelihood ratio tests for model selection and non-nesteed hypotheses," *Econometrica*, 57, 307-333.
[62] White, J.S. (1958) "The limiting distribution of the serial correlation coefficient in the explosive case", *The Annals of Mathematical Statistics*, 29: 1188-1197.
[63] Wold, H., (1938) *A Study in the Analysis of Stationary Time Series* (Doctoral dissertation), Almqvist & Wiksell, Uppsala.
[64] Yule, G. U. (1927) "On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers", *Philosophical Transactions of the Royal Society of London*, Series A, 226: 267-298.
[65] Yule, G.U. and M.G. Kendall (1950) *An Introduction to the Theory of Statistics*, 14th ed., Griffin, London.

6 Appendix A: The 'asymptotic rendering' of AR(1)

To avoid repetition, the discussion focuses on the AR(1)-(C) model:

$$Y_t{=}\beta_0{+}\beta_1 t{+}\alpha_1 Y_{t-1}{+}\varepsilon_t,\ \ \varepsilon_t \backsim \mathsf{NIID}(0,\sigma_\varepsilon^2),\ t{\in}\mathbb{N}, \tag{53}$$

since the other two models (A) and (B) are special cases arising by imposing zero restrictions on (μ_0, μ_1, μ_2). The stochastic difference equation in (53) is 'solved using recursive substitution to yield (Hamilton, 1994, Enders, 2004):

$$\boxed{Y_t = \alpha_1^t Y_0 + \beta_0 \left(\sum_{i=0}^{t-1} \alpha_1^i\right) + \beta_1 \left(\sum_{i=0}^{t-1} \alpha_1^i (t-i)\right) + \sum_{i=0}^{t-1} \alpha_1^i \varepsilon_{t-i},\ t \in \mathbb{N}.} \tag{54}$$

This can be used in conjunction with assumptions [i]-[iv] in (17) to derive:

$$\begin{aligned} &E(Y_t) = \alpha_1^t E(Y_0) + \beta_0 \left(\sum_{i=0}^{t-1} \alpha_1^i\right) + \beta_1 \left(\sum_{i=0}^{t-1} \alpha_1^i (t-i)\right), \\ &Cov(Y_t, Y_{t+\tau}) = \alpha_1^{2t+|\tau|} Var(Y_0) + \sigma^2 \alpha_1^{|\tau|} \left(\sum_{i=|\tau|}^{t-1} \alpha_1^{2i}\right),\ |\tau| \geq 0. \end{aligned} \tag{55}$$

It turns out that the summations yield different answers for $|\alpha_1| < 1$ and $|\alpha_1| = 1$.

6.1 The AR(1) model: $|\alpha_1| < 1$

Evaluating the summations in (55) for $|\alpha_1| < 1$ yields:

$$\sum_{i=0}^{t-1} \alpha_1^i = \frac{(1-\alpha_1^t)}{1-\alpha_1},\ \ \sum_{i=0}^{t-1} \alpha_1^i (t-i) = \frac{(\alpha_1^{t+1} - \alpha_1) + (1-\alpha_1)t}{(1-\alpha_1)^2},\ \ \sum_{i=0}^{t-1} \alpha_1^{2i} = \frac{(1-\alpha_1^{2t})}{1-\alpha_1^2}, \tag{56}$$

giving rise to the first two moments of $\{Y_t,\ t \in \mathbb{N}\}$:

$$\begin{aligned} E(Y_t) &= \alpha_1^t E(Y_0) + \beta_0 (\tfrac{1-\alpha_1^t}{1-\alpha_1}) + \tfrac{\beta_1 \alpha_1 (\alpha_1^t - 1)}{(1-\alpha_1)^2} + \tfrac{\beta_1 t}{(1-\alpha_1)}, \\ Cov(Y_t, Y_{t+\tau}) &= \alpha_1^{2t+|\tau|} Var(Y_0) + \sigma_0^2 \alpha_1^{|\tau|} [(1-\alpha_1^{2t}) / (1-\alpha_1^2)],\ |\tau| \geq 0, \end{aligned} \tag{57}$$

indicating that both moments are t-varying, rendering $\{Y_t,\ t \in \mathbb{N}\}$ nonstationary.

The conventional wisdom, going back to Wold (1938), invokes the stipulation $\langle \mathbf{f} \rangle$ 'as $t \to \infty$' (Diebold and Nerlove, 1990):

$$\lim_{t\to\infty} \sum_{i=0}^{t-1} \alpha_1^i = \frac{1}{1-\alpha_1},\ \ \lim_{t\to\infty} \sum_{i=0}^{t-1} \alpha_1^i (t-i) = \frac{(1-\alpha_1)t - \alpha_1}{(1-\alpha_1)^2},\ \ \lim_{t\to\infty} \sum_{i=0}^{t-1} \alpha_1^{2i} = \frac{1}{1-\alpha_1^2},$$

to yield the so-called 'asymptotic rendering' of the moments in (57) (Hamilton, 1994):

$$\lim_{t\to\infty} E(Y_t) = \frac{\beta_0}{1-\alpha_1} - \frac{\beta_1 \alpha_1}{(1-\alpha_1)^2} + \frac{\beta_1 t}{(1-\alpha_1)},\ \ \lim_{t\to\infty} Var(Y_t) = (\frac{\sigma^2}{1-\alpha_1^2}),\ \ \lim_{t\to\infty} Cov(Y_t, Y_{t+1}) = \sigma^2 (\frac{\alpha_1}{1-\alpha_1^2}).$$

The above resolution of the t-varying moments problem is contrived and probabilistically ad hoc. It inserts (unnecessarily) Y_0 into the sample $\mathbf{Y}$, and then replaces the enlarged index set $\mathbb{N}_{\{0\}} := (0, 1, 2, ..., n, ...)$ with $\mathbb{Z} := (0, \pm 1, \pm 2, ..., \pm n, ...)$, which seems particularly dubious for economic time series. Worse, it has nothing to do with the traditional asymptotics 'as $n \to \infty$', since $t \to \infty$ takes the index t backwards to $-\infty$, beclouding the fact that the t-varying problem in (57) is an artifact stemming from misusing the aPP perspective to guide empirical modeling and inference.

For a pertinent probabilistic 'solution' to the t-varying problem consider a simple rearrangement of the terms in (57) to separate α_1^t, shown below:

$$\begin{aligned} E(Y_t) &= \alpha_1^t E(Y_0) + (\tfrac{\beta_0}{1-\alpha_1}) - \tfrac{\beta_0 \alpha_1^t}{1-\alpha_1} - \tfrac{\beta_1 \alpha_1}{(1-\alpha_1)^2} + \tfrac{\beta_1 \alpha_1^{t+1}}{(1-\alpha_1)^2} + \tfrac{\beta_1 t}{(1-\alpha_1)} = \\ &= \boxed{\alpha_1^t \left[E(Y_0) - (\tfrac{\beta_0}{1-\alpha_1} - \tfrac{\beta_1 \alpha_1}{(1-\alpha_1)^2})\right]} + \tfrac{\beta_0}{(1-\alpha_1)} - \tfrac{\beta_1 \alpha_1}{(1-\alpha_1)^2} + (\tfrac{\beta_1}{(1-\alpha_1)})t = \\ &= \alpha_1^t [E(Y_0) - \mu_0] + \mu_0 + \mu_1 t, \end{aligned} \tag{58}$$

where the model parameters relate to the moments of $\{Y_t,\ t\in\mathbb{N}\}$ via:

$$\mu_0=[\{\beta_0/(1-\alpha_1)\}-\alpha_1\beta_1/(1-\alpha_1)^2],\quad \mu_1=[\beta_1/(1-\alpha_1)]. \tag{59}$$

By viewing Y_0 as a typical element of the process $\{Y_t,\ t\in\mathbb{N}\}$ with $E(Y_t)=\mu_0+\mu_1 t$, $t\in\mathbb{N}_{\{0\}}$, implies $E(Y_0)=\mu_0$, which addresses the t-varying problem. As expected from the statistical perspective, $E(Y_t)=\mu_0+\mu_1 t$ coincides with that of $f(y_t, y_{t-1};\boldsymbol{\phi})$ in Table 4. When (59) is solved for β_0 and β_1, yields the parametrizations in Table 5.

Similarly, rearranging the terms of $Cov(Y_t, Y_{t+\tau})$ in (57) to isolate α_1^{2t} yields:

$$\begin{aligned}Cov(Y_t, Y_{t+\tau}) &=\alpha_1^{2t+|\tau|}Var(Y_0)+\sigma_\varepsilon^2\alpha_1^{|\tau|}(\tfrac{1-\alpha_1^{2t}}{1-\alpha_1^2})=\\ &=\alpha_1^{|\tau|}\left\{\alpha_1^{2t}\left[Var(Y_0)-(\tfrac{\sigma_\varepsilon^2}{1-\alpha_1^2})\right]+(\tfrac{\sigma_\varepsilon^2}{1-\alpha_1^2})\right\},\ |\tau|\geq 0,\end{aligned} \tag{60}$$

which indicates that for $Var(Y_0)=(\frac{\sigma_\varepsilon^2}{1-\alpha_1^2}):=\sigma(0),\ Cov(Y_t, Y_{t+\tau})=\alpha_1^{|\tau|}(\frac{\sigma_\varepsilon^2}{1-\alpha_1^2})$, yielding:

$$Var(Y_t)=(\sigma_\varepsilon^2/)1-\alpha_1^2))=\sigma(0) \text{ and } Cov(Y_t, Y_{t-1})=\alpha_1\sigma(0)=\sigma(1),\ t\in\mathbb{N}. \tag{61}$$

Similarly, treating Y_0 as a typical element of $\{Y_t,\ t\in\mathbb{N}_{\{0\}}\}$, i.e. for $E(Y_0)=\mu_0$, $Var(Y_0)=\sigma(0)$, the t-varying problem disappears, and solving $\alpha_1\sigma(0):=\sigma(1)$ and $[\sigma_\varepsilon^2/(1-\alpha_1^2))=\sigma(0)$, for α_1 and σ_ε^2, yields the parametrizations of the key parameters $\alpha_1=[\sigma(1)/\sigma(0)]$, $\sigma_\varepsilon^2=\sigma(0)(1-\alpha_1^2)$ (Table 5).

6.2 The UR(1) model: $|\alpha_1|=1$

In the case $|\alpha_1|=1$ the summations in (56) yield:

$$\textstyle\sum_{i=0}^{t-1}\alpha_1^i=t,\quad \sum_{i=0}^{t-1}\alpha_1^i=(t-1),\quad \sum_{i=0}^{t-1}\alpha_1^i(t-i)=\frac{1}{2}t\,(t+1)\,,\quad \sum_{i=0}^{t-1}\alpha_1^{2i}=t,$$

and thus (54), after replacing β_0, β_1 with γ_0, γ_1 to avoid confusions, yields:

$$E(Y_t)\ =E(Y_0)+\gamma_0 t+\gamma_1\left(\tfrac{1}{2}t\,(t+1)\right)=\mu_0+\mu_1 t+\mu_2 t^2,\ t\in\mathbb{N}, \tag{62}$$

$$\text{where } \mu_0=E(Y_0),\ \mu_1=\left(\tfrac{1}{2}\gamma_1+\gamma_0\right),\ \mu_2=\tfrac{1}{2}\gamma_1. \tag{63}$$

Using (62), the covariance of $(Y_t-E(Y_t))=\sum_{i=0}^{t-1}\varepsilon_{t-i}$ is:

$$Cov(Y_t, Y_{t-1})=E\left[(\textstyle\sum_{i=0}^{t-1}\varepsilon_{t-i})(\sum_{i=1}^{t-1}\varepsilon_{t-i})\right]=\sum_{i=1}^{t-1}E(\varepsilon_{t-i}^2)=\sigma_\varepsilon^2(t-1).$$

For $\sigma(0)=\sigma_\varepsilon^2$, this implies that: $Var(Y_t)=\sigma(0)t,\ Cov(Y_t, Y_{t-1})=\sigma(0)(t-1)$, where the choice of $E(Y_0)=\mu_0$ and $Var(Y_0)=0$, stem again from viewing Y_0 as typical element of the sample $\mathbf{Y}_0$. The first two moments indicate that $\{Y_t,\ t\in\mathbb{N}\}$ is a Wiener process (Table 6), yielding the statistical UR(1) model in Table 7.

In conclusion, it is important to emphasize that the above derivations give rise to the statistical parametrizations associated with the AR(1)-(B) model when $\mu_1=\mu_2=0$, for the statistical AR(1) model in Table 2 and $\mu_2=0$ for the UR(1) model in Table 3.

7 Appendix B: Derivations - the LR and LM tests

AR(1)-TV model 3 in (45): $\mathcal{M}_1$: $Y_t=\beta_0+\delta t+\alpha_1 Y_{t-1}+u_t,\ u_t\backsim \mathsf{NMd}(0,\sigma^2(t)),\ t\in\mathbb{N}.$

7.1 Likelihood Ratio (LR) test

The unrestricted *likelihood function* (LF) takes the form:

$$L(\widehat{\boldsymbol{\theta}})\ =e^{-\frac{n}{2}}\left[(2\pi)^{-\frac{n}{2}}\right]\left[\textstyle\prod_{t=1}^{n}\widehat{\sigma}^2(\widehat{\mu}_1,\widehat{\alpha}_1)(\widehat{\alpha}_1^2(1-t)+t)\right]^{-\frac{1}{2}},$$

where the unrestricted MLEs are denoted by '$\widehat{\theta}$'.

The restricted likelihood is: $L(\widetilde{\boldsymbol{\theta}})=e^{-\frac{n}{2}}\left((2\pi)^{-\frac{n}{2}}\right)(\widetilde{\sigma}^2)^{-\frac{n}{2}},\ \widetilde{\boldsymbol{\theta}}:=(\widetilde{\mu}_1,1,\widetilde{\sigma}^2),$ where the restricted MLEs are: $\widetilde{\sigma}^2=\frac{1}{n}\sum_{t=1}^n(\Delta Y_t-\widetilde{\mu}_1)^2,\quad \widetilde{\mu}_1=\frac{1}{n}\sum_{t=1}^n\Delta Y_t.$

The LR test statistic takes the general form:

$$\lambda(\mathbf{y})=\frac{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}(L(\boldsymbol{\theta}))}{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}(L(\boldsymbol{\theta}))}=\frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})}\quad=\frac{e^{-\frac{n}{2}}\left[(2\pi)^{-\frac{n}{2}}\right]\left[\widetilde{\sigma}^2\right]^{-\frac{n}{2}}}{e^{-\frac{n}{2}}\left[(2\pi)^{-\frac{n}{2}}\right]\left[\prod_{t=1}^n\left[\widehat{\sigma}^2(\widehat{\mu}_1,\widehat{\alpha}_1)\right]\left[\widehat{\alpha}_1^2(1-t)+t\right]\right]^{-\frac{1}{2}}}=$$

$$=\left[\widetilde{\sigma}^2\right]^{-\frac{n}{2}}\left[\prod_{t=1}^n\left[\widehat{\sigma}^2(\widehat{\mu}_1,\widehat{\alpha}_1)\right](\widehat{\alpha}_1^2(1-t)+t)\right]^{\frac{1}{2}},$$

$$-2\ln\lambda(\mathbf{y})\quad=n(\ln\widetilde{\sigma}^2-\ln\widehat{\sigma}^2(\widehat{\mu}_1,\widehat{\alpha}_1))-\sum_{t=1}^n\ln(\widehat{\alpha}_1^2(1-t)+t),$$

where the last component takes the explicit form:

$$\sum_{t=1}^n\ln(\alpha_1^2(1-t)+t)=n\ln\left((1-\alpha_1)(1+\alpha_1)\right)+\ln\left[\Gamma\left(\frac{n(\alpha_1^2-1)-1}{(\alpha_1-1)(\alpha_1+1)}\right)/\Gamma\left(-\frac{1}{(\alpha_1-1)(\alpha_1+1)}\right)\right].$$

7.2 Lagrange Multiplier (LM) test

Derivation. The log-likelihood function of the AR(1)-TV model in Table 19 is:

$$\ln L(\boldsymbol{\theta})=\text{const.}-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2}\sum_{t=1}^n\ln c_t-\frac{1}{2\sigma^2}\sum_{t=1}^n\frac{u_t^2}{c_t},\ \boldsymbol{\theta}:=(\mu_1,\alpha_1,\sigma^2),$$
$$c_t:=\left[\alpha_1^2(1-t)+t\right],\ u_t:=(Y_t-\alpha_1\mu_1-(1-\alpha_1)\mu_1 t-\alpha_1 Y_{t-1}).$$

The first derivatives leading to the unrestricted MLEs are as follows:

$$\frac{\partial\ln L}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{t=1}^n\frac{u_t^2}{c_t}\ \Rightarrow\widehat{\sigma}^2(\widehat{\mu}_1,\widehat{\alpha}_1)=\frac{1}{n}\sum_{t=1}^n\frac{(Y_t-\widehat{\mu}_1[\widehat{\alpha}_1(1-t)+t]-\widehat{\alpha}_1Y_{t-1})^2}{\left[\widehat{\alpha}_1^2(1-t)+t\right]},$$

$$\frac{\partial\ln L}{\partial\alpha_1}=\frac{1}{\sigma^2}\sum_{t=1}^n\left(\frac{\alpha_1(1-t)}{c_t^2}\right)\left[u_t^2-c_t\sigma^2\right]+\frac{1}{\sigma^2}\sum_{t=1}^n\frac{[\mu_1(1-t)+Y_{t-1}]u_t}{c_t},$$

$$\frac{\partial\ln L}{\partial\mu_1}=\frac{1}{\sigma^2}\sum_{t=1}^n\frac{[\alpha_1(1-t)+t]u_t}{c_t}\Rightarrow\widehat{\mu}_1(\widehat{\alpha}_1)=\left(\sum_{t=1}^n\frac{[\widehat{\alpha}_1(1-t)+t]^2}{\left[\widehat{\alpha}_1^2(1-t)+t\right]}\right)^{-1}\left(\sum_{t=1}^n\frac{[\widehat{\alpha}_1(1-t)+t](Y_t-\widehat{\alpha}_1Y_{t-1})}{\left[\widehat{\alpha}_1^2(1-t)+t\right]}\right).$$

NOTE that the first order conditions for (μ_1,σ^2) can be solved explicitly, but that of α_1 can only be solved numerically. The first derivatives under H_0: $\alpha_1=1$:

$$\left.\frac{\partial\ln L}{\partial\sigma^2}\right|_{\alpha_1=1}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{t=1}^n u_t^2,$$

$$\left.\frac{\partial\ln L}{\partial\alpha_1}\right|_{\alpha_1=1}=\frac{1}{\sigma^2}\sum_{t=1}^n\left[u_t^2-\sigma^2\right]\xi_t+\frac{1}{\sigma^2}\sum_{t=1}^n\psi_t u_t,$$

$$\left.\frac{\partial\ln L}{\partial\mu_1}\right|_{\alpha_1=1}=\frac{1}{\sigma^2}\sum_{t=1}^n u_t,\ \text{where }\psi_t:=\left[\mu_1\xi_t+Y_{t-1}\right],\ \xi_t:=(1-t),$$

yielding the score: $\boldsymbol{\eta}(\boldsymbol{\theta}):=\left(\left(\left.\frac{\partial\ln L(\boldsymbol{\theta})}{\partial\sigma^2}\right|_{\alpha_1=1}\right),\ \left(\left.\frac{\partial\ln L(\boldsymbol{\theta})}{\partial\alpha_1}\right|_{\alpha_1=1}\right),\ \left(\left.\frac{\partial\ln L(\boldsymbol{\theta})}{\partial\mu_1}\right|_{\alpha_1=1}\right)\right).$

The estimated first derivatives under H_0: $\alpha_1=1$, are:

$$\left.\frac{\partial\ln L(\widehat{\boldsymbol{\theta}})}{\partial\sigma^2}\right|_{\alpha_1=1}=-\frac{n}{2\widetilde{\sigma}^2}+\frac{1}{2(\widetilde{\sigma}^2)^2}\sum_{t=1}^n\widetilde{u}_t^2,\ \text{where }\widetilde{u}_t=(\Delta Y_t-\widetilde{\mu}_1),\ \widetilde{\mu}_1=\frac{1}{n}\sum_{t=1}^n\Delta Y_t,$$

$$\left.\frac{\partial\ln L(\widehat{\boldsymbol{\theta}})}{\partial\alpha_1}\right|_{\alpha_1=1}=\frac{1}{\widetilde{\sigma}^2}\sum_{t=1}^n\xi_t\left[\widetilde{u}_t^2-\widetilde{\sigma}^2\right]+\frac{1}{\widetilde{\sigma}^2}\sum_{t=1}^n\widetilde{\psi}_t\widetilde{u}_t,\ \text{where }\widetilde{\sigma}^2=\frac{1}{n}\sum_{t=1}^n(\Delta Y_t-\widetilde{\mu}_1)^2,$$

$$\left.\frac{\partial\ln L(\widehat{\boldsymbol{\theta}})}{\partial\mu_1}\right|_{\alpha_1=1}=\frac{1}{\widetilde{\sigma}^2}\sum_{t=1}^n\widetilde{u}_t,\ \text{where }\widetilde{\psi}_t:=\left[\widetilde{\mu}_1\xi_t+Y_{t-1}\right],\ \xi_t:=(1-t),$$

and thus the estimated score vector $\boldsymbol{\eta}(\widetilde{\boldsymbol{\theta}}){:=}\boldsymbol{\eta}(\widehat{\boldsymbol{\theta}})\Big|_{\alpha_1=1}$ takes the form:

$$\boldsymbol{\eta}(\widetilde{\boldsymbol{\theta}}){:=}\Big(\;0,\;\; (\tfrac{1}{\widetilde{\sigma}^2}\textstyle\sum_{t=1}^{n}\widetilde{\psi}_t\widetilde{u}_t{+}\tfrac{1}{\widetilde{\sigma}^2}\sum_{t=1}^{n}\xi_t\left[\widetilde{u}_t^2{-}\widetilde{\sigma}^2\right]),\;\; 0\;\Big).$$

The second derivatives of the log-likelihood function are:

$$\frac{\partial^2 \ln L}{\partial\sigma^4} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{t=1}^{n}\frac{u_t^2}{c_t},$$

$$\frac{\partial^2 \ln L}{\partial\alpha_1^2} = -\sum_{t=1}^{n}\left[\frac{(1-t)}{c_t}\right] + 2\sum_{t=1}^{n}\frac{[\alpha_1(1-t)]^2}{c_t^2} - \frac{1}{\sigma^2}\sum_{t=1}^{n}\frac{[\mu_1(1-t)+Y_{t-1}]^2}{c_t} -$$
$$-\frac{4\alpha_1}{\sigma^2}\sum_{t=1}^{n}\frac{(1-t)[\mu_1(1-t)+Y_{t-1}]u_t}{c_t^2} + \frac{1}{\sigma^2}\sum_{t=1}^{n}(1-t)\frac{u_t^2}{c_t^2} - \frac{4\alpha_1^2}{\sigma^2}\sum_{t=1}^{n}\frac{(1-t)^2u_t^2}{c_t^3},$$

$$\frac{\partial^2 \ln L}{\partial\mu_1^2} = -\frac{1}{\sigma^2}\sum_{t=1}^{n}\frac{[\alpha_1(1-t)+t]^2}{c_t},$$

$$\frac{\partial^2 \ln L}{\partial\sigma^2\partial\alpha_1} = -\frac{1}{\sigma^4}\left[\sum_{t=1}^{n}\frac{[\mu_1(1-t)+Y_{t-1}]u_t}{c_t} + \sum_{t=1}^{n}[\alpha_1(1-t)]\frac{u_t^2}{c_t^2}\right],$$

$$\frac{\partial^2 \ln L}{\partial\sigma^2\partial\mu_1} = -\frac{1}{\sigma^4}\sum_{t=1}^{n}[\alpha_1(1-t)+t]\frac{u_t}{c_t},$$

$$\frac{\partial^2 \ln L}{\partial\mu_1\partial\alpha_1} = \frac{-2}{\sigma^2}\sum_{t=1}^{n}\frac{[\alpha_1(1-t)+t][\alpha_1(1-t)]u_t}{c_t^2} + \frac{1}{\sigma^2}\sum_{t=1}^{n}\frac{[(1-t)]u_t}{c_t} - \frac{1}{\sigma^2}\sum_{t=1}^{n}\frac{[\mu_1(1-t)+Y_{t-1}][\alpha_1(1-t)+t]}{c_t}.$$

The Fisher information matrix under $H_0{:}\alpha_1{=}1$ is:

$$\mathbb{I}_n(\boldsymbol{\theta})|_{\alpha_1=1} := E\left(-\begin{pmatrix} \frac{\partial^2 \ln L}{\partial\mu_1^2} & \frac{\partial^2 \ln L}{\partial\mu_1\partial\alpha_1} & \frac{\partial^2 \ln L}{\partial\mu_1\partial\sigma^2} \\ \frac{\partial^2 \ln L}{\partial\mu_1\partial\alpha_1} & \frac{\partial^2 \ln L}{\partial\alpha_1^2} & \frac{\partial^2 \ln L}{\partial\sigma^2\partial\alpha_1} \\ \frac{\partial^2 \ln L}{\partial\mu_1\partial\sigma^2} & \frac{\partial^2 \ln L}{\partial\sigma^2\partial\alpha_1} & \frac{\partial^2 \ln L}{\partial\sigma^4}\end{pmatrix}\right)\Bigg|_{\alpha_1=1} = \begin{pmatrix} \frac{n}{\sigma^2} & \sum\limits_{t=1}^{n}\frac{\psi_t}{\sigma^2} & 0 \\ \sum\limits_{t=1}^{n}\frac{\psi_t}{\sigma^2} & 2\sum\limits_{t=1}^{n}\xi_t^2+\sum\limits_{t=1}^{n}\frac{\psi_t^2}{\sigma^2} & \frac{1}{\sigma^2}\sum\limits_{t=1}^{n}\xi_t \\ 0 & \frac{1}{\sigma^2}\sum\limits_{t=1}^{n}\xi_t & \frac{n}{2\sigma^4}\end{pmatrix}.$$

Hence, for $\mathbb{I}_n(\widetilde{\boldsymbol{\theta}}){:=}\mathbb{I}_n(\widehat{\boldsymbol{\theta}})\Big|_{\alpha_1=1}$, the LM test statistic is:

$$LM(\mathbf{y}){=}\boldsymbol{\eta}(\widetilde{\boldsymbol{\theta}})^\top\left(\mathbb{I}_n(\widetilde{\boldsymbol{\theta}})\right)^{-1}\boldsymbol{\eta}(\widetilde{\boldsymbol{\theta}}) \;=\frac{\left(\frac{1}{\widetilde{\sigma}^2}\sum\limits_{t=1}^{n}\widetilde{\psi}_t\widetilde{u}_t+\frac{1}{\widetilde{\sigma}^2}\sum\limits_{t=1}^{n}\xi_t\left[\widetilde{u}_t^2-\widetilde{\sigma}^2\right]\right)^2}{\left[2\sum_{t=1}^{n}(\xi_t-\overline{\xi})^2+\left(\frac{1}{\widetilde{\sigma}^2}\right)\sum_{t=1}^{n}(\widetilde{\psi}_t-\overline{\psi})^2\right]},\;\; \xi_t{:=}(1{-}t).$$

It is interesting to note that the determinant of the restricted information matrix:

$$\det(\mathbb{I}_n(\boldsymbol{\theta})|_{\alpha_1=1}){=}\left(\frac{n^2}{2\sigma^6}\right)\left[2\sum_{t=1}^{n}(1-t)^2+\sum_{t=1}^{n}\frac{\psi_t^2}{\sigma^2}\right] - \left(\frac{n}{\sigma^2}\right)\left(\frac{1}{\sigma^2}\sum_{t=1}^{n}(1-t)\right)^2 - \frac{n}{2\sigma^4}\left(\sum_{t=1}^{n}\frac{\psi_t}{\sigma^2}\right)^2$$
$$=\left(\frac{n^2}{2\sigma^6}\right)\left[2\sum_{t=1}^{n}(\xi_t-\overline{\xi})^2+\left(\frac{1}{\sigma^2}\right)\sum_{t=1}^{n}(\psi_t-\overline{\psi})^2\right]^2,$$

which brings out a direct connection between the $\det(\mathbb{I}_n(\boldsymbol{\theta})|_{\alpha_1=1})$ and $LM(\mathbf{y})$.

8 Appendix C: Simulated Data plots

8.1 Simulated data t-plots for the AR(1)-TV model

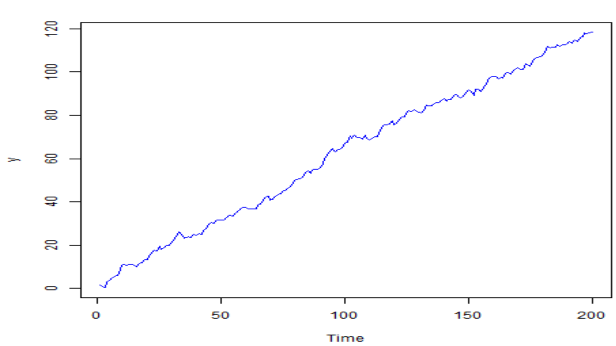

Fig. A1: $y_t{=}.5{+}1.0y_{t-1}{+}\epsilon_t,\;\epsilon_t{\backsim}\mathsf{N}(0{,}1)$

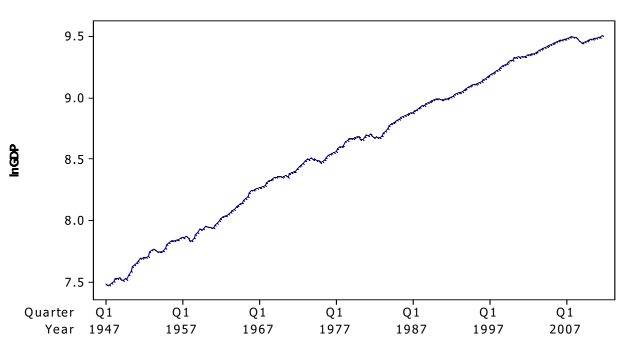

Fig. A2: lnGDP (quarterly data 1947-2007)

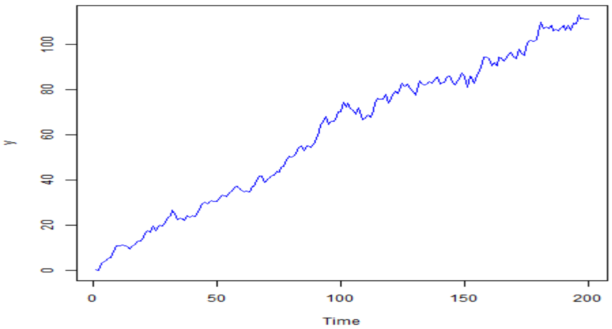

Fig. A3: y_t=.515+.005t+.99y_{t-1}+$\sigma(t)\epsilon_t$

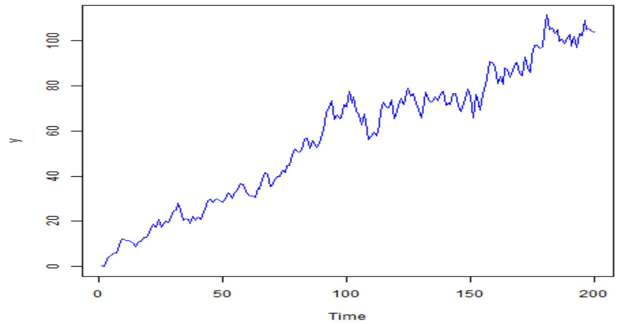

Fig. A4: y_t=.575+.025t+.95y_{t-1}+$\sigma(t)\epsilon_t$

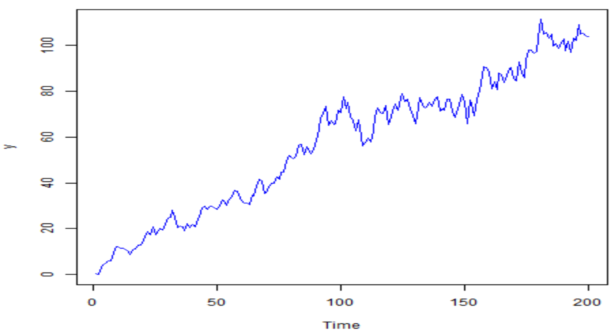

Fig. A5: y_t=.65+.05t+.9y_{t-1}+$\sigma(t)\epsilon_t$

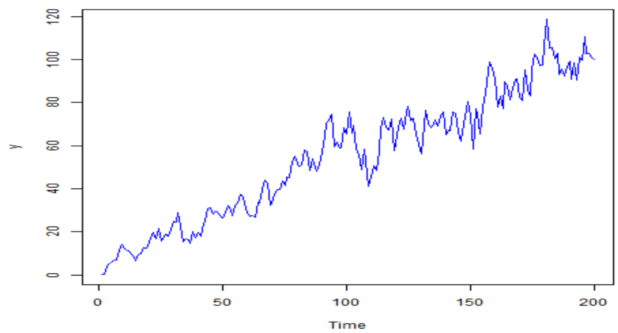

Fig. A6: y_t=.725+.075t+.85y_{t-1}+$\sigma(t)\epsilon_t$

8.2 Simulated data plots for the LR and LM tests

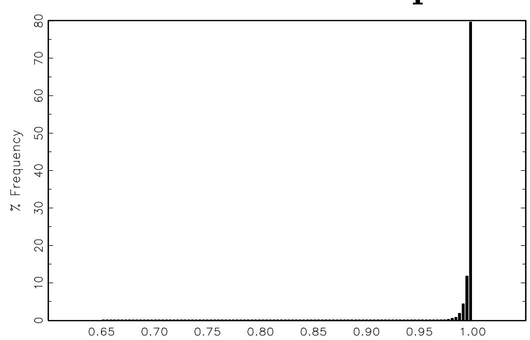

Fig. A9: Histogram: MLE estimator of α_1

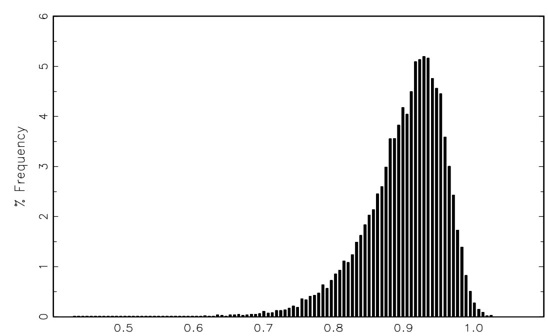

Fig. A10: Histogram: OLS estimator of α_1

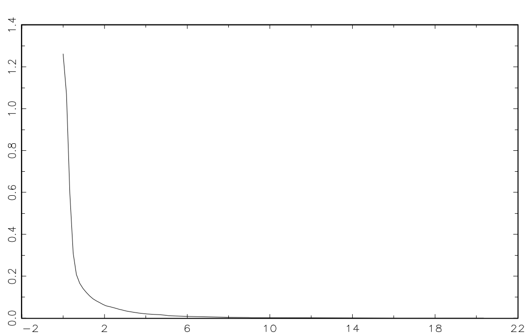

Fig. A11: Smoothed histogram of LR test

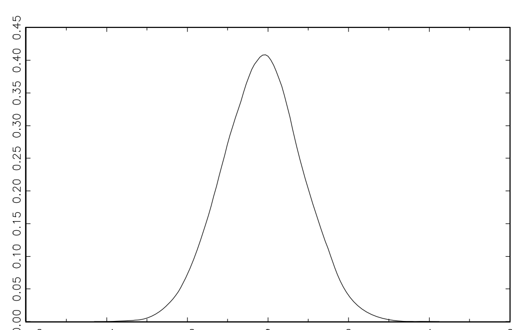

Fig. A12: Smoothed histogram of $\sqrt{\text{LM}}$ test