Uniform and distribution-free inference with general autoregressive processes*

Tassos Magdalinos
University of Southampton

Katerina Petrova
Universitat Pompeu Fabra and Barcelona School of Economics

——— December 19, 2022 ———

Abstract

A unified theory of estimation and inference is developed for an autoregressive process with root in $(-\infty, \infty)$ that includes the stable, unstable, explosive and all intermediate regions. The discontinuity of the limit distribution of the t-statistic along autoregressive regions and its dependence on the distribution of the innovations in the explosive regions $(-\infty, -1) \cup (1, \infty)$ are addressed simultaneously. A novel estimation procedure, based on a data-driven combination of a near-stationary and a mildly explosive endogenously constructed instrument, delivers an asymptotic mixed-Gaussian theory of estimation and gives rise to an asymptotically standard normal t-statistic across all autoregressive regions independently of the distribution of the innovations. The resulting hypothesis tests and confidence intervals are shown to have correct asymptotic size (uniformly over the parameter space) both in autoregressive and in predictive regression models, thereby establishing a general and unified framework for inference with autoregressive processes. Extensive Monte Carlo experimentation shows that the proposed methodology exhibits very good finite sample properties over the entire autoregressive parameter space $(-\infty, \infty)$ and compares favourably to existing methods within their parametric $(-1, 1]$ validity range. We demonstrate that a first-order difference equation for the number of infections with an explosive/stable root results naturally after linearisation of an SIR model at the outbreak and apply our procedure to Covid-19 infections to construct confidence intervals on the model's parameters, including the epidemic's basic reproduction number, across a panel of countries without *a priori* knowledge of the model's stability/explosivity properties.

Keywords: Uniform inference, Central limit theory, Autoregression, Predictive regression, Instrumentation, Mixed-Gaussianity, t-statistic, Confidence intervals.

*We would like to thank Peter Phillips, Kirill Evdokimov, Adam McCloskey and Bo Zhou for valuable comments and discussions. Petrova acknowledges support by the General Directorate for Research in the Government of Catalonia through the Beatriu de Pinós grant 2019/BP/00239 and the Spanish Ministry of Science and Innovation, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

1 Introduction

Inference in the first-order autoregressive process, arguably the prototypical time series model, has a long history dating back to at least Mann and Wald (1943) for stationary autoregression, White (1958) for explosive autoregression and Phillips (1987a) for unit-root autoregression. The variety of stochastic behaviour arising from different autoregressive regimes has resulted in a number of important applications in macroeconomics and finance: nonstationary autoregressive processes played a fundamental role in the development of the theory of cointegration and causal inference in systems of macroeconomic and financial variables. Autoregressive processes with coefficients in the explosive region $(1, \infty)$ have been employed for the modelling of phenomena whose temporal evolution exhibits stochastic exponential growth, from the rate of epidemic infection to the formation of financial and commodity bubbles during periods of market exuberance.

While convenient from a modelling point of view, the different stochastic properties arising from different regions of the autoregressive parameter space present a major challenge for inference, with standard econometric methodology (such as least squares or maximum likelihood) applying only under *a priori* knowledge of the parameter region, with misspecification resulting to asymptotically invalid confidence intervals (CIs) and hypothesis tests. Early work on obtaining CIs for an autoregressive coefficient in $(-1, 1]$, thereby accommodating stationary autoregressions and unit root processes, includes Stock (1991), Andrews (1993), Hansen (1999) and Romano and Wolf (2001). Mikusheva (2007) develops the first general methodology for establishing uniform properties of CIs in autoregressive processes with root in $(-1, 1]$ and proposes a correction of Stock (1991)'s method that achieves uniform asymptotic validity. Subsequent work by Andrews and Guggenberger (2009, 2014) establishes methodology for CI construction with correct asymptotic size uniformly over the above region under the potential presence of conditional heteroskedasticity of unknown form. Uncertainty over the persistence degree of a stochastic regressor poses similar difficulties for hypothesis testing in a regression model and a literature on inference in a predictive regression with a near-nonstationary regressor was developed in parallel with the aforementioned advances in autoregressive inference. Notable contributions include Campbell and Yogo (1996), Jansson and Moreira (2006) as well as bootstrap methods based on the theoretical results of Cavaliere and Georgiev (2020). Hypothesis testing procedures that achieve robust inference with time series regressors with persistence ranging from stationarity to (near) unit root nonstationarity are those of Elliott, Müller and Watson (2015) and Kostakis, Magdalinos and Stamatogiannis (2015). The latter paper builds on the IVX procedure of Phillips and Magdalinos (2009), which has been extended in a number of directions by Breitung and Demeterscu (2015), Yang, Long, Peng and Cai (2020), Magdalinos and Phillips (2020), Demeterscu, Georgiev, Rodrigues and Taylor (2022).

Both strands of the literature on inference in autoregressions and predictive regressions discussed above restrict the autoregressive parameter space to $(-1, 1]$; the aim of this paper is to develop hypothesis tests and CIs with uniform asymptotic validity over the entire autoregressive parameter space $(-\infty, \infty)$ and over the space of a wide class of innovation distribution functions.

We propose a novel data-generated instrumental variable (IV) procedure that tackles two important inference problems in autoregressions and predictive regressions simultaneously: firstly, it delivers a unified asymptotic theory of inference and CI construction that covers the entire autoregressive spectrum of stationary, nonstationary, explosive processes and all intermediate regions; secondly, it provides a solution to the long-standing problem of distribution-free asymptotic inference in explosive autoregressions[1].

The key idea of our approach is to filter the regressor's autoregressive data generating process (DGP) through a time series that acts as an endogenously generated instrument constructed to behave asymptotically as: (i) a near-stationary process[2] when the DGP lies close to the stationary region; (ii) a mildly explosive process when the DGP lies close to the explosive region; (iii) a random linear combination of (i) and (ii) when the DGP is in the near-nonstationary region defined by at most local departures from unity. The resulting IV estimator inherits the desirable asymptotic properties of near-stationary and mildly explosive processes and is asymptotically mixed-Gaussian along the entire autoregressive parameter space $(-\infty, \infty)$ independently of the distribution of the innovations of the autoregressive process. The asymptotic mixed-Gaussianity property implies that the IV-based t-statistic is asymptotically standard normal and can be employed for CI construction based on $\mathcal{N}(0,1)$ quantiles. Moreover, we show that the proposed IV-based test has uniformly correct size and gives rise to CIs with uniformly correct asymptotic coverage. To our knowledge, our procedure provides the first unified, distribution-free treatment of first-order autoregression exhibiting arbitrary stochastic characteristics ranging from stationarity to explosivity.

Extensive Monte Carlo experimentation reveals good finite sample properties for the proposed IV-based hypothesis tests and CIs that compare favourably to the leading procedures for inference in autoregression (Andrews and Guggenberger (2014)) and predictive regression (Elliott et al. (2015)) in their parametric validity range $(-1, 1]$ while providing correct inference in $(-\infty, -1] \cup (1, \infty)$, where no existing alternative approach has general asymptotic validity.

Oscillating processes with roots in $(-\infty, 0)$ arise naturally in series which exhibit seasonality at certain frequency, and seasonal unit root tests are routinely used to test hypotheses on whether shocks have permanent effect on the seasonal pattern of the series. Autoregressive processes with roots potentially exceeding unity for a non-trivial fraction of the sample are popular for modelling and date stamping of financial and commodity price bubbles (Phillips and Yu (2011), Phillips, Wu and Yu (2011) among others). Further empirically relevant applications include series that exhibit stochastic exponential growth, for example, epidemiological models of disease transmission. In this paper, we consider a susceptible-infected-removed (SIR) model of temporal evolution of disease transmission and show that, upon linearisation around the disease-free equilibrium, the

[1] Anderson (1959) shows that, in the explosive case, the limit distributions of the OLS estimator and the associated t-statistic are not invariant to deviations from the assumptions of i.i.d. Gaussian errors and zero initial condition; in general, they are of unknown form driven by the distribution of the innovations in the autoregression.

[2] Near-stationary and mildly explosive processes, introduced by Phillips and Magdalinos (2007), are AR(1) processes with sample-size dependent root θ_n satisfying $\theta_n \to 1$ and: $n(\theta_n - 1) \to -\infty$ in the near-stationary case or $n(\theta_n - 1) \to \infty$ in the mildly explosive case.

model-implied number of active infections evolves as a first order autoregressive process with an explosive (stable) root whenever the basic reproduction number r_0 is above (below) unity. In Section 6, we employ our procedure to model the early dynamics of the Covid-19 epidemic across a panel of countries and construct CIs for r_0 and the other epidemiological parameters of the model without *a priori* knowledge of whether the epidemic is in a controllable or uncontrollable stage, i.e. without restricting the parameter space.

The paper is organised as follows: Section 2 presents a general framework of autoregression (Section 2.1), predictive regression (Section 2.2) and sets out the dynamic behaviour of a basic SIR epidemiological model (Section 2.3). Section 3.1 introduces our novel IV procedure of combined near-stationary/mildly explosive instrumentation. Section 3.2 presents the main results on uniform asymptotic inference in autoregression and predictive regression (Theorems 1 and 2) and applies them to the SIR model of Section 2.3 (Corollary 1). Section 3.3 establishes the asymptotic mixed-Gaussianity property of the IV estimators that drives the asymptotic results of Section 3.2. Section 4 discusses implementation of the procedure and conducts Monte Carlo experiments to assess the finite sample properties of our CIs and hypothesis tests in comparison to the leading existing inference procedures in autoregressions and predictive regressions. Section 5 applies the CIs of Corollary 1 to Covid-19 infections across a panel of countries and Section 6 concludes. The proofs of the main results (Theorems 1-3) are provided in Appendix A. The proofs of all remaining statements of the paper (Lemmata 1-6 and Corollary 1) are provided in the supplementary online Appendix B, which also contains auxiliary mathematical results and additional simulation results.

2 A model of general autoregressive dependence

2.1 Probabilistic framework for autoregression

We consider a first order autoregressive process with an intercept

$$x_t = \mu + X_t, \quad X_t = \rho_n X_{t-1} + u_t, \quad t \in \{1, ..., n\} \tag{1}$$

with (possibly sample-size-dependent) autoregressive root ρ_n in $(-\infty, \infty)$, with an innovation sequence $(u_t)_{t\in\mathbb{N}}$ and an initialisation X_0.

Assumptions maintained on ρ_n, $(u_t)_{t\in\mathbb{N}}$ and X_0 are presented in Assumptions 1, 2 and 3 below.

Assumption 1a (AR parameter space). *The parameter space of the autoregressive parameter in (1) has the following form:* $(\rho_n)_{n\in\mathbb{N}}$ *is a sequence of real numbers satisfying* $\rho_n \to \rho \in (-\infty, \infty)$.

Assumption 1a considers drifting sequences of autoregressive parameters of sufficient generality to ensure the uniform asymptotic validity of critical regions and CIs over an autoregressive parameter space consisting of an arbitrary closed subinterval of $(-\infty, \infty)$: see (25) and Theorems 1 and 2 below. In order to establish the asymptotic theory of estimation of Theorem 3, it is convenient to strengthen Assumption 1a in a way that categorises autoregressive processes according to their stochastic properties.

Assumption 1b (AR categories). *In addition to* $(\rho_n)_{n\in\mathbb{N}}$ *satisfying Assumption 1a, the limit* $c := \lim_{n\to\infty} n\left(|\rho_n| - 1\right)$ *exists in* $\mathbb{R} \cup \{-\infty, \infty\}$.

Under Assumption 1b, the process x_t in (1) belongs to one of the following classes:

C(i) *near-stationary* processes if $(\rho_n)_{n\in\mathbb{N}}$ in (1) satisfies Assumption 1b with $c = -\infty$

C(ii) *near-nonstationary* processes if $(\rho_n)_{n\in\mathbb{N}}$ in (1) satisfies Assumption 1b with $c \in \mathbb{R}$

C(iii) *near-explosive* processes if $(\rho_n)_{n\in\mathbb{N}}$ in (1) satisfies Assumption 1b with $c = \infty$.

Each of the above autoregressive classes may be partitioned into a *regular* subclass when $\rho \geq 0$ and an *oscillating* subclass when $\rho < 0$: $C(i) = C_+(i) \cup C_-(i)$, $C(ii) = C_+(ii) \cup C_-(ii)$ and $C(iii) = C_+(iii) \cup C_-(iii)$, with $\mathrm{C}_+(.)$ denoting the relevant subclass of processes when $\rho \geq 0$ and $\mathrm{C}_-(.)$ denoting the relevant subclass of processes when $\rho < 0$. We further denote the subclass of C(i) consisting of purely *stationary* processes and the subclass of C(iii) consisting of purely *explosive* processes by:

C$_0$(i) $(\rho_n)_{n\in\mathbb{N}}$ in (1) satisfies $\rho_n \to \rho \in (-1, 1)$

C$_0$(iii) $(\rho_n)_{n\in\mathbb{N}}$ in (1) satisfies $\rho_n \to \rho$ with $|\rho| > 1$.

When $|\rho| = 1$ in Assumption 1a, Assumption 1b is more restrictive than Assumption 1a (Assumptions 1a and 1b are equivalent when $|\rho| \neq 1$): for example, the sequence

$$\rho_n = 1 + (-1)^n / k_n \quad k_n \to \infty \tag{2}$$

satisfies Assumption 1a but not Assumption 1b. However, sequences of autoregressive parameters satisfying Assumption 1a satisfy Assumption 1b subsequentially, in the following sense.

Lemma 1. *Let $(\rho_n)_{n\in\mathbb{N}}$ satisfy Assumption 1a. For any subsequence $(\rho_{m_n})_{n\in\mathbb{N}}$ of $(\rho_n)_{n\in\mathbb{N}}$ there exists a further subsequence $(\rho_{s_n})_{n\in\mathbb{N}}$ of $(\rho_{m_n})_{n\in\mathbb{N}}$ such that $(\rho_{s_n})_{n\in\mathbb{N}}$ satisfies Assumption 1b.*

We will see in Sections 3.2 and 3.3 below that, while Assumption 1b is needed to establish the asymptotic mixed-Gaussianity of the proposed IV estimator of Theorem 3, studentisation and Lemma 1 may be employed to weaken the requirement on $(\rho_n)_{n\in\mathbb{N}}$ to Assumption 1a for the (uniform) asymptotic validity of the test statistics and CIs of Theorems 1 and 2.

Assumption 2 (innovation sequence). *Given a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, u_t in (1) is an $\mathcal{F}_t$-martingale difference sequence such that $\mathbb{E}_{\mathcal{F}_{t-1}}(u_t^2) = \sigma^2$ for all but finitely many t a.s.,*

$$\liminf_{t\to\infty} \mathbb{E}_{\mathcal{F}_{t-1}} |u_t| > 0 \ \ a.s. \tag{3}$$

and $(u_t^2)_{t\in\mathbb{Z}}$ is a uniformly integrable sequence.

Assumption 3 (initial condition). *The initial condition $X_0(n)$ of the stochastic difference equation (1) is a $\mathcal{F}_0$-measurable random process $X_0(n)$ satisfying*

$$X_0(n) = \max\left\{O_p(1), o_p\left(\kappa_n^{1/2}\right)\right\}, \ \ \textit{where} \ \ \kappa_n := \left||\rho_n| - 1\right|^{-1} \wedge n. \tag{4}$$

Under C_0(iii) *assume that $X_0(n) \to_p X_0$ where X_0 is a $\mathcal{F}_0$-measurable random variable.*

We provide a brief discussion of the model in (1) and Assumptions 1-3. The process generated by (1) consists of all types of first-order autoregressive processes employed in statistics and econometrics. The parametrisation of Assumption 1b follows Andrews and Guggenberger (2012). The class C(i) of near-stationary processes consists of the subclass of autoregressions in (1) that behave asymptotically as ergodic processes, in the sense that $n^{-1}(1-\rho_n^2)\sum_{t=1}^n x_t^2$ satisfies a law of large numbers and $n^{-1/2}(1-\rho_n^2)^{1/2}\sum_{t=1}^n x_{t-1}u_t$ satisfies a central limit theorem and consists of stable and near-stable processes. It was introduced by Phillips and Magdalinos (2007) and the autoregressive parametrisation was generalised by Giraitis and Phillips (2006) and Andrews and Guggenberger (2012). Limit theory of non-linear functionals of near-stationary processes has been derived by Duffy and Kasparis (2021). For the class C(ii) of near-nonstationary processes, introduced by Phillips (1987b) and Chan and Wei (1987), the above ergodicity property is lost and limit theory of estimation and inference is non-Gaussian. The class C(iii) constitutes the class of first-order autoregressive processes exhibiting stochastic exponential growth: Phillips and Magdalinos (2007) show that processes in C(iii) satisfy $x_n \asymp (\rho_n - 1)^{-1/2} \rho_n^n$ when $\rho_n \to 1$, the same rate that applies under the prototypical explosive autoregression C_0(iii) of White (1958) and Anderson (1959) which is a subclass of C(iii). The validity of CI methods for an autoregressive parameter in $(-1, 1]$ (covering the autoregressive regions C(i) and the part of C_+(ii) to the left of unity) has been established by Mikusheva (2007) and by Andrews and Guggenberger (2014); the current paper proposes a CI for the autoregressive parameter with uniform coverage probability over arbitrary closed subintervals of $(-\infty, \infty)$.

The modelling choice on the intercept μ in (1) yields an autoregressive process

$$
\begin{aligned}
x_t &= \mu (1 - \rho_n) + \rho_n x_{t-1} + u_t && (5)\\
&= \mu + (X_0(n) - \mu)\, \rho_n^t + x_{0t}, \quad x_{0t} = \textstyle\sum_{j=1}^n \rho_n^{t-j} u_j && (6)
\end{aligned}
$$

where x_{0t} denotes the autoregression (1) when $\mu = 0$ and $X_0 = 0$. This autoregressive specification, designed to introduce an intercept while maintaining the stochastic structure of a nonstationary autoregression[3] by reducing the contribution of the intercept as the autoregressive parameter approaches unity, is standard in the literature: see Andrews (1993), Mikusheva (2007), Andrews and Guggenberger (2009, 2014). Since $\rho_n \to -1$ does not increase the order of magnitude of the non-stochastic component of x_t that is driven by the intercept (as happens when $\rho_n \to 1$), no adjustment is required in the nonstationary oscillating case.

Assumption 2 requires u_t to be a conditionally homoskedastic[4] martingale difference sequence that satisfies a uniform integrability assumption for (u_t^2). The above conditions guarantee the validity of a law of large numbers $n^{-1} \sum_{t=1}^n u_t^2 \to_{L_1} \sigma^2$ and a functional central limit theorem for the partial sum process of (u_t). Condition (3) together with $\mathbb{E}_{\mathcal{F}_{j-1}}\left(u_j^2\right) = \sigma^2$ ensure that in the explosive case C_0(iii) the random variable

$$
X_\infty = \left(\rho^2 - 1\right)^{1/2} \left(\textstyle\sum_{j=1}^\infty \rho^{-j} u_j + X_0 - \mu\right) \qquad (7)
$$

is non-zero *a.s.*: see Corollary 2 of Lai and Wei (1983).

An additional complication to the different rates of convergence and limit distributions among the autoregressive classes C(i)-C(iii) arises from the fact that, within class C(iii), the subclass C_0(iii) of purely explosive processes exhibits different asymptotic behaviour than mildly explosive

[3]It is wel-known that a process of the form $x_t = \mu + \rho x_{t-1} + u_t$ behaves asymptotically as a linear deterministic trend when $\rho = 1$. Our procedure for confidence interval construction can accommodate such degeneracies of autoregressive stochastic behaviour (in the sense that Theorem 1 continues to hold) but we omit the details as such deterministic trends have limited relevance for economic modelling.

[4]The main results of the paper continue to hold under stationary conditional heteroskedasticity, e.g. when u_t is a stationary GARCH process, at the cost of higher moment assumptions: see Andrews and Guggenberger (2012), Magdalinos (2020) and Hu, Kasparis and Wang (2021).

processes. The asymptotic distribution of the OLS estimator in the explosive case, when it exists, is entirely driven by the distribution of the innovation process (u_t): no central limit theory applies and sample moments converge as L_2-bounded martingales to random variables such as X_∞ in (7) whose distribution changes with the distribution of (u_t). As Anderson (1959) shows, the well known Cauchy limit distribution for the normalised and centred OLS estimator and the corresponding standard normal limit distribution for the t-statistic only apply when the innovation process u_t in (1) is i.i.d. Gaussian and the explosive time series is initialised at $X_0 = 0$. For a non-identically distributed sequence of innovations, the distributional limit of the t-statistic based on the OLS may not even exist. On the other hand, the class of mildly explosive autoregressions (processes in C(iii) satisfying $|\rho_n| \to 1$) behaves more regularly, with sample moments converging in distribution via a martingale central limit theorem established by Phillips and Magdalinos (2007) and extended in various directions by Aue and Horvath (2007), Magdalinos (2012) and Arvanitis and Magdalinos (2019). The subsequent Cauchy and standard normal limit distributions for the OLS estimator and the t-statistic respectively are invariant to the distribution of the innovations u_t, the (stationary) dependence structure of u_t and the initialisation X_0. These desirable properties of mildly explosive autoregressions are employed by our instrument in the estimation procedure of Section 3 below to "regularise" the asymptotic behaviour of sample moments generated by explosive time series into a distribution-free asymptotic mixed-Gaussian framework.

Assumption 3 is standard for C(i) and C(ii) and mildly explosive processes but removes the $X_0 = 0$ condition employed in the explosive case C_0(iii): see Wang and Yu (2015) for the effect of X_0 in the limit distributions of OLS estimators and test statistics in the explosive case.

2.2 Predictive regression framework

In many economic and financial applications, the econometric model takes the form

$$y_t = \gamma + \beta x_{t-1} + \varepsilon_t \tag{8}$$

driven by an autoregressive process x_t in (1). While the parameter of interest in such models is β and the AR root of (1) is a nuisance parameter, the validity of inference procedures on β is subject to a degree of knowledge of the stochastic properties of x_t; see e.g. Campbell and Yogo (1996). Recent inference procedures that provide valid inference on β when the process x_t lies in the regions C(i) and C_+(ii) include: Jansson and Moreira (2006), Phillips and Magdalinos (2009), Elliott et al. (2015), Magdalinos and Phillips (2020) and Hu, Kasparis and Wang (2021). The nonstationary and near-explosive regions C_-(ii) and C(iii) are not considered by the above papers, and the right side of the local-to-unity region C_+(ii) is also ruled out in most of the literature. OLS-based inference on β in the purely explosive region C_0(iii) suffers from the same problem as OLS-based inference on ρ_n, with standard inference applying only under i.i.d. Gaussian innovations ε_t. The inference procedure on β in the predictive regression model (1) and (8) proposed in this paper can accommodate regressors along the entire spectrum of autoregressive processes, as defined by Assumption 1a, and we establish its asymptotic validity uniformly over the autoregressive regime. Inference on β is possible under weaker assumptions on u_t than maintained by Assumption 2.

Assumption 4. *The innovation sequence* $(u_t)_{t\in\mathbb{N}}$ *in (1) is a stationary linear process of the form* $u_t = \sum_{j=0}^{\infty} c_j e_{t-j}$*, where* $(c_j)_{j\geq 0}$ *is a sequence of constants satisfying* $\sum_{j=0}^{\infty}|c_j| < \infty$*,* $\sum_{j=0}^{\infty} j c_j^2 < \infty$*,* $c_0 = 1$ *and* $C(1) = \sum_{j=0}^{\infty} c_j \neq 0$*; for* $|\rho| > 1$ $C_\rho(1) = \sum_{j=0}^{\infty} \rho^{-j} c_j \neq 0$*. Given a filtration* $(\mathcal{F}_t)_{t\in\mathbb{Z}}$*, the sequence* $v_t := (\varepsilon_t, e_t)'$ *is an* $\mathcal{F}_t$*-martingale difference sequence satisfying* $\mathbb{E}_{\mathcal{F}_{t-1}}(v_t v_t') = \Sigma_v > 0$ *a.s. for all* t*, (3) with* u_t *replaced by* e_t*, and* $\left(\|v_t\|^2\right)_{t\in\mathbb{Z}}$ *is a uniformly integrable sequence.*

2.3 An epidemiological model of infection growth

Variants of the susceptible-infected-removed (SIR) model, originally introduced by Kermack and McKendrick (1927), constitute the main paradigm for modelling the evolution of epidemics. In this section, we consider a standard discrete-time SIR model and demonstrate that upon linearisation around the disease-free equilibrium (DFE), whenever the model's *basic reproduction number* (BRN) is above unity, the model-implied dynamics for the number of infected will necessarily display a first-order difference equation with an explosive root, implying an exponential growth for infections at the outbreak of the epidemic. Moreover, we show that at the DFE, the dynamics of the first differences of the number of recovered and deceased are both characterised by a predictive regression with the lag of the (potentially explosive) process of infections as regressor.

We briefly describe the model below. The number of *infected*, *susceptible*, *recovered* and *deceased* individuals at time t, denoted by I_t, S_t, R_t and D_t respectively, evolves according to the following non-linear system of difference equations:

$$\begin{aligned} I_{t+1} &= I_t\left(1 + \theta S_t/N - \gamma - \delta\right) \qquad (9)\\ S_{t+1} &= S_t\left(1 - \theta I_t/N\right),\ R_{t+1} = R_t + \gamma I_t,\ D_{t+1} = D_t + \delta I_t \end{aligned}$$

with non-negative initial conditions S_0, I_0, R_0, D_0 satisfying $S_t + I_t + R_t + D_t = N$ for all t, where N denotes the constant population size (births or deaths by other causes are ruled out or cancel perfectly in each period). Since at each period, S_t is a linear combination of the remaining states $S_t = N - I_t - R_t - D_t$, we substitute this identity in the equation for I_{t+1} and work with the reduced system of I_t, R_t and D_t. The choice for removing S_t facilitates estimation since data on susceptibles are unavailable.

The model's dynamics is governed by the parameters $\theta, \gamma, \delta \in (0, 1]$: θ is the contact rate, i.e. the average number of individuals an infected person passes the infection in a period; γ is the recovery rate and δ is the death rate. There is no heterogeneity, each individual is equally likely to contract the disease with no possibility of re-infection. The model's dynamics is driven by the BRN which in the model (9) is given by

$$r_0 = \theta/(\gamma + \delta), \qquad (10)$$

measuring the number of infections per infected individual. When $r_0 \geq 1$ the underlying disease escalates into an epidemic and continues to spread and when $r_0 < 1$ the growth of infections can be contained. Epidemiologists consider r_0 the key parameter for determining whether an epidemic is controllable and for understanding its transmission mechanism.

In order to study the dynamics implied by this basic dynamic nonlinear model, we use next generation matrix (NGM) approach and linearise the system in (9) around the DFE ($I = R =$

$D = 0, S = N)$[5]. Such an approximation is accurate at early stages of an epidemic outbreak, when the number of susceptibles is large relatively to the total number of infected, recovered and deceased. The resulting linear system takes a triangular form $Y_t = JY_{t-1}$, with $Y_t = [I_t, R_t, D_t]'$ and J the Jacobian matrix evaluated at the DFE:

$$J = \begin{bmatrix} 1+\theta-\gamma-\delta & 0 & 0 \\ \gamma & 1 & 0 \\ \delta & 0 & 1 \end{bmatrix},$$

where the equation for I_t is a first-order difference equation with root $\rho = 1+\theta-\gamma-\delta$, which (in view of (10)) satisfies the following: $\rho > 1$ whenever $r_0 > 1$, $\rho = 1$ whenever $r_0 = 1$, and $\rho < 1$ whenever $r_0 < 1$. In other words, at an outbreak of an epidemic, the number of infections displays exponential growth, a general result that applies to a variety of models (see e.g. Theorem 2.1 in Allen and Van den Driessche (2008)).

The standard way to add a stochastic component to the model is by adding zero-mean measurement error to the system, which corresponds to assuming that the linearised model holds on average. The resulting stochastic system that we take to the data is

$$\begin{bmatrix} I_t \\ \Delta R_t \\ \Delta D_t \end{bmatrix} = \begin{bmatrix} 1+\theta-\gamma-\delta & 0 & 0 \\ \gamma & 0 & 0 \\ \delta & 0 & 0 \end{bmatrix} \begin{bmatrix} I_{t-1} \\ R_{t-1} \\ D_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \tag{11}$$

with stochastic dynamic behaviour, formalised by the following assumption, which combines Assumptions 1 and 3 for the autoregressive process I_t and a vector-valued version of Assumption 2 for the innovation sequence in (11).

Assumption 5. *The autoregressive parameter $\rho_n := 1+\theta-\gamma-\delta$ of I_t in (11) satisfies Assumption 1a; I_0 satisfies Assumption 3. The innovation sequence $u_t = [u_{1t}, u_{2t}, u_{3t}]'$ in (11) is an $\mathcal{F}_t$-martingale difference sequence satisfying $\mathbb{E}_{\mathcal{F}_{t-1}}(u_t u_t') = \Sigma > 0$ for all t a.s., (3) with $|u_t|$ replaced by $|u_{1t}|$ and $\left(\|u_t\|^2\right)_{t\in\mathbb{Z}}$ is a uniformly integrable sequence.*

The advantage of the inference procedure developed by this paper over existing procedures is that it is valid for any $\rho_n \to \rho \in (0,\infty)$, which includes all three parameter regions of empirical interest and relevance during the Covid-19 epidemic outbreak. While this is a simple stylised model, it serves as a demonstration of the scope of the inference procedure of this paper and the advantages that its robustness and distribution-free properties provide. We are not aware of any alternative statistical procedure which can achieve this throughout the range $\rho \in (0,\infty)$ without restricting attention to a particular region of the parameter space and without imposing parametric assumptions on the distribution of u_t in the explosive region $(1,\infty)$.

3 General asymptotic inference with autoregressions

3.1 Combined near-stationary/explosive instrumentation

This section introduces new estimators of the autoregressive root ρ_n of (1) and the slope parameter β in (8) that deliver a unified asymptotic theory of hypothesis testing and CI construction

[5] The model can be linearised at any other point ($I = iN, R = rN, D = dN, S = (1-i-r-d)N$) for fractions i, r and d of the population N of I_t, R_t and D_t respectively, but the DFE is usually the chosen for early analysis.

for ρ_n and β over the entire parameter space defined in Assumption 1a. The idea behind the estimation procedure is to filter the autoregression x_t in (1) through a time series that acts as an instrument and is constructed to behave asymptotically as: a near-stationary process when x_t belongs to the near-stationary class C(i); a mildly explosive process when x_t belongs to the near-explosive class C(iii); a (random) linear combination of the above when x_t belongs to the near-nonstationary class C(ii). In addition, the instrument process is designed to emulate the regular/oscillating behaviour of x_t for all autoregressive classes apart from the purely stationary subclass C_0(i) (where employing both regular and oscillating near-stationary instrument results to IV estimators that are asymptotically equivalent to OLS). The resulting instrumental variable estimator inherits the desirable asymptotic properties of near-stationary/mildly-explosive processes and is asymptotically mixed-Gaussian along all autoregressive classes C(i)-C(iii), independently of the distribution of the innovations u_t in (1). Large sample distributional invariance is crucial for the explosive region C_0(iii), where the asymptotic distributions of the OLS estimator and the resulting t-statistic are of unknown form that depends on the innovations' distribution.

Successful instrumentation based on a combined near-stationary/near-explosive process requires statistical information separating the near-stationary autoregressive class C(i) from the near-explosive class C(iii) in large samples. Such information is available in the OLS estimator $\hat{\rho}_n$ of ρ: for each $n \in \mathbb{N}$, define the event

$$F_n = \{n(|\hat{\rho}_n| - 1) \leq 0\}, \tag{12}$$

its complement $\bar{F}_n$, and the events $F_n^+ = F_n \cap \{\hat{\rho}_n \geq 0\}$, $F_n^- = F_n \cap \{\hat{\rho}_n < 0\}$, $\bar{F}_n^+ = \bar{F}_n \cap \{\hat{\rho}_n \geq 0\}$ and $\bar{F}_n^- = \bar{F}_n \cap \{\hat{\rho}_n < 0\}$. Clearly, $\left\{F_n^+, F_n^-, \bar{F}_n^+, \bar{F}_n^-\right\}$ is a collection of disjoint events that partitions the sample space that supports the autoregressive process x_t. Asymptotic separation of the C(i) and C(iii) classes can be achieved by employing (12): under C(i) $n(|\hat{\rho}_n| - 1) \rightarrow_p -\infty$ which implies that $\mathbf{1}_{\bar{F}_n} = 0$ for all but finitely many n with probability tending to 1 (w.p.t. 1) whereas under C(iii) $n(|\hat{\rho}_n| - 1) \rightarrow_p \infty$ which implies that $\mathbf{1}_{F_n} = 0$ for all but finitely many n w.p.t. 1. On the other hand, asymptotic separation between regular and oscillating autoregressions outside the stationary region $(-1, 1)$ may be achieved by employing the event $\{\hat{\rho}_n \geq 0\}$ and its complement. Lemma 2 below shows that this asymptotic separation, required for successful instrumentation, is achieved at arbitrary rate.

Lemma 2. *Let $(m_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of positive numbers such that $m_n \rightarrow \infty$. Under Assumption 4:* **(i)** *if $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(i) then $m_n \mathbf{1}_{\bar{F}_n^+} \rightarrow_p 0$ and $m_n \mathbf{1}_{\bar{F}_n^-} \rightarrow_p 0$;* **(ii)** *if $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(iii) then $m_n \mathbf{1}_{F_n^+} \rightarrow_p 0$ and $m_n \mathbf{1}_{F_n^-} \rightarrow_p 0$;* **(iii)** *if $\rho_n \rightarrow \rho \geq 1$ then $m_n \mathbf{1}_{F_n^-} \rightarrow_p 0$ and $m_n \mathbf{1}_{\bar{F}_n^-} \rightarrow_p 0$;* **(iv)** *if $\rho_n \rightarrow \rho \leq -1$ then $m_n \mathbf{1}_{F_n^+} \rightarrow_p 0$ and $m_n \mathbf{1}_{\bar{F}_n^+} \rightarrow_p 0$.*

We now present our instrumentation procedure. Given a sequence (v_t), we denote $\bar{v}_{n-j} := n^{-1} \sum_{t=j+1}^{n} v_{t-j}$ and $\underline{v}_{t-j} := v_{t-j} - \bar{v}_{n-j}$ for $j \in \{0, ..., t-1\}$. Subtracting $\bar{x}_n$ from (5):

$$\underline{x}_t = \rho_n \underline{x}_{t-1} + \underline{u}_t, \tag{13}$$

and the OLS estimator for ρ_n and the resulting residuals are given by

$$\hat{\rho}_n = \left(\sum_{t=1}^{n} \underline{x}_{t-1}^2\right)^{-1} \sum_{t=1}^{n} \underline{x}_t \underline{x}_{t-1} \text{ and } \hat{u}_t = \underline{x}_t - \hat{\rho}_n \underline{x}_{t-1}. \tag{14}$$

Recalling the events $\{F_n^+, F_n^-, \bar{F}_n^+, \bar{F}_n^-\}$ below (12) and letting $\nabla v_t := v_t + v_{t-1}$, we define

$$\tilde{u}_t = \Delta x_t \mathbf{1}_{F_n^+} + \nabla \underline{x}_t \mathbf{1}_{F_n^-} + \hat{u}_t \mathbf{1}_{\bar{F}_n^+} + \hat{u}_t \mathbf{1}_{\bar{F}_n^-} \tag{15}$$

$$\rho_{nz} = \varphi_{1n} \mathbf{1}_{F_n^+} + \varphi_{1n}^- \mathbf{1}_{F_n^-} + \varphi_{2n} \mathbf{1}_{\bar{F}_n^+} + \varphi_{2n}^- \mathbf{1}_{\bar{F}_n^-} \tag{16}$$

where $(\varphi_{1n})_{n\in\mathbb{N}}$ and $(\varphi_{1n}^-)_{n\in\mathbb{N}}$ are *chosen* sequences in C(i) (so that $n(|\varphi_{1n}|-1) \to -\infty$ and $n(|\varphi_{1n}^-|-1) \to -\infty$) with $\varphi_{1n} \to 1$ and $\varphi_{1n}^- \to -1$; $(\varphi_{2n})_{n\in\mathbb{N}}$ and $(\varphi_{2n}^-)_{n\in\mathbb{N}}$ are *chosen* sequences in C(iii) (so that $n(|\varphi_{2n}|-1) \to \infty$ and $n(|\varphi_{2n}^-|-1) \to \infty$) with $\varphi_{2n} \to 1$ and $\varphi_{2n}^- \to -1$. We construct an instrument process by accumulating the stochastic sequence $\tilde{u}_t$ in (15) according to a first order autoregressive process

$$\tilde{z}_t = \rho_{nz}\tilde{z}_{t-1} + \tilde{u}_t = \sum_{j=1}^t \rho_{nz}^{t-j}\tilde{u}_j \tag{17}$$

with chosen root ρ_{nz} in (16), initialised at $\tilde{z}_0 = 0$. It is easy to see that the instrument process in (17) admits an orthogonal decomposition

$$\tilde{z}_t = \tilde{z}_{1t}\mathbf{1}_{F_n^+} + \tilde{z}_{1t}^-\mathbf{1}_{F_n^-} + \tilde{z}_{2t}\mathbf{1}_{\bar{F}_n^+} + \tilde{z}_{2t}^-\mathbf{1}_{\bar{F}_n^-} \tag{18}$$

where $\tilde{z}_{1t}$ employs a root φ_{1n} chosen in the regular near-stationary region C_+(i), $\tilde{z}_{1t}^-$ employs a root φ_{1n}^- chosen in the oscillating near-stationary region C_-(i), $\tilde{z}_{2t}$ employs a root φ_{2n} chosen in the regular near-explosive region C_+(iii) and $\tilde{z}_{2t}^-$ employs a root φ_{2n}^- chosen in the oscillating near-explosive region C_-(iii):

$$\tilde{z}_{1t} = \varphi_{1n}\tilde{z}_{1t-1} + \Delta x_t \quad \text{and} \quad \tilde{z}_{1t}^- = \varphi_{1n}^-\tilde{z}_{1t-1}^- + \nabla \underline{x}_t \tag{19}$$

$$\tilde{z}_{2t} = \varphi_{2n}\tilde{z}_{2t-1} + \hat{u}_t \quad \text{and} \quad \tilde{z}_{2t}^- = \varphi_{2n}^-\tilde{z}_{2t-1}^- + \hat{u}_t. \tag{20}$$

The proposed estimator for ρ_n after instrumenting x_t by $\tilde{z}_t$ takes the form of a standard instrumental variable (IV) estimator:

$$\tilde{\rho}_n = \frac{\sum_{t=1}^n \underline{x}_t \tilde{z}_{t-1}}{\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{t-1}} = \tilde{\rho}_{1n}\mathbf{1}_{F_n^+} + \tilde{\rho}_{1n}^-\mathbf{1}_{F_n^-} + \tilde{\rho}_{2n}\mathbf{1}_{\bar{F}_n^+} + \tilde{\rho}_{2n}^-\mathbf{1}_{\bar{F}_n^-} \tag{21}$$

where $\tilde{\rho}_{jn} = \sum_{t=1}^n \underline{x}_t\tilde{z}_{jt-1}/\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{jt-1}$ and $\tilde{\rho}_{jn}^- = \sum_{t=1}^n \underline{x}_t\tilde{z}_{jt-1}^-/\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{jt-1}^-$ employ the (regular/oscillating) near-stationary instruments in (19) for $j = 1$ and the (regular/oscillating) near-explosive instruments in (20) for $j = 2$. Filtering in (17) and (21) is similar in spirit to the IVX procedure of Phillips and Magdalinos (2009) and the instrument process $\tilde{z}_{1t}$ in (18) is precisely the IVX instrument on the aforementioned paper. However, the IVX instrument $\tilde{z}_{1t}$ is designed to achieve robust inference in the C_+(i)-C_+(ii) classes of regular near-stationary and near-nonstationary processes, and it is invalid for oscillating processes in C_-(i) when $\rho = -1$ as well as for near-nonstationary processes in C_-(ii). For such cases, $\tilde{z}_{1t}^-$ is designed to use $\nabla \underline{x}_t$ instead of the first difference Δx_t as 'residuals' for the instrument construction[6] and employs a root $\varphi_{1n}^- \to -1$ so that the instrument process emulates the oscillation of the original process x_t. The selection between $\tilde{z}_{1t}$ and $\tilde{z}_{1t}^-$ becomes irrelevant when x_t belongs to the stationary subclass C_0(i) with $\rho \in (-1, 1)$ because the IV estimators $\tilde{\rho}_{1n}$ and $\tilde{\rho}_{1n}^-$ based on regular and oscillating near-stationary instruments are both asymptotically equivalent to OLS. Moreover, the IVX procedure of Phillips and Magdalinos (2009) is not suited for inference in near-explosive classes C_+(iii) and C_-(iii), and the new mildly explosive instrument process $\tilde{z}_{2t}$ in (18) is designed for conducting inference in the near-explosive class C_+(iii) and local-to-unity class C_+(ii) while the oscillating

[6] If $\rho = -1$ in (13), $\nabla \underline{x}_t = u_t - \bar{u}_n = u_t + O_p(n^{-1/2})$, behaves asymptotically as an innovation.

mildly explosive instrument $\tilde{z}_{2t}^{-}$ in (18) is designed for conducting inference in C_{-}(iii) and C_{-}(ii). The IV estimators based on $\tilde{z}_{2t}$ and $\tilde{z}_{2t}^{-}$ differ from the IVX estimator based on $\tilde{z}_{1t}$ (and its oscillating version $\tilde{z}_{1t}^{-}$) in two important ways: firstly, the instrument construction is based on the OLS residuals $\hat{u}_t$ which (unlike Δx_t and $\nabla \underline{x}_t$) approximate well the true innovation process u_t in (1) in class C(iii); secondly, a mildly explosive (instead of a near-stationary) root is employed in the instrument generation.

The genuine novelty of the approach lies not so much in the construction of these three new instrument processes, but in the data-driven combination of the novel near-explosive instruments $\tilde{z}_{2t}$ and $\tilde{z}_{2t}^{-}$ for regions C(ii) and C(iii) with the near-stationary instruments $\tilde{z}_{1t}$ and $\tilde{z}_{1t}^{-}$ in order to achieve correct asymptotic inference for autoregressive roots in $(-\infty, \infty)$ without *a priori* knowledge of which persistence region the true process belongs to.

Combining $\tilde{z}_{1t}$ with $\tilde{z}_{2t}$ (and the oscillating $\tilde{z}_{1t}^{-}$ with $\tilde{z}_{2t}^{-}$) to unify inference on both sides of (negative) unity and combining the regular instruments $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ with their oscillating versions $\tilde{z}_{1t}^{-}$ and $\tilde{z}_{2t}^{-}$ to unify inference on regular and oscillating processes is intuitively appealing but the asymptotic validity of such an approach is not obvious: the asymptotic mixed-Gaussianity (AMG) property of the estimator in (21) is established in Section 3.3. An essential step is provided by the asymptotic separation property of Lemma 2, which implies that the asymptotic behaviour of $\tilde{\rho}_n$ in (21) will be driven exclusively by: (i) the component $\tilde{\rho}_{1n}$ ($\tilde{\rho}_{1n}^{-}$) involving the near-stationary (oscillating) instrument $\tilde{z}_{1t}$ ($\tilde{z}_{1t}^{-}$) when x_t is in C_{+}(i) (C_{-}(i)) and (ii) the component $\tilde{\rho}_{2n}$ ($\tilde{\rho}_{2n}^{-}$) involving the mildly explosive (oscillating) instrument $\tilde{z}_{2t}$ ($\tilde{z}_{2t}^{-}$) when x_t is in C_{+}(iii) (C_{-}(iii)). The fact that the contribution of both components (or their oscillating versions) in the near-nonstationary case C_{+}(ii) (C_{-}(ii)) preserves the AMG property of the IV estimator in (21) is a central result of Theorem 3.

For the predictive regression model (1), (8), the instrument (17) gives rise to the estimator

$$\tilde{\beta}_n = \frac{\sum_{t=1}^{n} \underline{y}_t \tilde{z}_{t-1}}{\sum_{t=1}^{n} \underline{x}_{t-1} \tilde{z}_{t-1}} = \tilde{\beta}_{1n} \mathbf{1}_{F_n^{+}} + \tilde{\beta}_{1n}^{-} \mathbf{1}_{F_n^{-}} + \tilde{\beta}_{2n} \mathbf{1}_{\bar{F}_n^{+}} + \tilde{\beta}_{2n}^{-} \mathbf{1}_{\bar{F}_n^{-}} \tag{22}$$

where $(\tilde{\beta}_{jn}, \tilde{\beta}_{jn}^{-})$ are defined in the same way as $(\tilde{\rho}_{jn}, \tilde{\rho}_{jn}^{-})$ with $\underline{x}_t$ replaced by $\underline{y}_t$ in the numerator.

3.2 Asymptotic inference for autoregression and predictive regression

Estimation of the autoregressive root in (1) by the IV estimator in (21) has the advantage that the limit distribution of the normalised and centred estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ belongs to the mixed-Gaussian family of distributions in all C(i)-C(iii) cases, independently of the distribution of the innovations u_t and ε_t. This is in contrast to the OLS estimators which do not have a mixed-Gaussian limit distribution in the near-nonstationary cases C_{+}(ii) and C_{-}(ii) and whose asymptotic behaviour is entirely driven by the distribution of the innovations (u_t) in the explosive case C_0(iii). We defer the technical exposition of the AMG property of the IV estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ to Section 3.3. In this section, we focus on the main implication of the AMG property: that the t-statistics based on $\tilde{\rho}_n$ in (21) and $\tilde{\beta}_n$ in (22) have a standard normal limit distribution and achieves uniform inference along the entire autoregressive parameter space independently of the distribution of the innovations in (1).

Denoting the lagged data vectors $X = (x_1, ..., x_{n-1})'$, $\tilde{Z} = (\tilde{z}_1, ..., \tilde{z}_{n-1})'$ and $\underline{X} = (x_1 - \bar{x}_{n-1}, ..., x_{n-1} - \bar{x}_{n-1})'$, we define a t-statistic based on $\tilde{\rho}_n$ as follows:

$$\tilde{T}_n(\rho_n) = \frac{\left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{1/2}}{\hat{\sigma}_n} (\tilde{\rho}_n - \rho_n) \tag{23}$$

where $P_{\tilde{Z}} = \tilde{Z}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'$ and $\hat{\sigma}_n^2$ is the OLS estimator of the variance of u_t in (1). The t-statistic in (23) can be used to test hypotheses or to construct a $(1-\alpha)\,\%$ CI:

$$I_n(\tilde{\rho}_n, \alpha) = [\tilde{\rho}_n - c_n(\alpha), \tilde{\rho}_n + c_n(\alpha)], \quad c_n(\alpha) = \left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{-1/2} \Phi^{-1}(1-\alpha/2)\, \hat{\sigma}_n \tag{24}$$

where Φ denotes the $\mathcal{N}(0,1)$ distribution function. By combining the AMG property of $\tilde{\rho}_n$ (established by Theorem 3 below under Assumption 1b) and Lemma 1, the t-statistic in (23) is asymptotically $\mathcal{N}(0,1)$ along all drifting sequences of autoregressive parameters $(\rho_n)_{n\in\mathbb{N}}$ satisfying Assumption 1a. Consequently, the t-test in (23) has uniform asymptotic size and the CI in (24) have uniform asymptotic coverage over the parameter space

$$\Theta = [-M, M] \ \textit{for any}\ M > 0. \tag{25}$$

Denote by $\mathcal{R}_n = \left\{\left|\tilde{T}_n(\rho_n)\right| > \Phi^{-1}(1-\alpha/2)\right\}$ the critical region (CR) of the t-test in (23).

Theorem 1. *Consider the process (1) satisfying Assumptions 2 and 3, the process $\tilde{z}_t$ defined in (15)-(17) and the IV estimator $\tilde{\rho}_n$ in (21).*

(i) *Under Assumption 1a, the t-statistic in (23) satisfies $\tilde{T}_n(\rho_n) \to_d \mathcal{N}(0,1)$ as $n \to \infty$.*

(ii) *The CR $\mathcal{R}_n$ is asymptotically similar with correct asymptotic size over the parameter space Θ in (25):* $\liminf_{n\to\infty} \inf_{\rho\in\Theta} \mathbb{P}_\rho(\mathcal{R}_n) = \limsup_{n\to\infty} \sup_{\rho\in\Theta} \mathbb{P}_\rho(\mathcal{R}_n) = \alpha$.

(iii) *The CI in (24) satisfies* $\lim_{n\to\infty} \inf_{\rho\in\Theta} \mathbb{P}_\rho[\rho \in I_n(\tilde{\rho}_n, \alpha)] = 1 - \alpha$.

Remarks.

1. Theorem 1 shows that the methodology of the paper delivers uniform and distribution-free inference over an autoregressive parameter space in (25) that consists of arbitrarily large closed subintervals of $(-\infty, \infty)$. In the terminology of Andrews, Cheng and Guggenberger (2020), the CR $\mathcal{R}_n$ and the CI $I_n(\tilde{\rho}_n, \alpha)$ are uniformly asymptotically similar over Θ. To our knowledge, this is the first procedure that provides a unified framework of inference and CI construction when data originate from autoregressive time series encompassing the stationary, nonstationary, explosive and all intermediate regions described in C_+(i)-C_+(iii) as well as their oscillating counterparts in C_-(i)-C_-(iii), without *a priori* knowledge or the need for pre-testing.

2. It is possible to extend the uniformity in Theorem 1 over a space of distribution functions on $\mathbb{R}^\infty$ of the innovation sequence $(u_t)_{t\in\mathbb{N}}$. The details of such an extension are elaborate and we omit them for brevity.

3. The unified asymptotic inference framework provided by Theorem 1 is achieved due to the crucial AMG property of the IV estimator $\tilde{\rho}_n$ in (21), established by Theorem 3. The instrumentation by a combination of a near-stationary and mildly explosive process and their oscillating counterparts in (17) serves this purpose by design: it employs information from a non-AMG procedure (the OLS estimator is not AMG in regions C(ii) and C_0(iii)) to construct an estimator (21) that enjoys the AMG property across all autoregressive classes C(i)-C(iii).

4. The inferential framework of (21), (23) and (24) constitutes the first procedure that achieves inference with general asymptotic validity in the explosive region C_0(iii). This provides a solution to a long-standing problem in explosive autoregression, pointed out by Anderson (1959), namely that the asymptotic distribution of estimators and tests based on least squares (when they exist) are entirely driven by the distribution of the innovations (u_t) in (1). Wang and Yu (2015) derive explicit expressions of the dependence of the standard OLS t-statistic limit distribution on the distribution of the innovations of (1) and the initial condition X_0 for the case $\rho > 1$. As Theorem 3 shows, the IV estimator $\tilde{\rho}_n$ in (21) has the AMG property irrespective of the distribution of (u_t) giving rise to a distribution-free and correctly-sized asymptotic CI in (24). To our knowledge, the t-statistic in (23) and the associated CI in (24) provide the first solution to the problem of distribution-free asymptotic inference in the explosive (oscillating) autoregression.

5. The asymptotic normality result of Theorem 1 includes oscillating sequences for ρ_n under Assumption 1a for which the t-statistic based on the OLS estimator may not converge in distribution. As an example, consider the sequence $(\rho_n)_{n\in\mathbb{N}}$ in (2) with $k_n = n$. The standard t-statistic $T_n(\rho_n)$ based on the OLS estimator $\hat{\rho}_n$ does not converge in distribution: $T_{2n}(\rho_{2n}) \to_d R_1$ and $T_{2n-1}(\rho_{2n-1}) \to_d R_{-1}$ where $R_c = \left(\sigma^2 \int_0^1 J_c(r)^2 dr\right)^{-1/2} \int_0^1 J_c(r)\, dB(r)$. While the IV estimator $\tilde{\rho}_n - \rho_n$ in (21) also has, after normalisation, two accumulation points in distribution, both distributions have the AMG property; as a result, both subsequences $\tilde{T}_{2n}(\rho_{2n})$ and $\tilde{T}_{2n-1}(\rho_{2n-1})$ converge in distribution to $\mathcal{N}(0,1)$, implying that $\tilde{T}_n(\rho_n) \to_d \mathcal{N}(0,1)$. The proof of Theorem 1 employs Lemma 1 to show that the above asymptotic behaviour of the t-statistic in (23) is typical and only requires the weaker Assumption 1a.

6. The generality of our methodology makes it suitable for various empirical application such as testing for episodes of bubbles in financial asset prices, where the existing approaches (e.g. Phillips and Yu (2011), Phillips et al. (2011)) model bubbles as mildly explosive episodes but assume away the purely explosive case $\rho > 1$, due to lack of asymptotic validity of existing approaches in this region. Another important application involves the stochastic evolution of epidemics, e.g. Covid-19, where the basic reproduction number r_0 of infections has widely been reported in the explosive region in highly infectious periods and in the stationary region in periods of remission; see Section 5 for further details.

For the predictive regression model in (1) and (8), we employ a similar studentisation to (23) based on the IV estimator $\tilde{\beta}_n$ in (22):

$$\tilde{T}_n(\beta_n) = \frac{\left(\underline{X}' P_{\underline{\tilde{Z}}} \underline{X}\right)^{1/2}}{\hat{\sigma}_\varepsilon} \left(\tilde{\beta}_n - \beta_n\right) \tag{26}$$

where $\hat{\sigma}_\varepsilon^2$ is the OLS estimator of the variance of ε_t in (8), $\sigma_\varepsilon^2 = \mathbb{E}(\varepsilon_t^2)$. While the t-statistic in (26) is shown to be asymptotically standard normal in Theorem 2 below, the estimation of the intercept in (8) induces a finite sample size distortion when x_t has a positive unit root and a near-stationary instrument is employed, as documented by Kostakis et al. (2015), Hosseinkouchack and Demetrescu (2021) and Harvey, Leybourne and Taylor (2021). The problem occurs because the sample moment that drives mixed normality is given by $\sum_{t=1}^n \tilde{z}_{1t-1}\varepsilon_t - n\bar{z}_{1n-1}\bar{\varepsilon}_n$ and, while the first term on the right-hand side dominates and is asymptotically normally distributed, $n\bar{z}_{n-1}\bar{\varepsilon}_n$ is not asymptotically mixed-Gaussian and has more pronounced finite sample effects when x_t is a positive unit root process (in which case $n\bar{x}_{n-1}\bar{\varepsilon}_n$ contributes to the OLS limit distribution, see Remark 2 below). Given that the finite sample distortion only occurs very close to 1, one solution is to employ the fully-modified (FM) transformation of Phillips and Hansen (1990) that orthogonalises the innovations ε_t of (8) with respect to the innovations u_t of (1) and, hence, transform the non-AMG lower order term $\bar{z}_{n-1}\bar{\varepsilon}_n$ into an AMG component for regressors very close to a unit root. The FM-corrected IV estimator $\tilde{\beta}_{1n}$ in (22) (the component of $\tilde{\beta}_n$ generated by a near-stationary regular instrument) takes the form

$$\beta^*_{1n} = \left(\sum_{t=1}^n \underline{y}_t \tilde{z}'_{1t-1} + \hat{\rho}_{\varepsilon u}\frac{\hat{\sigma}_\varepsilon}{\hat{\omega}_u} x_n \bar{z}'_{1n-1}\right)\left(\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}'_{1t-1}\right)^{-1}$$

where $\hat{\sigma}^2_\varepsilon$, $\hat{\omega}^2_u$ and $\hat{\rho}_{\varepsilon u}$ are consistent estimators of σ^2_ε, $\omega^2_u = \sum_{k=-\infty}^{\infty}\mathbb{E}\left(u_t u_{t-k}\right)$ and $\rho_{\varepsilon u} = corr\left(\varepsilon_t, u_t\right)$. With the above correction, the IV estimator of β becomes

$$\beta^*_n = \mathbf{1}_{F_n^+}\beta^*_{1n} + \tilde{\beta}^-_{1n}\mathbf{1}_{F_n^-} + \tilde{\beta}_{2n}\mathbf{1}_{\bar{F}_n^+} + \tilde{\beta}^-_{2n}\mathbf{1}_{\bar{F}_n^-} \tag{27}$$

with $\tilde{\beta}^-_{1n}$, $\tilde{\beta}_{2n}$ and $\tilde{\beta}^-_{2n}$ are defined as in (22). A computation of the standard error of the estimator β^*_n above gives rise to the following t-statistic:

$$T^*_n\left(\beta_n\right) = N_n\left(\beta^*_n - \beta_n\right), \quad N_n := \hat{\sigma}_\varepsilon^{-1}\left(\sum_{t=1}^n \tilde{z}^2_{t-1} - n\bar{z}^2_{1,n-1}\left(1-\hat{\rho}^2_{\varepsilon u}\right)\mathbf{1}_{F_n^+}\right)^{-1/2}\sum_{j=1}^n \underline{x}_{j-1}\tilde{z}_{j-1}. \tag{28}$$

Denote by $\tilde{\mathcal{R}}_n = \left\{\left|\tilde{T}_n\left(\beta_n\right)\right| > \Phi^{-1}\left(1-\alpha/2\right)\right\}$ and $\mathcal{R}^*_n = \left\{\left|T^*_n\left(\beta_n\right)\right| > \Phi^{-1}\left(1-\alpha/2\right)\right\}$ the CRs of the t-tests in (26) and (28) and the corresponding CIs by $I_n\left(\tilde{\beta}_n,\alpha\right)$ and $I_n\left(\beta^*_n,\alpha\right)$: $I_n\left(\tilde{\beta}_n,\alpha\right)$ is defined as in (24) with $\left(\tilde{\rho}_n,\hat{\sigma}_n\right)$ replaced by $\left(\tilde{\beta}_n,\hat{\sigma}_\varepsilon\right)$; $I_n\left(\beta^*_n,\alpha\right)$ is defined as in (24) with $\left(\tilde{\rho}_n, c_n\left(\alpha\right)\right)$ replaced by $\left(\beta^*_n, \Phi^{-1}\left(1-\alpha/2\right)N_n^{-1}\right)$.

Theorem 2. *Consider the predictive regression model (1) and (8) satisfying Assumptions 3 and 4, the filtered process $\tilde{z}_t$ defined by (15)-(17) and the IV estimators $\tilde{\beta}_n$ and β^*_n in (22) and (27).*

(i) *Under Assumption 1a, the statistics in (26) and (28) satisfy $\tilde{T}_n\left(\beta_n\right) \to_d \mathcal{N}\left(0,1\right)$ and $T^*_n\left(\beta_n\right) \to_d \mathcal{N}\left(0,1\right)$.*

(ii) *The CRs $\tilde{\mathcal{R}}_n$ and $\mathcal{R}^*_n$ associated with (26) and (28) are asymptotically similar with correct asymptotic size uniformly over Θ in (25): for $R_n \in \left\{\tilde{\mathcal{R}}_n, \mathcal{R}^*_n\right\}$*

$$\liminf_{n\to\infty}\inf_{\rho\in\Theta}\mathbb{P}_\rho\left(R_n\right) = \limsup_{n\to\infty}\sup_{\rho\in\Theta}\mathbb{P}_\rho\left(R_n\right) = \alpha.$$

(iii) *The CIs $I_n\left(\tilde{\beta}_n,\alpha\right)$ and $I_n\left(\beta^*_n,\alpha\right)$ associated with (26) and (28) have correct asymptotic coverage uniformly over Θ: $\inf_{\rho\in\Theta}\mathbb{P}_\rho[\beta \in I_n(\tilde{\beta}_n,\alpha)]$ and $\inf_{\rho\in\Theta}P_\rho\left[\beta \in I_n\left(\beta^*_n,\alpha\right)\right]$ both converge to $1-\alpha$ as $n\to\infty$.*

Remarks.

1. The asymptotic inference based on the t-statistics in (26) and (28) is invariant to the stochastic properties of the autoregression in (1) and is shown to give rise to correct asymptotic size and coverage probability uniformly over the autoregressive parameter space Θ. To our knowledge, this is the first inference procedure that accommodates this level of generality. Extension of the uniformity over a space of distribution functions of $\left(u_t\right)_{t\in\mathbb{N}}$ and $\left(\varepsilon_t\right)_{t\in\mathbb{N}}$ is possible (see Remark 2 to Theorem 1) but omitted for brevity.

2. While $T_n^*(\beta_n)$ and $\tilde{T}_n(\beta_n)$ have the same limit distribution, the test based on $\tilde{T}_n(\beta_n)$ may suffer from finite sample distortion due to the fact that the estimation of the intercept γ in (8) does not feature in the first-order asymptotic theory. This only becomes an issue under C_+(ii) where estimation of γ features more prominently: in particular, the contribution of the non-AMG term $n\bar{z}_{1n-1}\bar{\varepsilon}_n$ is not reflected in the limit distribution of Theorem 2. While this contribution is $o_p(1)$, $n\pi_n^{-1}\bar{z}_{1n-1}\bar{\varepsilon}_n = O_p(n^{-1/2}(1-\varphi_{1n})^{-1/2})$ under C_+(ii) in the notation of Theorem 3, $n\bar{z}_{1n-1}\bar{\varepsilon}_n$ is asymptotically equivalent to $(1-\varphi_{1n})^{-1}x_n\sum_{t=1}^n\varepsilon_t$ and the correlation between x_n and $\sum_{t=1}^n\varepsilon_t$ distorts mixed-Gaussianity in finite samples. As a result, the t-statistic based on $\tilde{T}_n(\beta_n)$ exhibits finite sample distortions when the following occur *jointly*: (i) the autoregressive root of x_t is very close to 1; (ii) $\rho_{\varepsilon u} = corr(\varepsilon_t, u_t)$ is close to 1 in absolute value; (iii) φ_{1n} is chosen close to 1. The FM transformation of Phillips and Hansen (1990), $\varepsilon_{0t} = \varepsilon_t - \omega^{-1}\mathbb{E}(\varepsilon_t u_t)u_t$, orthogonalises $n^{-1/2}\sum_{t=1}^n\varepsilon_{0t}$ and $n^{-1/2}\sum_{t=1}^n u_t$ asymptotically when x_t is a unit root process and transforms the non-AMG term $n\bar{z}_{1n-1}\bar{\varepsilon}_n$ into a AMG term $n\bar{z}_{1n-1}\bar{\varepsilon}_{0n}$ with a remainder that becomes smaller the closer x_t is to a unit root process, thereby addressing the issues in (i) and (ii) above simultaneously. The estimator β_n^* arising from employing the FM transformation and the corresponding t-statistic $T_n^*(\beta_n)$ have significantly improved finite sample properties whenever ρ_n is close to 1 with large $|\rho_{\varepsilon u}|$, while both $\tilde{T}_n(\beta_n)$ and $T_n^*(\beta_n)$ perform equally well in all other cases. It is worth noting that the terms arising from the estimation of the intercept in both IV and OLS, $n\bar{z}_{1n-1}\bar{\varepsilon}_n$ and $n\bar{x}_{n-1}\bar{\varepsilon}_n$, have reduced order of magnitude for an autoregressive root close to -1, so no finite sample adjustment is necessary under C_-(ii).

3. Practical implementation of the test procedures of Theorems 1 and 2 requires a choice for φ_{1n}, φ_{1n}^-, φ_{2n} and φ_{2n}^- in (15) for the construction of the instrument $\tilde{z}_t$. Since our procedure is designed to work across the autoregressive parameter space $(-\infty,\infty)$, we require a single instrument that will perform well across C(i)-C(iii) both for the autoregression and predictive regression problems. We base our choice for φ_{1n} and φ_{2n} on the principle of controlling the worst finite sample distortion that occurs in the case of a unit root regressor with large correlation $|\rho_{\varepsilon u}|$ (Remark 2 above). We conduct a grid search Monte Carlo to select the maximal values of φ_{1n} and φ_{2n} (by Theorem 3, these achieve maximal power) subject to a satisfactory test size in the above least favourable case and we set $\varphi_{1n}^- = -\varphi_{1n}$ and $\varphi_{2n}^- = -\varphi_{2n}$; a detailed analysis of the choice of φ_{1n} and φ_{2n} can be found in Section 4.1. We demonstrate that the proposed choice of instrument in Section 4.1 works very well (in terms of size and power) both in the autoregression and in the predictive regression setups and across all persistence regions considered.

4. The above methodology may be extended to multivariate predictive regression models where both x_t and y_t in (1) and (8) are vector-valued and the statistical problem consists of testing a set of q restrictions on $vec(\beta)$. A model along the lines of Magdalinos and Phillips (2020) (that assumes away cointegrating relationships between elements of the VAR(1) process for x_t) extended to account for regressors with autoregressive eigenvalues in $(-\infty,\infty)$ may be considered with the asymptotically $\mathcal{N}(0,1)$ t-statistics of Theorem 2 replaced by asymptotically $\chi^2(q)$ Wald statistics based on the combined (vector-valued) instrument (15)-(17) of Section 3.1. The fact that the methodology of this paper extends directly to multivariate systems is a major advantage over existing methods, including Campbell and Yogo (2006) and Elliott et al. (2015). A multivariate extension is not pursued here as it would be a deviation from the main focus of the paper (the construction of CIs for ρ and β with uniform asymptotic validity). Another major advantage of our inference procedures in Theorems 1 and 2 is their simplicity and ease of implementation: they employ closed-form linear estimators and $\mathcal{N}(0,1)$ critical values, rendering implementation of the procedures by practitioners natural and straightforward.

We now briefly turn to the problem of conducting inference for the parameters of the epidemiological model in (11) and, in particular, of constructing robust CIs for the BRN r_0 in (10) regardless of whether r_0 is above, equal or below unity. Denoting the autoregressive parameter of I_t in the first equation of (11) by $\rho_n := 1+\theta-\gamma-\delta$, (11) can be expressed as a system of three equations, $I_t = \rho_n I_{t-1} + u_{1t}$, $\Delta R_t = \gamma I_{t-1} + u_{2t}$ and $\Delta D_t = \delta I_{t-1} + u_{3t}$, with each equation being estimated using the instrumental variable procedure in (17)-(21):

$$\tilde{\rho}_n = \frac{\sum_{t=1}^n I_t \tilde{z}_{t-1}}{\sum_{t=1}^n I_{t-1}\tilde{z}_{t-1}}, \quad \tilde{\gamma}_n = \frac{\sum_{t=1}^n \Delta R_t \tilde{z}_{t-1}}{\sum_{t=1}^n I_{t-1}\tilde{z}_{t-1}} \text{ and } \tilde{\delta}_n = \frac{\sum_{t=1}^n \Delta D_t \tilde{z}_{t-1}}{\sum_{t=1}^n I_{t-1}\tilde{z}_{t-1}} \tag{29}$$

where the instrument $\tilde{z}_t$ is constructed from the first equation of (11) by

$$\tilde{z}_t = \rho_{nz}\tilde{z}_{t-1} + \tilde{u}_{1t}, \quad \tilde{u}_{1t} = \Delta I_t \mathbf{1}_{F_n^+} + \hat{u}_{1t}\mathbf{1}_{\bar{F}_n^+},$$

$\hat{u}_{1t}$ are the OLS residuals obtained from the first equation of (11), the events F_n^+ and $\bar{F}_n^+$ are defined in (12), ρ_{nz} is chosen as in (15) and we impose that $\mathbf{1}_{F_n^-} = \mathbf{1}_{\bar{F}_n^-} = 0$ since $\rho \geq 0$. The remaining parameters r_0 and θ may be estimated from the identity $r_0 = 1 + (\rho_n - 1)/(\gamma+\delta)$ (obtained by dividing ρ_n by $\gamma+\delta$) as:

$$\tilde{r}_n = 1 + (\tilde{\rho}_n - 1)/\left(\tilde{\gamma}_n + \tilde{\delta}_n\right) \text{ and } \tilde{\theta}_n = \tilde{\rho}_n - 1 + \tilde{\gamma}_n + \tilde{\delta}_n \tag{30}$$

where $\tilde{\rho}_n$, $\tilde{\gamma}_n$ and $\tilde{\delta}_n$ are the IV estimators in (29). Adjusting for the asymptotic variance of $\tilde{r}_n$ and $\tilde{\theta}_n$, we may construct studentised version of these estimators as follows:

$$\left[T_n(\tilde{r}_n), T_n\left(\tilde{\theta}_n\right), T_n(\tilde{\gamma}_n), T_n\left(\tilde{\delta}_n\right)\right] = (X'P_{\tilde{Z}}X)^{1/2}\left[\frac{\tilde{r}_n - r_0}{\hat{\sigma}_{r_0}}, \frac{\tilde{\theta}_n - \theta}{\hat{\sigma}_\theta}, \frac{\tilde{\gamma}_n - \gamma}{\hat{\sigma}_\gamma}, \frac{\tilde{\delta}_n - \delta}{\hat{\sigma}_\delta}\right] \tag{31}$$

where $X = [I_1, ..., I_{n-1}]'$, $\tilde{Z} = [\tilde{z}_1, ..., \tilde{z}_{n-1}]'$, $\hat{\sigma}_{r_0}^2 = \hat{v}_n'\hat{\Sigma}_n\hat{v}_n$, $\hat{\sigma}_\theta^2 = \iota'\hat{\Sigma}_n\iota$, $\hat{\sigma}_\gamma^2 = e_2'\hat{\Sigma}_n e_2$, $\hat{\sigma}_\delta^2 = e_3'\hat{\Sigma}_n e_3$, $\hat{\Sigma}_n = n^{-1}\sum_{t=1}^n \hat{u}_t\hat{u}_t'$ with $\hat{u}_t$ denoting the OLS residuals of (11), $\iota = [1,1,1]'$, $e_2 = [0,1,0]'$, $e_3 = [0,0,1]'$ and $\hat{v}_n = \left[1/(\hat{\gamma}_n + \hat{\delta}_n), (1-\hat{\rho}_n)/(\hat{\gamma}_n+\hat{\delta}_n)^2, (1-\hat{\rho}_n)/(\hat{\gamma}_n+\hat{\delta}_n)^2\right]'$ based on the OLS estimators $\hat{\rho}_n$, $\hat{\gamma}_n$ and $\hat{\delta}_n$ in (11). Letting $c_n^g(\alpha) = (X'P_{\tilde{Z}}X)^{-1/2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\hat{\sigma}_g$ for $g \in \{r_0, \theta, \gamma, \delta\}$ and denoting $[a \pm b] = [a-b, a+b]$ for brevity, we may construct CIs based on the studentised estimators in (31): $I_n(\tilde{r}_n, \alpha) = [\tilde{r}_n \pm c_n^{r_0}(\alpha)]$, $I_n(\tilde{\theta}_n, \alpha) = [\tilde{\theta}_n \pm c_n^\theta(\alpha)]$, $I_n(\tilde{\gamma}_n, \alpha) = [\tilde{\gamma}_n \pm c_n^\gamma(\alpha)]$ and $I_n\left(\tilde{\delta}_n, \alpha\right) = [\tilde{\delta}_n \pm c_n^\delta(\alpha)]$. The analysis leading to Theorem 1 yields the following.

Corollary 1. *Consider the model (I_t, R_t, D_t) in (11) satisfying Assumption 5 with parameters r_0, θ, γ and δ estimated in (29) and (30). The t-statistics in (31) all converge in distribution to $\mathcal{N}(0,1)$ and the CIs $I_n(\tilde{r}_n, \alpha)$, $I_n(\tilde{\theta}_n, \alpha)$, $I_n(\tilde{\gamma}_n, \alpha)$ and $I_n(\tilde{\delta}_n, \alpha)$ all have asymptotic probability of containment equal to $1-\alpha$.*

3.3 Asymptotic mixed-normality of the IV estimator

In this section, we establish the AMG property of the IV estimators of the paper. We begin by focusing on the regular regimes C_+(i)-C_+(iii): we first provide a brief discussion of the behaviour of the instrument process $\tilde{z}_t$ in (17) and then establish formally the asymptotic behaviour of sample moments of the instrument under C_+(i)-C_+(iii) (Lemmata 3-5). In Lemma 6, we show how the asymptotic behaviour of the oscillating classes C_-(i)-C_-(iii) may be derived from that of their regular counterparts C_+(i)-C_+(iii) via a simple transformation. The results in Lemmata 3-6 provide an insight into the mechanics of the instrument $\tilde{z}_t$ and facilitate the proof of the main result of this section: the AMG property of the normalised and centred IV estimators in (21), (22) and (27), formally stated in Theorem 3 below.

While the artificial instrument's autoregressive roots φ_{1n} and φ_{2n} in (16) may be chosen freely within the near-stationary/near-explosive range, the processes $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ in (19) are not near-stationary/near-explosive because the residuals Δx_t and $\hat{u}_t$ used in the instrument construction are not innovations. For x_t in the classes C_+(i)-C_+(ii), Magdalinos and Phillips (2020) show that: (i) $\tilde{z}_{1t}$ can be asymptotically approximated by a near-stationary process

$$z_{1t} = \varphi_{1n} z_{1t-1} + u_t = \sum_{j=1}^{t} \varphi_{1n}^{t-j} u_j \tag{32}$$

when the instrument in (19) is less persistent than the original process x_t in (1) (i.e. when ρ_n is closer to 1 than φ_{1n}) and (ii) $\tilde{z}_{1t}$ reduces asymptotically to the original process x_t (necessarily near-stationary by the choice of $(\varphi_{1n})_{n\in\mathbb{N}}$ in C_+(i)) when φ_{1n} is closer to 1 than ρ_n. The above property is a consequence of employing Δx_t is the construction of $\tilde{z}_{1t}$. On the other hand, as a consequence of employing the OLS residuals $\hat{u}_t$ in their construction, the instruments $\tilde{z}_{2t}$ and $\tilde{z}_{2t}^-$ in (20) are always approximated by the mildly explosive processes

$$z_{2t} = \varphi_{2n} z_{2t-1} + u_t \quad \text{and} \quad z_{2t}^- = \varphi_{2n}^- z_{2t-1}^- + u_t \tag{33}$$

in all sample moments. A precise statement on the approximation of $\tilde{z}_{2t}$ by z_{2t} can be found in part (iv) of Lemma B2 in Appendix B.

By Lemma 2, sample moments involving the near-stationary instrument $\tilde{z}_{1t}$ contribute asymptotically when the original process x_t belongs to the classes C_+(i)-C_+(ii) whereas sample moments involving the mildly-explosive instrument $\tilde{z}_{2t}$ make an asymptotic contribution for the classes C_+(ii)-C_+(iii). The next two results, Lemmata 3 and 4, discuss the asymptotic behaviour of sample moments involving $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ for classes C_+(i)-C_+(ii) and C_+(ii)-C_+(iii) respectively.

Under Assumption 4, denote the autocovariance function and long-run variance of (u_t) by $\gamma_u(\cdot)$ and $\omega^2 = \sum_{k=-\infty}^{\infty} \gamma_u(k) = C(1)^2 \sigma^2$ respectively and let

$$\Gamma_n = \sum_{k=1}^{\infty} \rho_n^{k-1} \gamma_u(k) \quad \text{and} \quad \Gamma = \lim_{n\to\infty} \Gamma_n = \sum_{k=1}^{\infty} \rho^{k-1} \gamma_u(k) \tag{34}$$

with $\Gamma \in \mathbb{R}$ by Assumptions 1a, 4 and the dominated convergence theorem. When $\rho = 1$, $\Gamma = \sum_{k=1}^{\infty} \gamma_u(k)$ is the one-sided long-run covariance of (u_t). Let $W(t)$ denote a standard Brownian motion on $[0,1]$ and $B(t) = \omega W(t)$; when $c \in \mathbb{R}$ in Assumption 1b, define

$$W_c(t) = \int_0^t e^{c(t-s)} dW(s), \; J_c(t) = \int_0^t e^{c(t-s)} dB(s) \text{ and } K_c = \int_0^1 J_c(r)\, dB(r) / \int_0^1 J_c(r)^2 dr. \tag{35}$$

Lemma 3. *The following hold under Assumptions 3 and 4 and C_+(i)-C_+(ii) of Assumption 1b:*

(i) $n^{-1}(1-\rho_n^2\varphi_{1n}^2) \sum_{t=1}^{n} \underline{x}_{t-1} \tilde{z}_{1t-1} = \tilde{\Psi}_n + o_p(1) \to_d \tilde{\Psi}(c)$ *where*

$\tilde{\Psi}(c) = \sigma^2 + 2\rho\Gamma + \left(J_c(1)^2 - 2J_c(1)\int_0^1 J_c(r)\,dr\right)\mathbf{1}\{c \in \mathbb{R}\}$, x_{0t} is defined in (6) and

$$\begin{aligned}\tilde{\Psi}_n &= (1+\rho_n)\left[\sigma^2 + 2\rho_n\Gamma_n + (2\rho_n - 1)\left(n^{-1}\textstyle\sum_{t=1}^n x_{0t-1}u_t - \Gamma_n\right)\right] \\ &\quad -\rho_n\left(1-\rho_n^2\right)n^{-1}\textstyle\sum_{t=1}^n x_{0t-1}^2 - 2\left(n^{-1/2}x_{0n}\right)n^{-3/2}\textstyle\sum_{j=1}^n x_{0j-1}. \end{aligned} \tag{36}$$

(ii) $n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)\sum_{t=1}^n \tilde{z}_{1t}^2 \to_p \sigma^2 + 2\rho\Gamma$

(iii) $n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{1/2}\sum_{t=1}^n \tilde{z}_{1t-1}e_t \to_d \mathcal{N}\left(0, \left(\sigma^2+2\rho\Gamma\right)\sigma_e^2\right)$

where Γ_n and Γ are defined in (34), $J_c(\cdot)$ in (35) and $\sigma_e^2 = \mathbb{E}e_t^2$.

Next, we turn to the discussion of the asymptotic behaviour of sample moments of $\tilde{z}_{2t}$. In order to maintain a common asymptotic development for autoregressions in the near-nonstationary and near-explosive classes C(ii)-C(iii), we define the convergence rates

$$\nu_n = \left(\rho_n^2 - 1\right)^{-1/2}|\rho_n|^n\,\mathbf{1}\{c=\infty\} + n^{1/2}\mathbf{1}\{c \in \mathbb{R}\} \quad \text{and} \quad \nu_{n,z} = \left(\phi_{2n}^2 - 1\right)^{-1/2}|\phi_{2n}|^n \tag{37}$$

where c denotes the limit in Assumption 1b and $\phi_{2n} = \varphi_{2n}\mathbf{1}\{\rho \geq 0\} + \varphi_{2n}^{-}\mathbf{1}\{\rho < 0\}$, and

$$s_n = \left(\rho_n\phi_{2n} - 1\right)^{-1}\nu_{n,z}\nu_n. \tag{38}$$

Under C_+(ii)-C_+(iii), the limit theory for the mildly explosive instrument's sample moments will be driven by the stochastic sequences

$$\left[Y_n, Y_n^{\varepsilon}, Z_n\right] := \left(\varphi_{2n}^2 - 1\right)^{1/2}\left[\textstyle\sum_{t=1}^n \varphi_{2n}^{-(n-t)-1}u_t, \sum_{t=1}^n \varphi_{2n}^{-(n-t)-1}\varepsilon_t, \sum_{j=1}^n \varphi_{2n}^{-j}u_j\right]. \tag{39}$$

By Anderson (1959), Phillips (1987b) and Phillips and Magdalinos (2007), the autoregressive sample moments will be driven by

$$X_n := \frac{x_n}{\nu_n} = \left(\rho_n^2 - 1\right)^{1/2}\left(\textstyle\sum_{j=1}^n \rho_n^{-j}u_j + X_0(n) - \mu\right)\mathbf{1}\{c = \infty\} + \frac{x_n}{\sqrt{n}}\mathbf{1}\{c \in \mathbb{R}\}. \tag{40}$$

The following result characterises the joint asymptotic behaviour of the sequences Y_n, Z_n and X_n and the instrument sample moment asymptotics for the autoregressive classes C_+(ii)-C_+(iii).

Lemma 4. *Let X_∞ be the random variable defined in (7) and Y_n, Y_n^{ε}, Z_n and X_n be the stochastic sequences in (39) and (40), let Y, Z, X denote $\mathcal{N}(0,\omega^2)$ random variables and let Y^{ε}* be *$\mathcal{N}(0,\sigma_{\varepsilon}^2)$. Under Assumptions 3 and 4 and C_+(ii)-C_+(iii) of Assumption 1b, the following hold as $n \to \infty$:*

(i) $[Y_n, Z_n] \to_d [Y, Z]$, $[Y_n^{\varepsilon}, Z_n] \to_d [Y^{\varepsilon}, Z]$ *with Z independent of (Y, Y^{ε}) and*

$$\left[\frac{\varphi_{2n}^2 - 1}{\varphi_{2n}^n}\sum_{t=1}^n z_{2t-1}u_t, \frac{\left(\varphi_{2n}^2-1\right)^2}{\varphi_{2n}^{2n}}\sum_{t=1}^n z_{2t-1}^2, s_n^{-1}\sum_{t=1}^n x_{t-1}z_{2t-1}\right] = \left[Y_nZ_n, Z_n^2, X_nZ_n\right] + o_p(1).$$

(ii) *Under C_+(iii) with $\rho_n \to 1$, $[Y_n, X_n] \to_d [Y, X]$ and $[Y_n^{\varepsilon}, X_n] \to_d [Y^{\varepsilon}, X]$, with X independent of (Y, Y^{ε}).*

(iii) *Under C_+(iii) with $\rho_n \to \rho > 1$, $X_n \to_p X_\infty$, $X_\infty \neq 0$ a.s.; for any continuous function $g : \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $g(X_n)Y_n \to_d g(X_\infty)Y$ and $g(X_n)Y_n^{\varepsilon} \to_d g(X_\infty)Y^{\varepsilon}$ where both $g(X_\infty)Y$ and $g(X_\infty)Y^{\varepsilon}$ are $\mathcal{MN}\left(0, \omega^2 g^2(X_\infty)\right)$ and $\mathcal{MN}\left(0, \sigma_{\varepsilon}^2 g^2(X_\infty)\right)$ variables.*

Part (iii) of Lemma 4 deserves special attention because it establishes a CLT to a mixed-Gaussian distribution in the purely explosive case C_0(iii) when $\rho > 1$ and it is precisely the result that allows us to incorporate the purely explosive case in the distribution-free mildly explosive framework of asymptotic inference. To provide some insight into the role of the result for inference, we will see that the normalised and centred estimator $\tilde{\rho}_n$ in (21) behaves asymptotically as Y_n/X_n in Theorem 3 below. The conclusion of Lemma 4(iii) implies that the ratio Y_n/X_n has a $\mathcal{MN}\left(0, \sigma^2/X_\infty^2\right)$ limit distribution in the explosive case, establishing the AMG property of $\tilde{\rho}_n$

independently of the distribution of the innovation sequence in (1).

Establishing the AMG property of $\tilde{\rho}_n$ in the local-to-unity class C$_+$(ii) is more challenging as both components $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ of the instrument in (18) feature in the limit theory, their relative contribution weighted by the sequence of events F_n^+ and $\bar{F}_n^+$ in (12). Additional complication is introduced by the randomness of the signals' limits ($2\Psi_n \to_d \sigma^2 + J_c(1)^2$ from $\tilde{z}_{1t}$ and $X_n Z_n \to_d J_c(1) Z$ from $\tilde{z}_{2t}$) which are required to be independent from the Gaussian distributional limit of the normalised $\sum_{t=1}^n z_{t-1}u_t$ ($U(1)$ and Y below) for AMG property of $\tilde{\rho}_n$. Since, by standard local-to-unity manipulations (see Phillips (1987b) or Chan and Wei (1987)), the sequences F_n^+, Ψ_n and X_n in (12), (36) and (40) can be expressed as non-stochastic functionals of the partial sum process $B_n(\cdot)$ of u_t and $B_n(s) \Rightarrow B(s)$ on $D[0,1]$, it suffices to prove the independence of $[U(1), Y]$ and the Brownian motion B; this is established in the following result.

Lemma 5. *Define the following random elements in $D[0,1]$: $B_n(s) = n^{-1/2}\sum_{t=1}^{\lfloor ns \rfloor} u_t$, $U_n(s) = (n(1-\varphi_{1n}^2)^{-1})^{-1/2}\sum_{t=1}^{\lfloor ns \rfloor} z_{1t-1}e_t$ and $Y_n(s) = (\varphi_{2n}^2 - 1)^{1/2}\sum_{t=1}^{\lfloor ns \rfloor}\varphi_{2n}^{-(\lfloor ns \rfloor - t)-1}u_t$. Under Assumption 4, $[U_n(s), B_n(s), Y_n(s)] \Rightarrow [U(s), B(s), Y]$ on $D[0,1]$, where $U(s)$ and $B(s)$ are independent Brownian motions with $\mathbb{E}U(s)^2 = s\sigma_e^2\omega^2$ and $\mathbb{E}B(s)^2 = s\omega^2$, and Y is a $\mathcal{N}(0,\omega^2)$ random variable independent of $[U(s), B(s)]$.*

We have thus far concentrated our attention to the asymptotic behaviour of sample moments based on the regular autoregressive classes C$_+$(i)-C$_+$(iii). The asymptotic analysis of an oscillating autoregression x_t in classes C$_-$(i)-C$_-$(iii) may be derived from its regular counterpart via the transformation $x_t \mapsto (-1)^t x_t$. Denoting by x_{0t} in (6) an oscillating (zero-mean) autoregression with $\rho_n < 0$, it is easy to see that $x_t^+ := (-1)^t x_{0t}$ satisfies the recursion

$$x_t^+ = |\rho_n| x_{t-1}^+ + (-1)^t u_t, \tag{41}$$

which defines a regular autoregression since $|\rho_n| \geq 0$, and the sequence $\left\{(-1)^t u_t : t \in \mathbb{N}\right\}$ satisfies Assumption 2 and 4 whenever $(u_t)_{t\in\mathbb{N}}$ does so. The same transformation applies to the oscillating instrument processes generated by x_{0t}: denoting by $\tilde{z}_{0t}^-$ the instrument $\tilde{z}_{1t}^-$ (19) when $\mu = x_0 = 0$ (so that $\nabla \underline{x}_t = \nabla x_{0t}$) and by z_{2t}^- the oscillating mildly explosive process in (33), the transformed processes $\tilde{z}_{0t}^+ := (-1)^t \tilde{z}_{0t}^-$ and $z_{2t}^+ := (-1)^t z_{2t}^-$ satisfy

$$\tilde{z}_{0t}^+ = \left|\varphi_{1n}^-\right| \tilde{z}_{0t-1}^+ + \Delta x_t^+ \quad \text{and} \quad z_{2t}^+ = \left|\varphi_{2n}^-\right| z_{2t-1}^+ + (-1)^t u_t \tag{42}$$

where $\hat{u}_t^+ = x_t^+ - \hat{\rho}_n^+ x_{t-1}^+$ and $\hat{\rho}_n^+ = \sum_{t=1}^n x_t^+ x_{t-1}^+ / \sum_{t=1}^n \left(x_{t-1}^+\right)^2$ is the OLS estimator from (41). In other words, the transformation $x_t \mapsto (-1)^t x_t$ has the property of transforming an oscillating autoregression to a regular autoregression and an oscillating instrument to a regular instrument.

Given the event F_n in (12), define the normalisation sequence

$$\pi_n = \begin{cases} n^{1/2}\left(1 - \rho_n^2\phi_{1n}^2\right)^{-1/2}, & \text{under C(i)} \\ n^{1/2}\left(1 - \phi_{1n}^2\right)^{-1/2}\mathbf{1}_{F_n} + 2n^{1/2}\left(\phi_{2n}^2 - 1\right)^{-1/2}\mathbf{1}_{\bar{F}_n}, & \text{under C(ii)} \\ \left(\phi_{2n}^2 - 1\right)^{1/2}\left(|\rho_n|\,|\phi_{2n}| - 1\right)^{-1}\left(\rho_n^2 - 1\right)^{-1/2}|\rho_n|^n, & \text{under C(iii).} \end{cases} \tag{43}$$

where $\phi_{1n} = \varphi_{1n}\mathbf{1}\{\rho \geq 0\} + \varphi_{1n}^-\mathbf{1}\{\rho < 0\}$, $\phi_{2n} = \varphi_{2n}\mathbf{1}\{\rho \geq 0\} + \varphi_{2n}^-\mathbf{1}\{\rho < 0\}$.

Let $\tilde{\rho}_{1n}^+ = \sum_{t=1}^n x_t^+ \tilde{z}_{0t-1}^+ / \sum_{t=1}^n x_{t-1}^+ \tilde{z}_{0t-1}^+$ and $\tilde{\rho}_{2n}^+ = \sum_{t=1}^n x_t^+ z_{2t-1}^+ / \sum_{t=1}^n x_{t-1}^+ z_{2t-1}^+$ be generated by (41) and (42), define $\left(F_n^{++}, \bar{F}_n^{++}\right)$ in the same way as $\left(F_n^+, \bar{F}_n^+\right)$ with $\hat{\rho}_n$ replaced by $\hat{\rho}_n^+$ and define $(\tilde{\beta}_{1n}^+, \tilde{\beta}_{2n}^+)$ in the same way as $\left(\tilde{\rho}_{1n}^+, \tilde{\rho}_{2n}^+\right)$ with x_t^+ replaced in the numerator by $y_t^+ = (-1)^{t-1} y_t$

under the restriction $\gamma = \mu = 0$ in (1)-(8).

The following result (Lemma 6) shows how to obtain the limit distribution of the IV estimators based on an oscillating instrument $\left(\tilde{\rho}_{1n}^{-}, \tilde{\rho}_{2n}^{-}\right)$ and $(\tilde{\beta}_{1n}^{-}, \tilde{\beta}_{2n}^{-})$ from the limit distribution of the IV estimators based on its regular instrument counterparts $\left(\tilde{\rho}_{1n}^{+}, \tilde{\rho}_{2n}^{+}\right)$ and $(\tilde{\beta}_{1n}^{+}, \tilde{\beta}_{2n}^{+})$.

Lemma 6. *Consider an oscillating autoregression (1) with $\rho_n < 0$, the OLS estimators $\hat{\rho}_n$ and the IV estimators $\tilde{\rho}_{jn}^{-} = \sum_{t=1}^{n} \underline{x}_t \tilde{z}_{jt-1}^{-} / \sum_{t=1}^{n} \underline{x}_{t-1} \tilde{z}_{jt-1}^{-}$ and $\tilde{\beta}_{jn}^{-} - \sum_{t=1}^{n} \underline{y}_t \tilde{z}_{jt-1}^{-} / \sum_{t=1}^{n} \underline{x}_{t-1} \tilde{z}_{jt-1}^{-}$ in (21) and (22) for $j \in \{1,2\}$. Under C_{-}(i)-C_{-}(iii) of Assumption 1b,*

$$\pi_n \left[\left(\tilde{\rho}_{1n}^{-} - \rho_n \right) \mathbf{1}_{F_n^{-}}, \left(\tilde{\rho}_{2n}^{-} - \rho_n \right) \mathbf{1}_{\bar{F}_n^{-}} \right] = -\pi_n \left[\left(\tilde{\rho}_{1n}^{+} - |\rho_n| \right) \mathbf{1}_{F_n^{++}}, \left(\tilde{\rho}_{2n}^{+} - |\rho_n| \right) \mathbf{1}_{\bar{F}_n^{++}} \right] + o_p(1)$$

$$\pi_n [(\tilde{\beta}_{1n}^{-} - \beta) \mathbf{1}_{F_n^{-}}, (\tilde{\beta}_{2n}^{-} - \beta) \mathbf{1}_{\bar{F}_n^{-}}] = -\pi_n [(\tilde{\beta}_{1n}^{+} - \beta) \mathbf{1}_{F_n^{++}}, (\tilde{\beta}_{2n}^{+} - \beta) \mathbf{1}_{\bar{F}_n^{++}}] + o_p(1) .$$

Next, we employ the limit theory of Lemmata 2-6 to establish the main result of this section: the AMG property of the IV estimators $\tilde{\rho}_n$ in (21) and $\tilde{\beta}_n$ in (22) along all the classes C(i)-C(iii). For $c \in \mathbb{R}$, define the random variables $\Psi_{+}(c) = W_c(1) - \int_0^1 W_c(r)\, dr \mathbf{1}\{\rho = 1\}$

$$\Psi_{-}(c) = \left(\sigma^2 + 2\rho\Gamma \right) / \omega^2 + W_c(1)^2 - 2 W_c(1) \int_0^1 W_c(r)\, dr \mathbf{1}\{\rho = 1\} \tag{44}$$

and $\Psi(c) = \Psi_{-}(c) \mathbf{1}_{F_c} + \Psi_{+}(c) \mathbf{1}_{\bar{F}_c}$, where $W_c(\cdot)$, K_c and the event $F_c = \{K_c + c \leq 0\}$ and its complement $\bar{F}_c$ are defined in (35).

Theorem 3. *Consider the autoregression (1) and the predictive regression model (8) under Assumptions 1b and 3, and the IV estimators $\tilde{\rho}_n$ in (21), $\tilde{\beta}_n$ in (22) and β_n^{*} in (27). Given the normalisation sequence π_n in (43), $\pi_n(\tilde{\rho}_n - \rho_n) \to_d \mathcal{L}_1 =_d \mathcal{MN}(0, V_1)$ under Assumption 2 and $\pi_n(\tilde{\beta}_n - \beta) \to_d \mathcal{L}_2 =_d \mathcal{MN}(0, V_2)$ under Assumption 4, where:*

(i) *Under part C(i) of Assumption 1b, $V_1 = 1$ and $V_2 = \sigma_\varepsilon^2 / (\sigma^2 + 2\rho\Gamma)$.*

(ii) *Under part C(ii) of Assumption 1b, $V_1 = \Psi(c)^{-2}$ and $V_2 = (\sigma_\varepsilon^2/\omega^2)\, \Psi(c)^{-2}$.*

(iii) *Under part C(iii) of Assumption 1b, $\mathcal{L}_1 = Y/X$, $\mathcal{L}_2 = \tilde{Y}/X$, $V_1 = \sigma^2/X^2$, and $V_2 = \sigma_\varepsilon^2/X^2$, where $Y =_d \mathcal{N}(0, \sigma^2)$, $\tilde{Y} =_d \mathcal{N}(0, \sigma_\varepsilon^2)$, and X is independent of $\left(Y, \tilde{Y}\right)$ with $X =_d \mathcal{N}(0, \omega^2)$ when $|\rho_n| \to 1$ and $X = X_\infty$ in (7) when $|\rho_n| \to |\rho| > 1$.*

Moreover, under parts C(i)-C(iii) of Assumption 1b, $\pi_n(\tilde{\beta}_n - \beta_n^{*}) \to_p 0$.

Remarks.

1. The IV procedure proposed in the paper guarantees that the resulting estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ in (21), and (22) respectively exhibit a AMG property along the entire spectrum of autoregressive regressor processes, including stationary, non-stationary, explosive processes, all intermediate regimes and their oscillating counterparts. Importantly, the AMG property is derived via central limit theory and does not depend on the distribution of the innovation sequences (u_t) and (ε_t) in (1) and (8): the only requirements imposed on (u_t) and (ε_t) are Assumption 2 and 4 respectively, which allow the innovations to be non-Gaussian, dependent, non-identically distributed and, as far as inference on β is concerned, u_t may be a linear process under Assumption 4. The only component that depends on the distribution of (u_t) is the mixing variate X_∞ in the explosive case C_0(iii) which does not the affect the AMG property and, upon studentisation of $\tilde{\rho}_n$ and $\tilde{\beta}_n$ is scaled out of the limit distribution of self-normalised test statistics, such as the t-statistic of Theorems 1 and 2. This desirable property of the proposed estimator $\tilde{\rho}_n$ and $\tilde{\beta}_n$ is in sharp contrast to the dependence of large sample OLS inference on the distribution of (u_t) and (ε_t) in explosive autoregressions. Hence, in addition to producing robust inference along all autoregressive classes, our proposed estimation procedure is the first to achieve distribution-free asymptotic inference in explosive autoregressions and is asymptotically invariant to the initialisation X_0 in (1).

2. The key element of the procedure that delivers the AMG property and the distributional invariance to the innovations across the autoregressive classes C(i)-C(iii) is the newly proposed combined instrument $\tilde{z}_t$ in (15)-(17). This instrument employs information from the OLS estimator of the AR parameter (through the events F_n and Lemma 2) to determine whether $c = \lim_{n\to\infty} n\left(|\rho_n| - 1\right)$ takes the value $-\infty$ or ∞. When $c = -\infty$, $\tilde{z}_t$ takes the form of a near-stationary instrument which is: (i) regular $\tilde{z}_{1t}$ when $\rho = 1$; (ii) oscillating $\tilde{z}_{1t}^{-}$ when $\rho = -1$; (iii) either $\tilde{z}_{1t}$ or $\tilde{z}_{1t}^{-}$ when $\rho \in (-1, 1)$, in which case, the IV estimator $\tilde{\rho}_n$ based on both is asymptotically equivalent to the (asymptotically normal) OLS estimator. When $c = \infty$, $\tilde{z}_t$ takes the form of: (i) a mildly-explosive instrument $\tilde{z}_{2t}$ when $\rho \geq 1$ and (ii) an oscillating mildly-explosive instrument $\tilde{z}_{2t}^{-}$ when $\rho \leq -1$. The resulting IV estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ based on $\tilde{z}_{2t}$ and $\tilde{z}_{2t}^{-}$ are shown to achieve distribution-free inference in the near-explosive region C(iii), including the purely explosive subregion C_0(iii). Finally, when $c \in \mathbb{R}$, the autoregression is of the local-to-unity type C(ii) in which case $\tilde{z}_t$ takes the form of: (i) a random linear combination of $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ when $\rho = 1$ and (ii) a random linear combination of $\tilde{z}_{1t}^{-}$ and $\tilde{z}_{2t}^{-}$ when $\rho = -1$. This random combination, reflected in the random normalisation π_n of part (ii) of Theorem 3, depends on the limit distribution of the OLS estimator $\hat{\rho}_n$ through the events F_n^{+} and F_n^{-} in (12) which, like the limit distribution of Ψ_n in (36), can be expressed as a non-stochastic functional of the Brownian motion B; the asymptotic independence of the normalised $\sum_{t=1}^{n} \tilde{z}_{t-1} u_t$ and the Brownian motion B, established by Lemma 5, implies that the additional randomness introduced by the combination of $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ (and $\tilde{z}_{1t}^{-}$ and $\tilde{z}_{2t}^{-}$ when $\rho < 0$) does not affect the AMG property of $\tilde{\rho}_n$ and $\tilde{\beta}_n$. The AMG property across the entire range of autoregressive classes C(i)-C(iii) of Theorem 3 is the key feature of our estimation procedure that delivers the uniform and distribution-free inference based on the t-statistic of Theorems 1 and 2.

3. It is worth providing a brief explanation of how distribution-free asymptotic inference is achieved in the explosive case $|\rho_n| \to |\rho| > 1$. By employing the residuals $\hat{u}_t$ and a (regular or oscillating) mildly explosive root φ_{2n} or φ_{2n}^{-} for the construction of the instrument $\tilde{z}_{2t}$ in (17), the instrumentation of this paper and Lemma 2 ensure that, under C(iii), the limit distribution of $\tilde{\rho}_n$ is driven by the mildly explosive component z_{2t} (or its oscillating counterpart z_{2t}^{-}) in (33) and inherits the desirable AMG property of mildly explosive martingale transforms even when x_t in (1) is a purely explosive process. The price paid for this asymptotic distributional invariance is a reduction in the convergence rate of $\tilde{\rho}_n - \rho_n$ by an order of $\left(\phi_{2n}^2 - 1\right)^{-1/2}$ compared to the ρ_n^n-OLS rate. Given that the above order satisfies $o\left(n^{1/2}\right)$ and that the exponential part ρ_n^n of the OLS rate is maintained in the convergence rate of Theorem 3 part (iii), the efficiency loss associated with employing the IV estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ is small compared to the benefit from an estimation procedure that gives rise to test statistics and CIs of general asymptotic validity. In the case when $|\rho_n| \to 1$ under C(iii), the limit distributions Y/X and $\frac{\omega}{\sigma_\varepsilon}\tilde{Y}/X$ are Cauchy.

4 Monte Carlo Simulations

In this section, we design a Monte Carlo exercise to study the finite sample properties of the IV estimators introduced in this paper and how they compare to alternative approaches. We first discuss the instrument selection and provide a simple guide on how to implement the proposed inference procedure in Section 4.1. We demonstrate that with the above instrument choice, our procedure exhibits good small sample properties for autoregressive regimes covering the entire range from stationarity to (oscillating) explosivity. In Section 4.2 we provide an illustration of the failure of general asymptotic inference based on the OLS estimator in the explosive regions: in particular, we show that misspecifying the variance of a single observation can have severe consequences for the size and coverage rates of OLS-based inference that do not improve with the sample size, both in the autoregressive and predictive regression models. On the other hand, we demonstrate that the IV procedure of Theorems 1 and 2 continues to provide correct inference. Next, we compare the finite sample properties of our procedure to leading existing approaches: in Section 4.3.1, we provide a comparison of our CIs in (24) for the AR parameter to Andrews and Guggenburger (2014)'s procedure; in Section 4.3.2, we compare the size and power of our testing procedure in (28) in the predictive regression setup to the procedure proposed by Elliott et al. (2015). In both cases, we demonstrate that the IV procedure delivers: (i) correct size across all autoregressive regimes considered, and (ii) superior power in all cases of roots in $[-1, 1]$ (including local-to-unity, near- and purely stationary regions) except for the case of exact unit root, where the differences in power are negligible. Crucially, our procedure also provides correct inference on the right side of unity and on the left side of -1, in the local-to-unity, mildly and purely explosive regions, where no existing alternative approach has general asymptotic validity.

4.1 Practical implementation and instrument selection

Practical implementation of our procedure requires a choice for φ_{1n}, φ_{1n}^{-}, φ_{2n} *and* φ_{2n}^{-} in (15) for the instrument construction in (17). While theoretically, any values of $\varphi_{1n} \to 1$ belonging to C_{+}(i), $\varphi_{1n}^{-} \to -1$ belonging to C_{-}(i), $\varphi_{2n} \to 1$ belonging to C_{+}(iii) and $\varphi_{2n}^{-} \to -1$ belonging to C_{-}(iii) deliver correct asymptotic inference, finite sample performance may vary considerably with the choice for particular values. For simplicity and symmetry, we set $\varphi_{1n}^{-} = -\varphi_{1n}$ and $\varphi_{2n}^{-} = -\varphi_{2n}$. Choosing

$$\varphi_{1n} = 1 - 1/n^{b_1}, \ \varphi_{2n} = 1 + 1/n^{b_2} \tag{45}$$

reduces the problem to selecting values for b_1 and b_2 in $(0, 1)$. We adopt a conservative approach: (i) Remark 2 to Theorem 2 indicates that inference based on $T_n^*(\beta_n)$ suffers the worst finite sample distortion in the predictive regression case when $\rho_n = 1$ with large correlation $\rho_{\varepsilon u}$ between the innovations ε_t and u_t in (1) and (8)[7]; (ii) Theorem 3 shows that the power of the t-tests $\tilde{T}_n(\rho_n)$

[7] When $\rho = -1$, such finite sample distortions are not present since the oscillating behaviour of x_t reduces the order of magnitude of $\bar{x}_n$ and $\bar{z}_{1n}^{-}$, and, hence, the distorting effect of the intercept, see Remark 2 to Theorem 2.

and $T_n^*(\beta_n)$ is always increasing with b_1 and is increasing with b_2 in the regions C(i)-C(ii) (in C(iii) the exponential rate ρ_n^n in π_n is independent of b_2, so the choice of b_2 has only a minor effect on power). Given (i) and (ii), we base our selection of b_1 and b_2 on the principle of selecting the maximal values of b_1 and b_2 for which the size in the worst case scenario (i) is controlled. This amounts to a two-dimensional grid search problem outlined below.

We consider a grid of values for b_1 and b_2 in (45) generated by (1)-(8) with a unit root $\rho_n = 1$ and very strong positive and negative correlation $\rho_{\varepsilon u} \in \{0.99, -0.99\}$. Tables B1 and B2 of Appendix B contain the empirical size of the two-sided t-test based on $T_n^*(\beta_n)$ in (28) for testing $\beta = 0$ for $n = 1,000$ based on $10,000$ replications for various combinations of b_1 and b_2. The power plots for the grid points can be found in Figure B1 of Appendix B and are increasing both in b_1 and b_2. Our task is to select the largest values for b_1 and b_2, subject to the size being close to the nominal 5%. Imposing a 5.99% threshold on empirical size for these most unfavourable cases, our grid search procedure yields the following selection in (45): $b_1 = 0.85$ and $b_2 = 0.7$. We recommend this choice for the implementation of Algorithm 1 and use it throughout the Monte Carlo section and the empirical application in Section 5. In Section 4.3. we demonstrate that this choice works well for all autoregressive specifications in both a PR and AR setup.

Implementation of our procedure can be summarised by the following algorithm.

—— **Algorithm 1** ——

1. Given a sample for x_t, compute the OLS estimator $\hat{\rho}_n$ and the OLS-based residuals $\hat{u}_t$.

2. Select φ_{1n}, φ_{1n}^-, φ_{2n} and φ_{2n}^- (e.g. from (45) with the recommended $b_1 = 0.85$ and $b_2 = 0.7$), compute ρ_{nz} in (15) and build recursively the instrument $\tilde{z}_t$ in (17) initialising at $\tilde{z}_0 = 0$.

3. Use the constructed instrument $\tilde{z}_t$ to compute the IV estimator $\tilde{\rho}_n$ in (21) for the AR setup, or, given a sample for y_t, compute the IV estimator β_n^* in (27) for the predictive regression setup.

4. Compute the IV-based t-statistic in (23) and the CI $I_n(\rho_n, \alpha)$ of Theorem 1 for the AR setup, or the IV-based t-statistic in (28) and the CI $I_n(\beta_n^*, \alpha)$ of Theorem 2 for the predictive regression setup; conduct inference using $\mathcal{N}(0,1)$ critical values.

We first implement our choice of instrument in the predictive regression setup (8) along different autoregressive regimes for x_t in (1):

$$\begin{aligned}\rho_n \in\; & \{-1.06,-1.04,-1.02,-(1+10/n^{0.75}),-(1+50/n),-(1+30/n),-(1+15/n),-1,-(1-15/n),-(1-30/n)\\ & -(1-50/n),-(1-10/n^{0.75}),-0.9,-0.7,-0.5,\ 0,\ 0.5,\ 0.7,\ 0.9,\ 1-10/n^{0.75},1-50/n,1-30/n,1-15/n,\ 1,\\ & 1+15/n,1+30/n,\ 1+50/n,\ 1+10/n^{0.75},\ 1.02,\ 1.04,\ 1.06\},\ \ X_0 = 0,\ \mu = \mu_y = 0, \qquad (46)\\ \varepsilon_t \sim\; & \mathcal{N}\left(0,\sigma_\varepsilon^2\right),\ u_t \sim \mathcal{N}\left(0,\sigma^2\right),\ \sigma_\varepsilon = \sigma = 1,\ \rho_{\varepsilon u} \in \{-0.9,-0.45,0,0.45,0.9\}. \qquad (47)\end{aligned}$$

For each specification, we compute the empirical size of the 95% two-sided test statistic in (28) based on 5,000 simulated samples for sample sizes $n \in \{200, 500, 1000\}$. Throughout the entire Monte Carlo section, we always use reduced[8] sample sizes $n \in \{100, 200, 500\}$ for the explosive

[8]We do this for two reasons: (i) it facilitates comparison since the exponential rate of convergence for these specifications implies extremely precise estimates with SEs of the range of 10^{-20} for $n = 500$, and (ii) it prevents Matlab rounding such SEs to 0 (resulting to point CIs) without the need for committing excessive memory.

specifications $\rho_n \in \pm\{1+50/n, 1+10/n^{0.75}, 1.02, 1.04, 1.06\}$. Figures 1-3 display the rejection probability of our test procedure in (28) under the null $\beta = 0$ for the different autoregressive regions with 95% confidence against the two-sided alternative $\beta \neq 0$ for different correlation between the innovations $\rho_{\varepsilon u} \in \{-0.9, 0, 0.9\}$. Figures 1-3 provide evidence that our procedure delivers satisfactory empirical size throughout the different autoregressive specifications converging to the nominal 5% as the sample size increases.

FIGURE 1

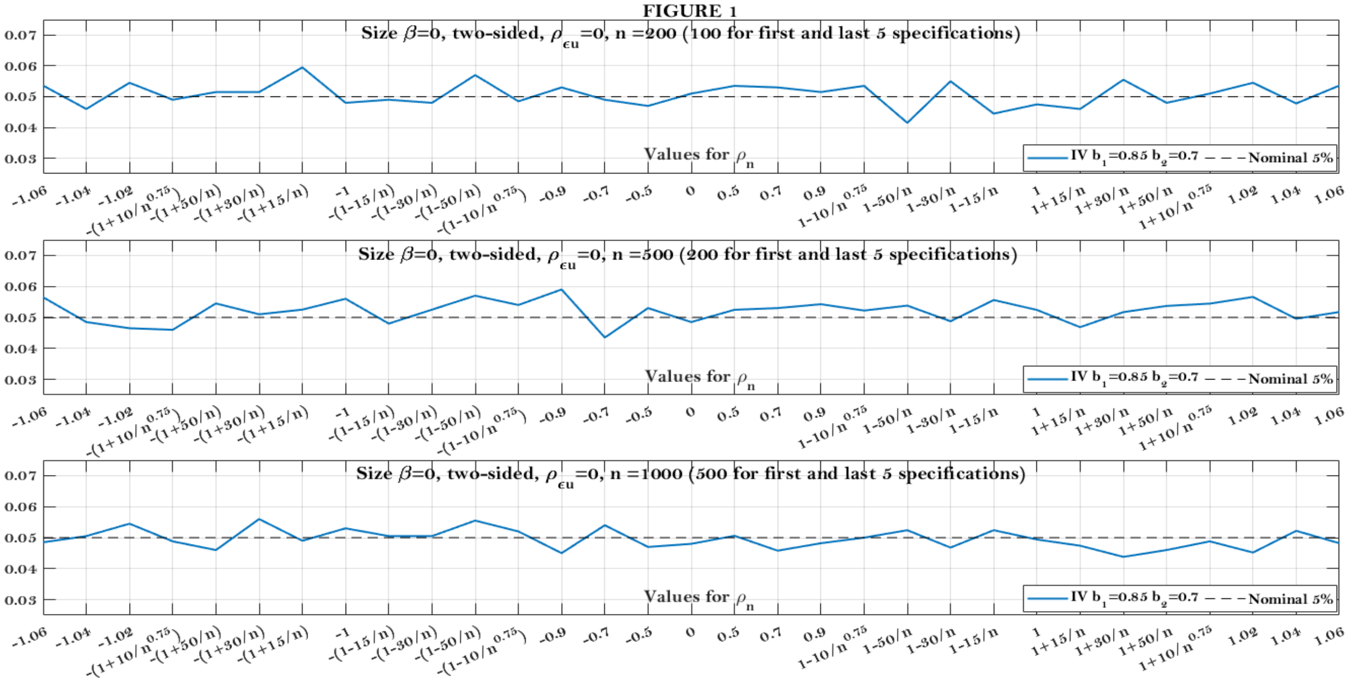
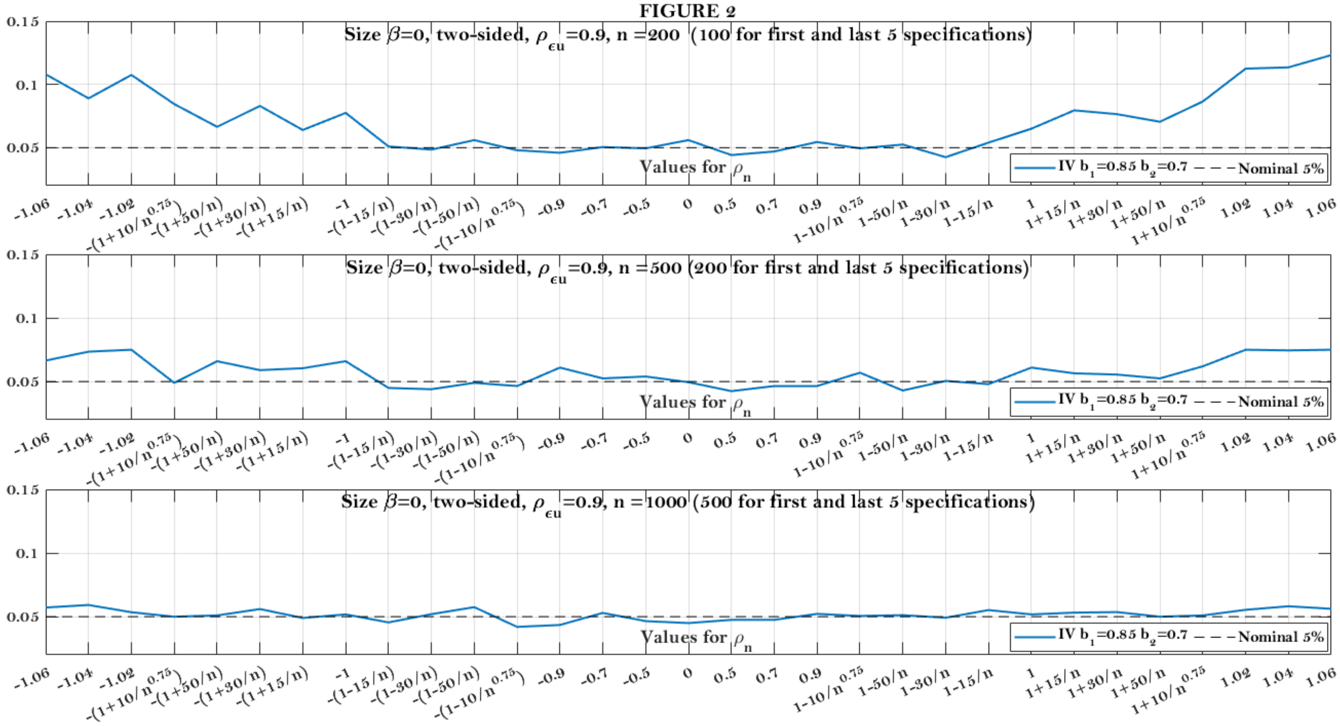
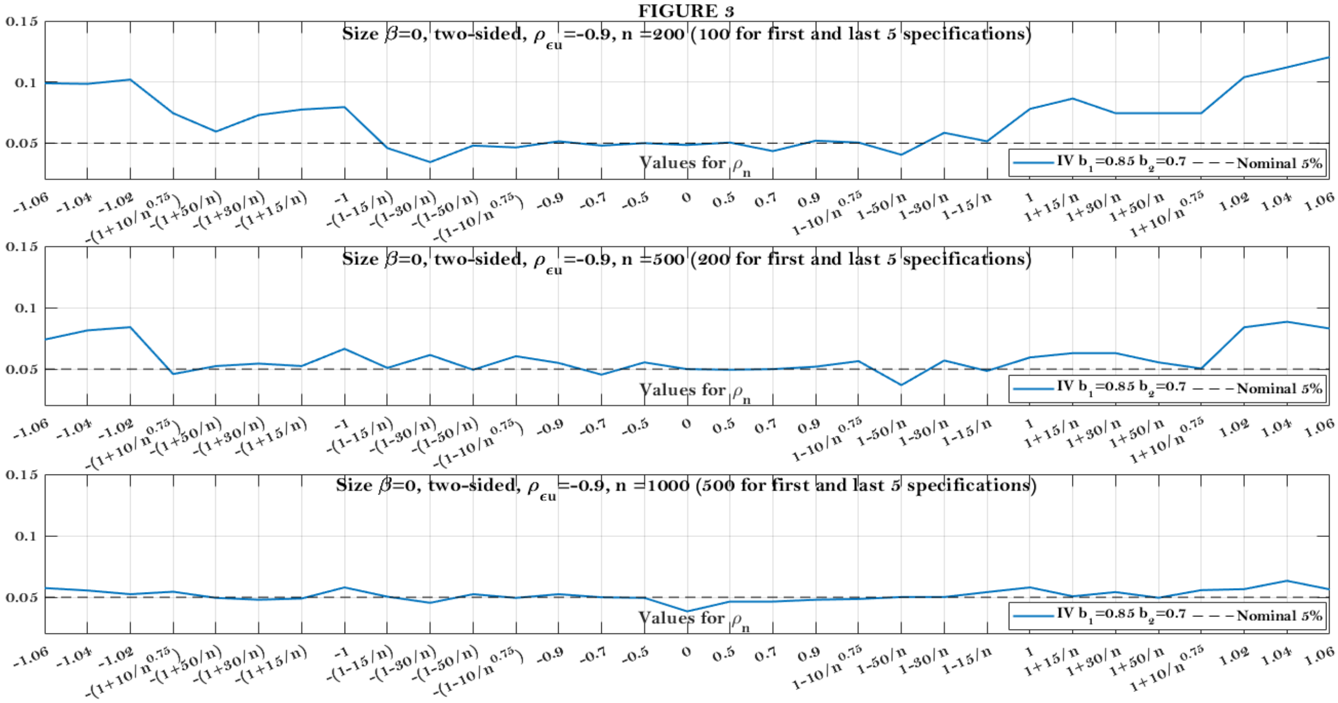


FIGURE 2



Appendix B contains two additional sets of results for moderate negative and positive correlation $\rho_{\varepsilon u} \in \{-0.45, 0.45\}$ as well as the proportion of times each of the instruments is chosen throughout the different autoregressive specifications. As expected, the (oscillating) mildly explosive instrument is never chosen in the stationary region C(i) even for small samples, and is chosen in the (negative) pure unit root case around 33% of the time (since the OLS distribution in this case is left-skewed with values below unity occurring with probability 2/3).

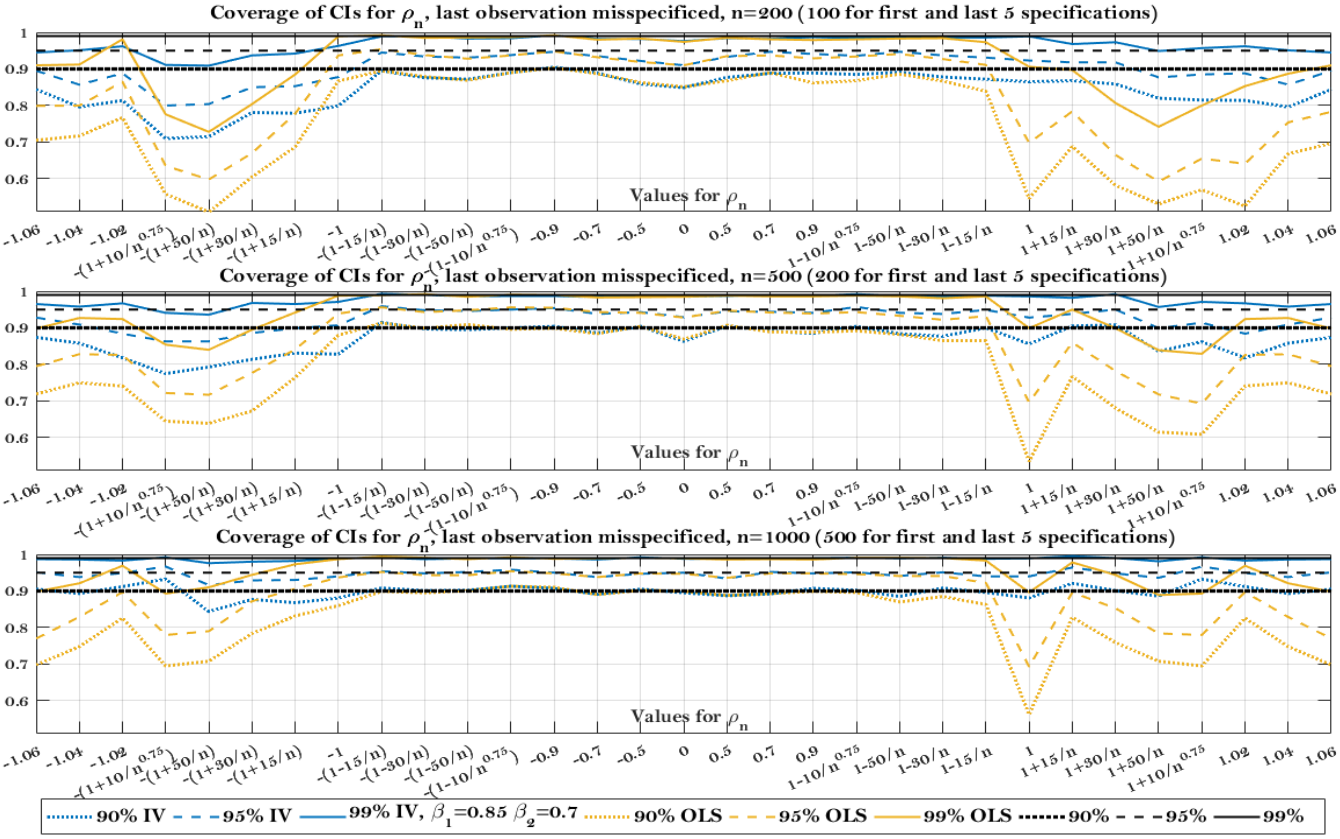
FIGURE 3



4.2 Invalidity of OLS in the explosive regions

In this section, we briefly discuss the relative performance of OLS and our procedure in the explosive regions $(-\infty, 1) \cup (1, \infty)$ and provide an illustration of the invalidity of OLS-based inference even in large samples. The lack of central limit theory for the numerator of the OLS estimators of ρ_n and β implies that the asymptotic distribution of the t-statistic based on the OLS is carried entirely by the last few observations for the innovations, and a change in the distribution of the last innovation in the sample, for example, distorts OLS-based inference even asymptotically. We simulate data from the predictive regression model in (8), with $\varepsilon_t \sim \mathcal{N}(0,1)$, $u_{t-1} \sim \mathcal{N}(0,1)$ for $t = 1, ..., n-1$ and we draw the last observation of the innovations from $\varepsilon_n \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $u_{n-1} \sim \mathcal{N}(0, \sigma^2)$ with $\sigma_\varepsilon = \sigma = 3$ instead. In the presence of CLT (as is the case for our IV estimator), misspecification of any finite number of terms will vanish asymptotically by virtue of uniform asymptotic negligibility (u.a.n.) implied by the CLT. In the absence of u.a.n. and hence a CLT (as is the case with OLS), this type of misspecification may affect the limit and invalidate inference. In Figure 4, we report the 90%, 95% and 99% coverage rates of the IV and OLS estimators of ρ_n respectively for different sample sizes (as in Section 4.1, we work with the autoregressive specifications in (46) and reduced sample sizes for the explosive processes). We compute the coverage rates as the proportion of time that the true ρ_n finds itself in the 90%, 95% and 99% CIs implied by the IV and OLS respectively, based on 5,000 replications.

FIGURE 4

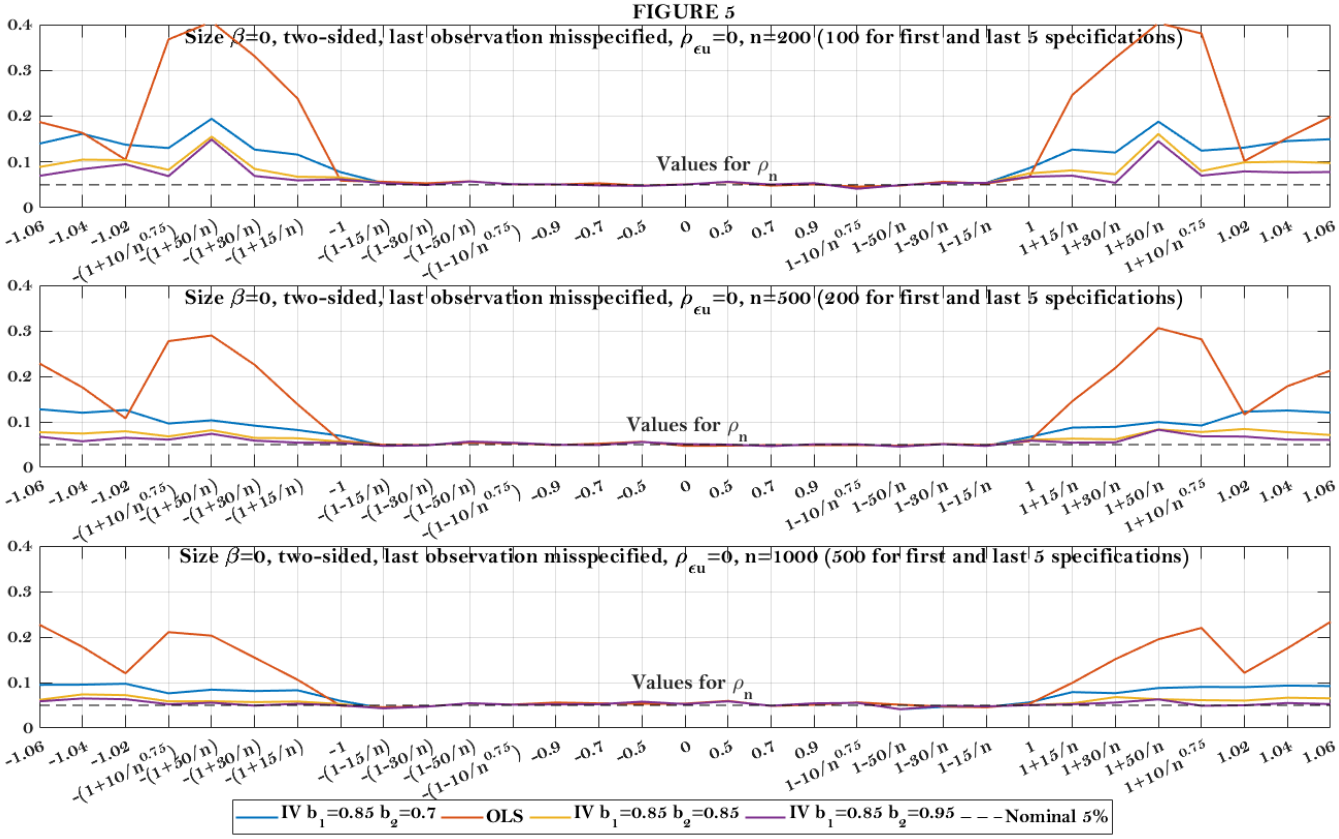


From Figure 4, it is clear that the OLS suffers large finite sample distortions in the local-to-unity region, as well as in the (negative) mildly and purely explosive regions. For sample size $n = 100$, the IV procedure is also affected by this end-of-sample problem and this is expected since our near-explosive instrument exhibits some explosive properties especially when n is small. However, as the sample size increases, the coverage rates of the IV procedure converge to the nominal levels, as Theorem 1 suggests. The coverage rates of OLS for the mildly explosive specification $\rho_n = \pm(1 + 10/n^{0.75})$ also improve as expected (although very slowly). Crucially, for the purely explosive DGPs, the OLS distortions do not improve even for larger samples. For example, when $\rho_n = \pm 1.06$, the 90% OLS CI contains the truth 70% of the time irrespective of increases in the sample size.

We find similar results in the predictive regression setup. In Figure 5, we report the rejection probability of the OLS under the null $\beta = 0$ against a two-sided alternative[9] for the same specifications and sample sizes. We present the rejection probability of the IV procedure for the choice of instrument in Section 4.1 as well as two other choices of instrument, increasing β_2 to 0.85 and 0.95 respectively. As it can be seen from Figure 5, the empirical size of the OLS for the purely explosive regions is distorted and crucially the distortions deteriorate as the sample size increases; the size of our procedure on the other hand converges to the nominal size as the sample size increases, as suggested by the theoretical results of Theorem 2.

[9] The online Appendix B contains additional comparison for the corresponding one-sided rejection probabilities.

FIGURE 5



4.3 Comparisons with alternative methods in the literature

4.3.1 Inference in the autoregressive model

In this section, we present a comparison of our procedure to current state-of-the-art methodology in the literature of robust inference in autoregressions and predictive regressions for $\rho \in (-1, 1]$. We first evaluate our proposed autoregressive CIs in (24) and we compare them to the procedure by Andrews and Guggenberger (2014)[10] (henceforth AG), which constructs the intervals by inverting the OLS t-statistic, which under the null is asymptotically nuisance-parameter-free.

In Figures 6 and 7, we report the coverage rates and lengths of the 90%, 95% and 99% CIs respectively for the IV estimator and AG procedure for ρ_n for different autoregressive regions and for different sample sizes. For the AG procedure, we use the symmetric two-sided intervals imposing homoskedasticity as we found these to perform best in terms of coverage especially in the local-to-unity regions; Appendix B also contains the equal-tailed two-sided intervals of Andrews and Guggenberger (2014). Figure 6 presents evidence that our IV-based CIs are comparable to the CIs based on the AG procedure in $[-1, 1]$, while also providing correct coverage for ρ_n in $(-\infty, -1] \cup [1, \infty)$ in the local-to-unity, mildly and purely explosive regions. In terms of interval length, Figure 7 shows that our intervals are shorter[11] than those of AG (which translates into higher power) for all specifications except for the exact (positive and negative) unit root case $|\rho| = 1$; the differences in length when $|\rho| = 1$ are not large and become negligible for large samples.

[10] The Gauss code for the procedure was kindly provided by Patrik Guggenberger and translated into Matlab.

[11] This result also holds for the equal-tailed two-sided intervals of AG; see Figures B6 and B7 of Appendix B.

FIGURE 6

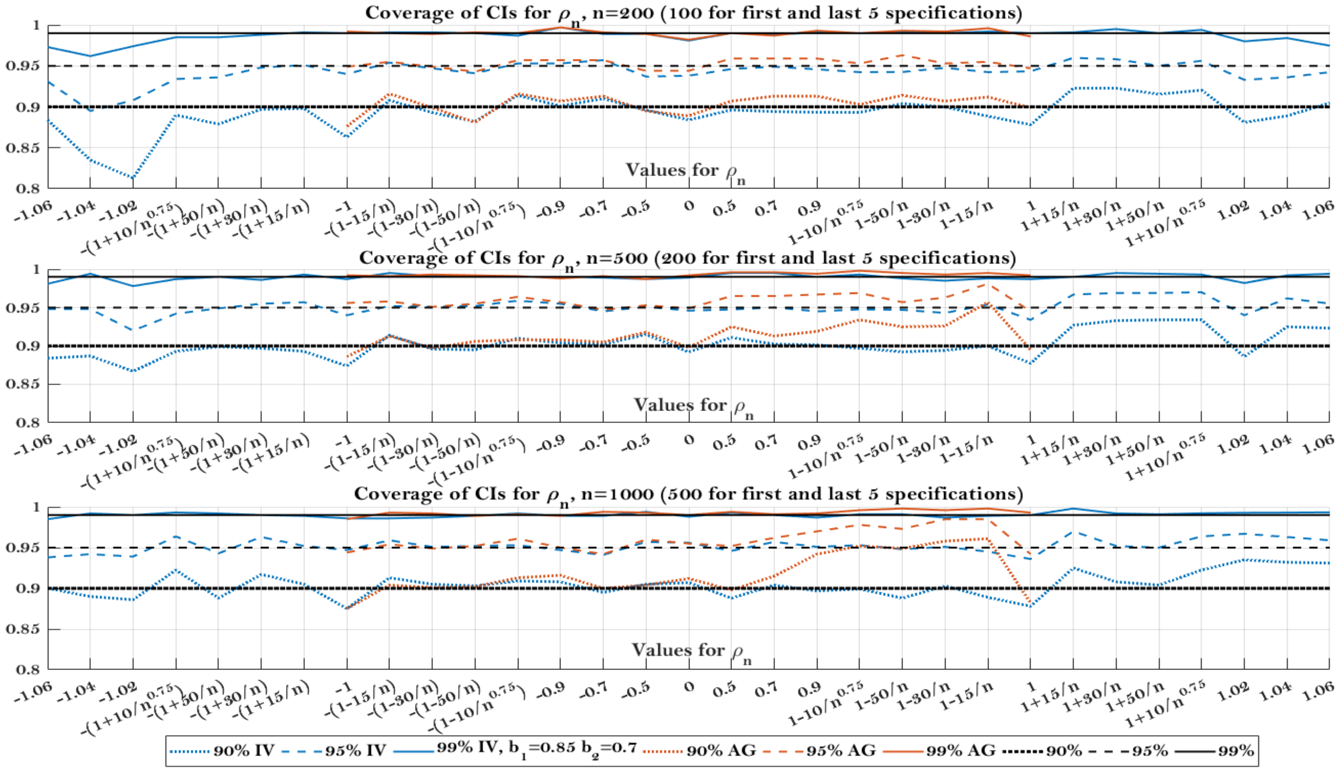
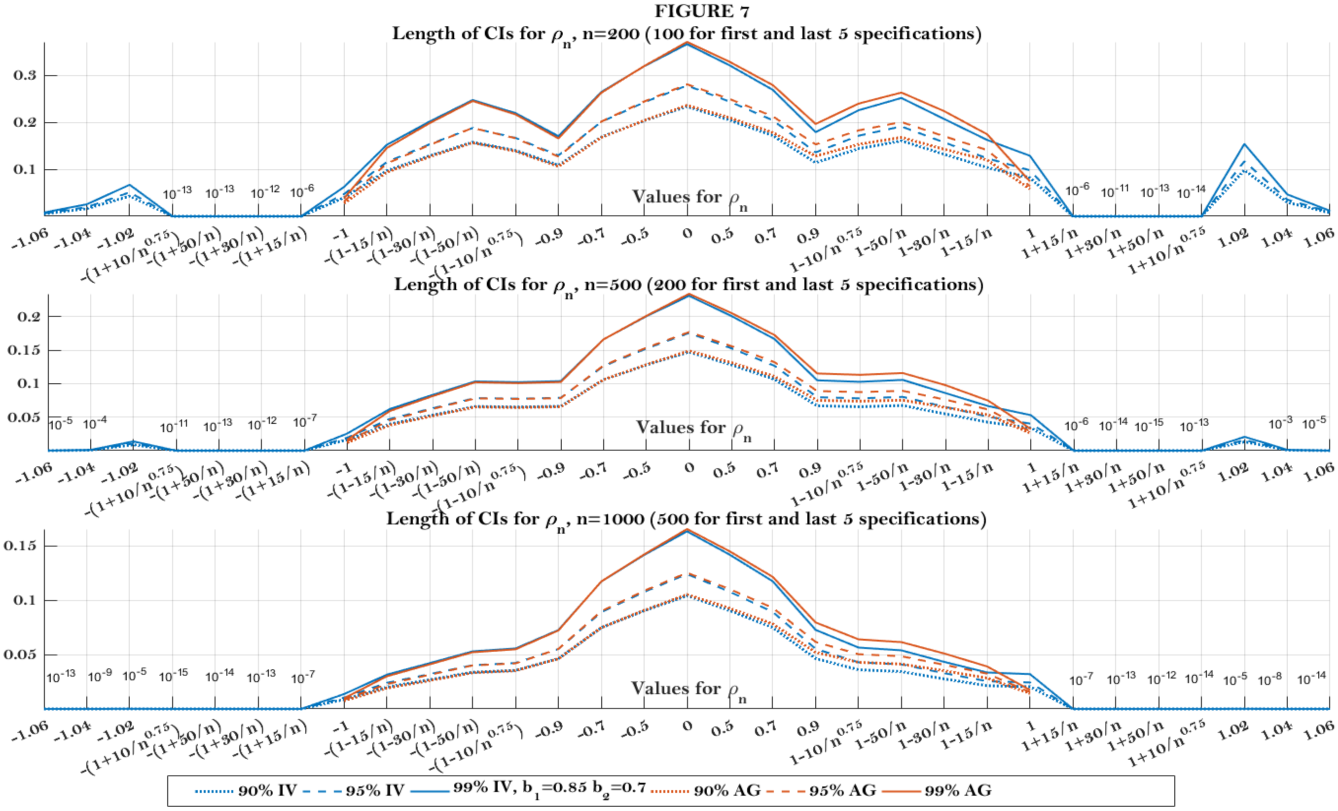


FIGURE 7



4.3.2 Size and power comparison in the predictive regression model

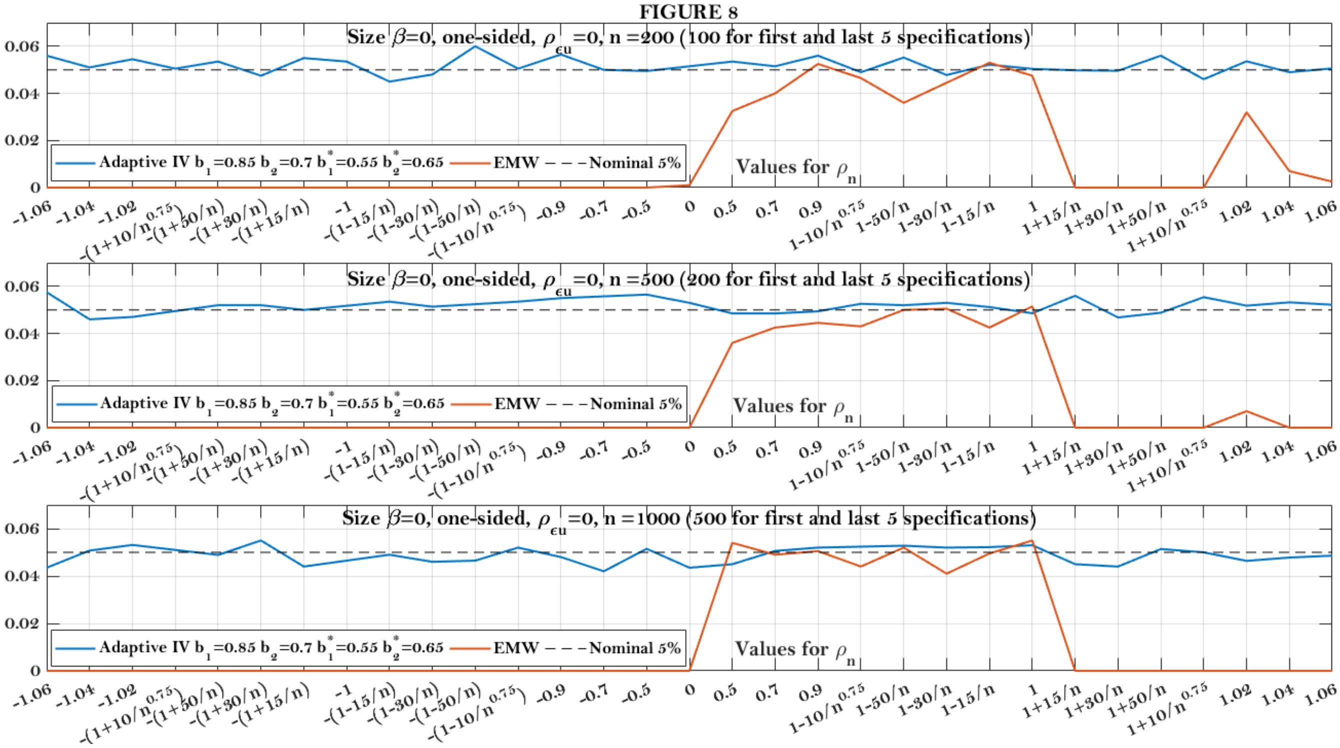
Next, we evaluate the performance of the IV-based t-statistic in (28) in the predictive regression setup (8) and we compare it to the one-sided test procedure by Elliott et al. (2015)[12], which, in the presence of a nuisance parameter, is nearly-optimal when the innovations of the model are Gaussian; Zhou et al. (2019) and Zhou and Werker (2021) provide extensions of this near-efficient testing procedure to non-Gaussian, fat-tailed or heteroskedastic innovations.

We generate data from the predictive model in (8) for the specifications of (46) and (47). We found that in the one-sided test setup, our choice of instrument works well in all but one scenario: the case with strong negative correlation, where our choice for b_1 and b_2 leads to small-sample oversizing in the pure unit root case. Since in all other cases, our choice of instrument from Section 4.1 delivers good size, we prefer not to repeat the selection exercise of Section 4.1, since selecting a more conservative instrument would lead to power loss even in cases where there is no size issue. Instead, we propose using the following adaptive t-statistic:

$$T_n^A(\beta_n) = \mathbf{1}\{\hat{\rho}_{\varepsilon u} \leq L \cap \hat{\rho}_n \geq 0\} T_n(\beta_n(\breve{z}_{1,t})) + \mathbf{1}\{\hat{\rho}_{\varepsilon u} > L \cup \hat{\rho}_n < 0\} T_n(\beta_n(\breve{z}_{2,t})) \tag{48}$$

where $T_n(\beta_n(\breve{z}_{1,t}))$ and $T_n(\beta_n(\breve{z}_{2,t}))$ are the t-statistics in (28) based on two different choices for instruments $\breve{z}_{1,t}$ and $\breve{z}_{2,t}$, $\hat{\rho}_{\varepsilon u}$ is the sample correlation coefficient between the OLS residuals $\hat{u}_t$ and $\hat{\varepsilon}_t$, $\hat{\rho}_n$ is the OLS estimator for ρ_n and L is a threshold level below which a more conservative instrument selection is triggered. In this way, we can resolve the size distortion in the positive unit root case under strong negative correlation, without affecting power in all other cases.

FIGURE 8



We set $L = -0.7$ and use the values $b_1 = 0.55$ and $b_2 = 0.65$ in (45) for the construction of the conservative instrument $\breve{z}_{1,t}$. For $\breve{z}_{2,t}$ we continue to use the choice of instrument from Section 4.1

[12] The Matlab code for the procedure was downloaded from Ulrich Müller's website and some additional procedures were kindly provided by Bo Zhou.

with $b_1 = 0.85$ and $b_2 = 0.7$. In the case of $\rho_{\varepsilon u} = -0.9$ in Figure 10, we display the rejection probability under the null (with 95% confidence against the one-sided alternative $\beta > 0$) of both the original choice of instrument and the new adapted procedure based on (48) to illustrate the effect of using the adaptive procedure.

FIGURE 9

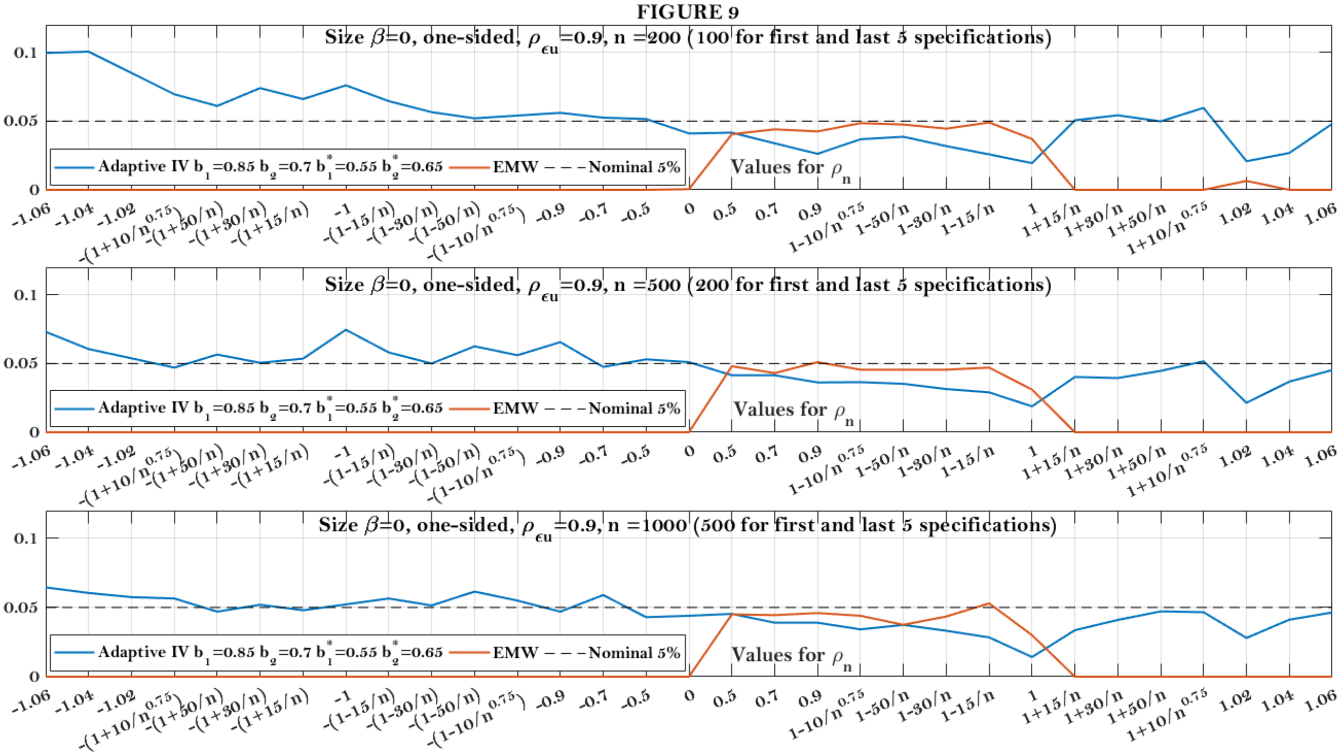


FIGURE 10

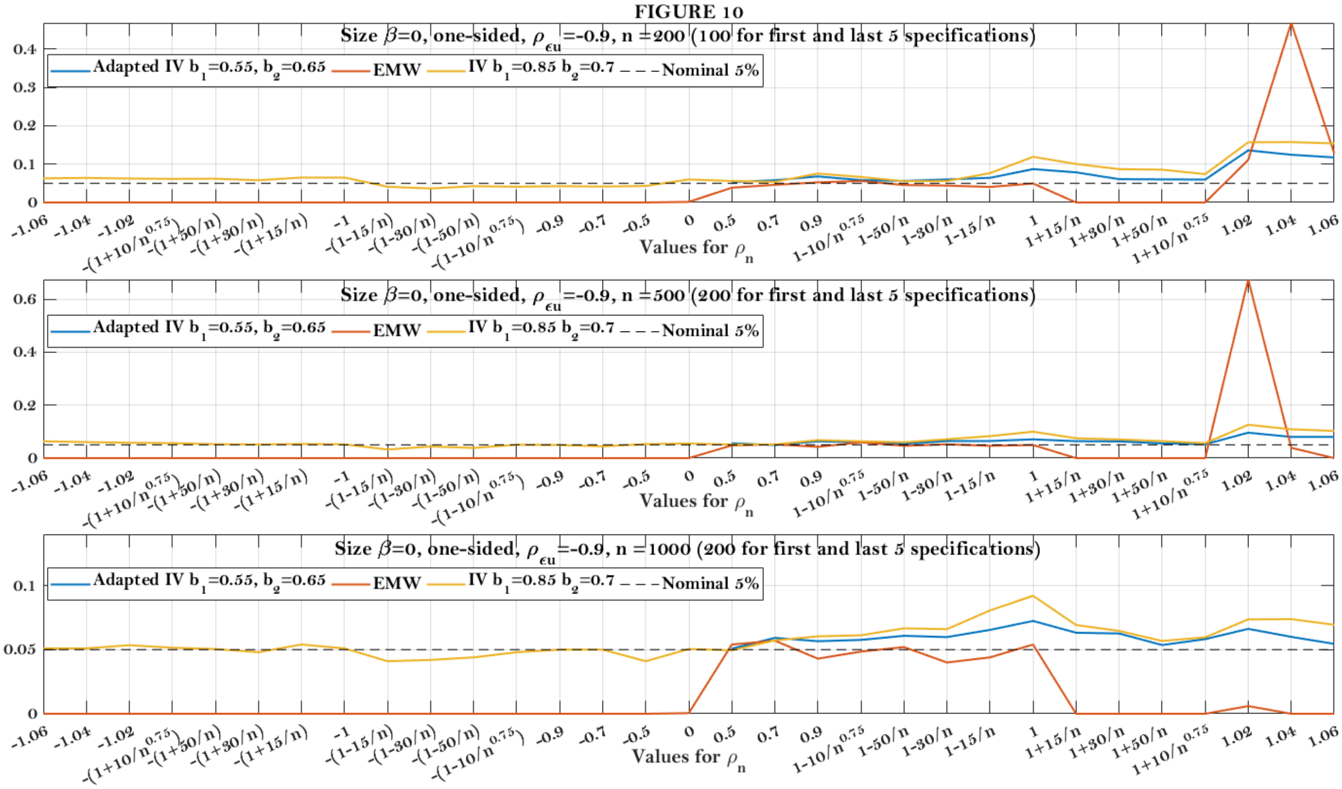


FIGURE 11

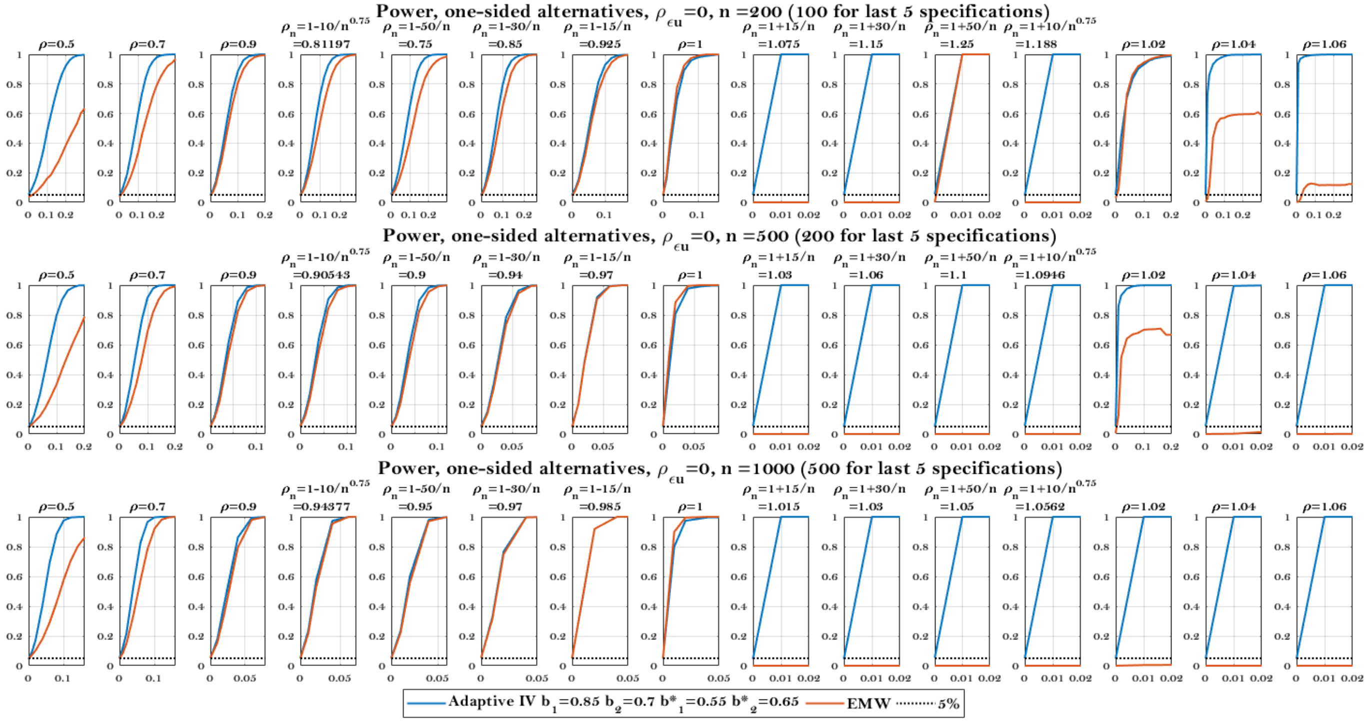
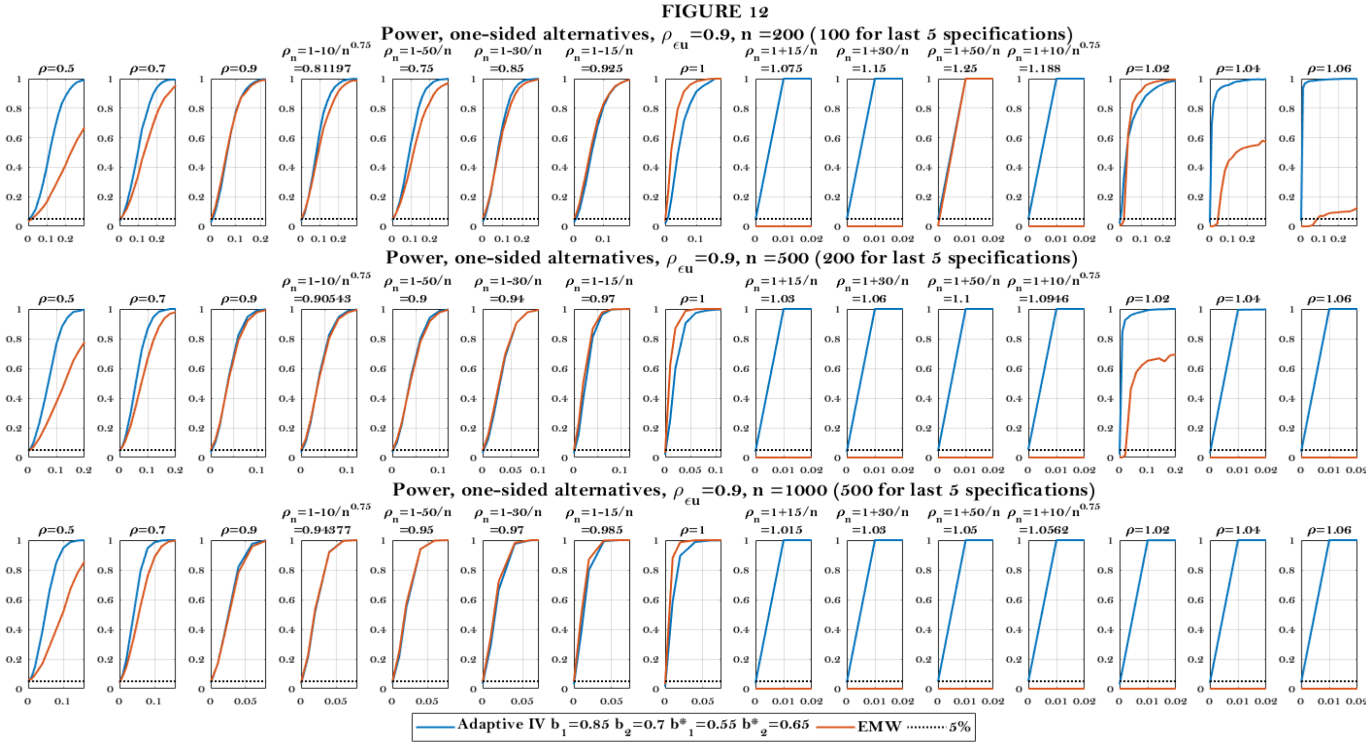


FIGURE 12



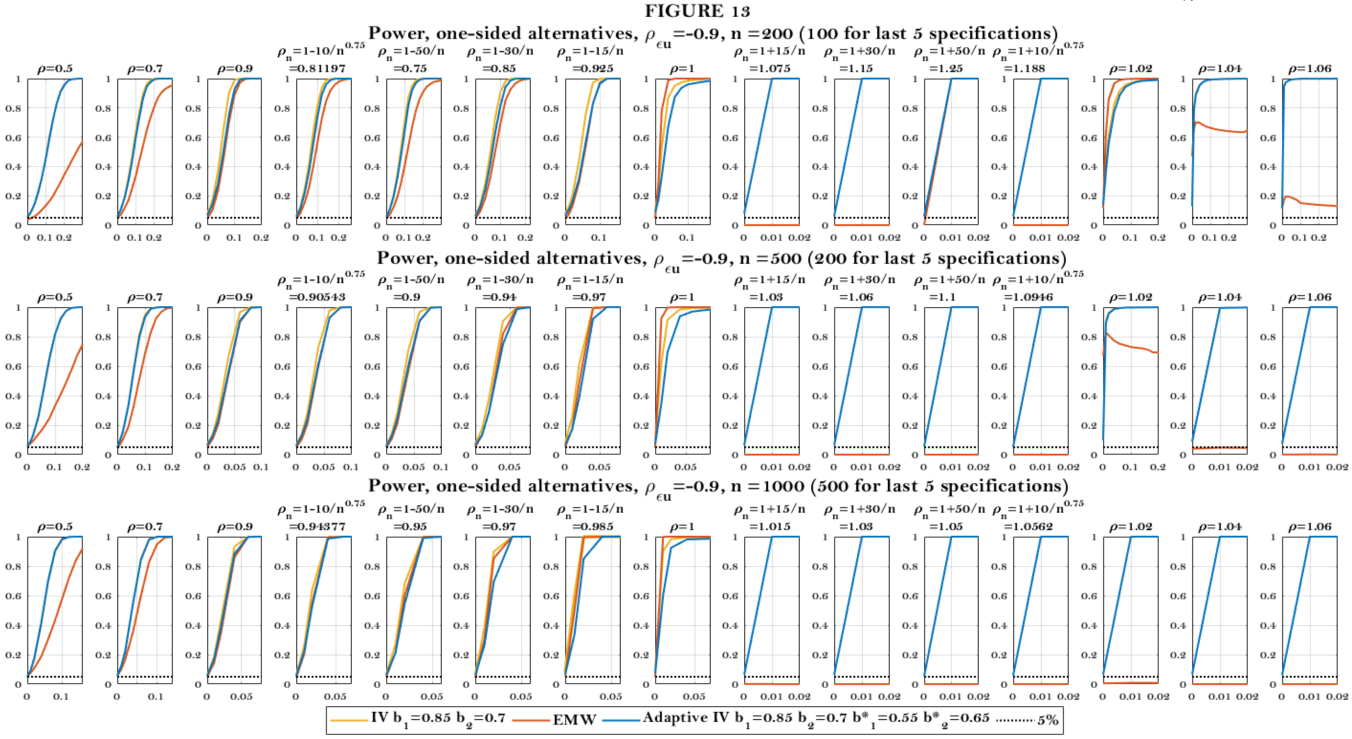
For all other cases, Figures 8-9, we display the rejection probability under the null based on the adaptive instrument which is nearly identical to the original choice of instrument in Section 4.1 since the sample correlation coefficient $\hat{\rho}_{\varepsilon u}$ almost always exceeds the threshold -0.7. Figures 11-13 present the corresponding power curves [13]. We apply the procedure by Elliott et al. (2015) (EMW) in all regions for comparison, stressing that their procedure is not designed to work (and

[13] Appendix B contains additional results for moderate negative and positive correlation $\rho_{\varepsilon u} = \pm 0.45$.

hence it is invalid) outside $(-1, 1]$.

There are several important conclusions from the size and power comparisons in Figures 8-13. First, our adaptive procedure in (48) performs well in terms of empirical size in all correlation cases and in all persistence regions for the regressor and, as the sample size increases, any small sample distortions vanish. Second, we find that the EMW procedure never rejects the null to the right of unity (when the null is true and when it is not), except for a few cases with a small sample; for example in the -0.9 correlation case, its size reaches 40% in the case of $\rho_n = 1.02$ when $n = 100$, but the oversizing disappears as n increases. Surprisingly, we find that the EMW procedure never rejects the null (even under the alternative) for stationary specifications with AR roots in $(-1, 0]$, where it is expected to be valid, since, to our knowledge, it is supposed to switch to OLS. For this reason, in Figures 11-13, we only present power comparison for the cases $\rho_n > 0$, since EMW has zero power for any alternative for all cases $\rho_n \leq 0$. For the regions $\rho_n \leq 0$, our IV procedure has power curves that are a near mirror image of the corresponding non-oscillating cases $\rho_n > 0$.

FIGURE 13



For the cases $\rho_n \in (0, 1]$, we reach a very similar conclusion to the one in the comparison with Andrews and Guggenberger (2014) in Section 4.3.1 in terms of power: namely, our procedure is always more powerful than EMW in all autoregressive specifications (stationary, near-stationary and left-side of local-to-unity regions) except in the case of an exact unit root. The differences in power in the unit root case are small particularly when the correlation in the innovations is moderate. Moreover, in the purely stationary specifications, the power gains of our IV procedure relative to EMW are very large even for large samples. This is another surprising feature of the EMW procedure, since, to our knowledge, it is expected to switch to OLS in these regions, and hence should be asymptotically equivalent to our IV procedure, which is also asymptotically equivalent to OLS in these cases. Crucially, our procedure provides correct inference extending to all cases

$\rho_n \leq 0$, including seasonal (near) unit root and oscillating (mildly) explosive processes as well as to the right of unity (the right-side of local-to-unity, mildly explosive and pure explosive regions) for which no alternative approaches are valid.

5 Inference in a linearised SIR model

In this section, we apply the procedure proposed in the paper to the linearised SIR model (11) on Covid-19 data in order to construct CIs for the parameters θ, γ, δ and for the BRN r_0 across a panel of countries. As discussed in Section 2.3, the triangular system in (11) implies that the dynamics of the number of infections follows an AR(1) process with root $\rho = 1 + \theta - \gamma - \delta$, which in the early stages of the Covid-19 outbreak, before any government intervention, is expected to be greater than unity (since $r_0 > 1$ implies $\theta > (\gamma + \delta)$), and the aim of containment policies was to reduce r_0 below unity. After the Covid-19 outbreak, there has been a lot of interest in epidemic modelling in econometrics, including versions of the SIR model (e.g., Liu, Moon and Schorfheide (2021) perform a fully parametric Bayesian estimation of a piece-wise linear approximation of a nonlinear SIR model, Li and Linton (2021) fit a nonstationary quadratic time trend model on the number of infections). Linearising the model at the DFE reveals the inherently nonstationary dynamics of the series at the outbreak and we stress that: (i) inference based on standard procedures such as OLS/MLE in (11) is only valid when $\rho < 1$ corresponding to the case $r_0 < 1$ which is not empirically relevant at the outbreak (since it implies absence of an epidemic) but may become relevant after government intervention, (ii) when $\rho > 1$, the series for I_t exhibit explosive behaviour with exponential growth and standard semi-parametric procedures such as OLS do not provide valid inference (e.g. CIs), unless i.i.d. Gaussianity assumption is imposed on u_{1t}, and (iii) when ρ is in vicinity of unity (i.e. when the contract rate θ is approximately equal to the removal rate $\gamma + \delta$), OLS/MLE procedures involve nonstandard unit root or local-to-unity asymptotics and so standard inference is invalid. Crucially, not only inference in the equation for I_t but also in the equations for ΔR_t and ΔD_t (which resemble predictive regressions with regressor I_t), and hence inference on γ and δ, is affected by the level of persistence of I_t, and consequently, OLS/MLE inference on γ and δ is only valid in the case $r_0 < 1$. On the other hand, the IV procedure proposed in this paper remains valid for all parameter regions for r_0 and without distributional or homogeneity assumptions of the innovations. Epidemiologists consider r_0 the key parameter for determining whether an epidemic is controllable and for understanding its transmission mechanism and, therefore, being able to construct CIs with correct coverage regardless of the value of $r_0 \in (0, \infty)$ and without distributional assumptions is of practical importance for policy makers.

We use a dataset on daily number of confirmed, recovered and deceased individuals obtained from the John Hopkins University database (https://github.com/CSSEGISandData/COVID-19) for Italy, Germany, Austria, Denmark, Israel and South Korea. The choice of countries is motivated by the availability and quality of series on the number of recovered (e.g. for many countries, recovered series are not reported, of poor quality or not updated at some point). We define the number of active infections as the number of confirmed cases minus the number of recovered cases and deaths at each period. Our sample spans from 22/01/2020 until 04/08/2021[14]. For each country, we start our sample from the date of the first reported death; and we split the remainder of the sample into four subperiods[15] (first reported death: 24/07/2020; 25/07/2020:26/11/2020, 27/11/2020:31/03/2021, 01/04/2020:04/08/2021). Our choice to conduct inference over subsamples is motivated by the unlikelihood that the model's parameters have remained constant over time; aggressive government policies aimed at containing the early epidemic's dynamics aimed at either reducing the number of new infections through imposing lockdowns and social distancing measures (reducing θ), through improved medical response to the outbreak: hospital bed availability, improved treatment (increasing γ, reducing δ), or later on, through vaccination by reducing the proportion of susceptibles S_0/N. We construct the CIs for θ,γ,δ and r_0 for each country and subsample, using the IV CIs in Corollary 1. For the instrument construction, we use (45) with $b_1 = 0.85$ and $b_2 = 0.7$, which we show work well for all AR regions in the Monte Carlo exercise.

FIGURE 14

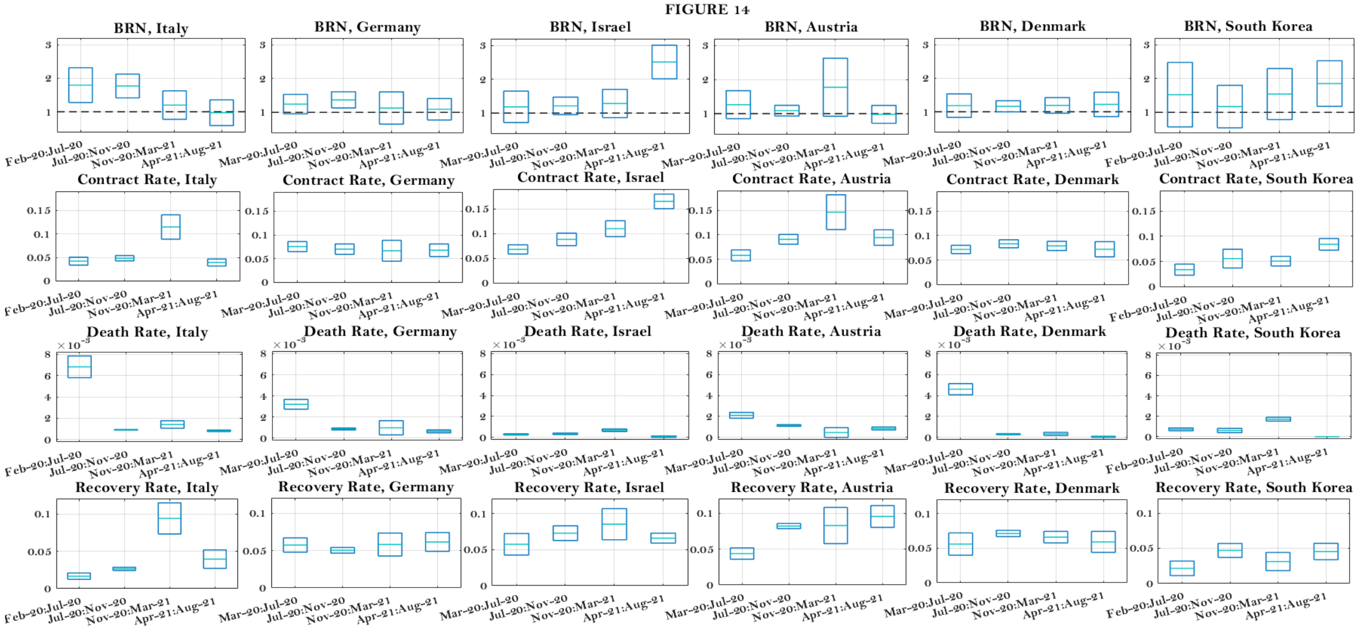


Figure 14 presents the IV estimates and 95% CIs for r_0, θ, δ, and γ for each country and subsample. There are three main conclusions from our empirical analysis. First, the death rate has considerably fallen over time in all countries, and the recovery rate has increased over time for most countries; both due to availability of better medical treatment for the virus (the overall effect of those two conflicting effects on the basic reproduction number r_0 depends on the relative change of $\delta+\gamma$). Second, the contract rate is constant over time for countries like Germany and Denmark, but increasing over time (especially during the winter of 2021) for Italy, Israel and Austria. Third, we find very different values for the basic reproduction number across countries: r_0 is relatively constant over time for countries like Denmark, South Korea, Austria and Germany and while its

[14] R_t series after 08/2021 are unavailable. In late 2021, many re-infections are observed due to mutations, so an SIS model (with probability of re-infection) may be more appropriate for analysis.

[15] To avoid arbitrary sample split, we use the same dates for all countries with roughly the same number of observations in each subsample. Our results are robust to alternative sample splits.

value is usually above unity, one is most of time included in the 95% CI. On the other hand, for Italy, we find that r_0 falls below unity in the period April-August 2021 while for Israel (whose experience has been very different due to an early vaccination programme), r_0 actually surges at the summer of 2021, when cases of re-infection begin to be reported.

While the linearised SIR model in (11) is a very simple and stylised model and the Covid-19 data have been shown to suffer from serious measurement errors and omissions, we make use of the basic SIR model to illustrate the usefulness and empirical relevance of the uniform inference procedure proposed in this paper. Its main advantage is that it gives rise to CIs for the parameters of SIR-type models with correct coverage rates in both highly infectious and remissive periods, a property of crucial empirical relevance as this section demonstrates: r_0 may take values in $(0, 1)$, $(1, \infty)$ as well as values in close vicinity to unity depending on the various stages of the epidemic.

6 Conclusion

The paper proposes a unified, distribution-free framework for inference in both autoregressive and predictive regression models, when the regressor's autoregressive root is in $(-\infty, \infty)$. This includes: (i) stable and near-stable processes, (ii) (seasonal) unit root and local-to-unity regressors, and (iii) regressors that exhibit stochastic exponential growth (e.g. explosive and mildly explosive).

The unified inference is based on a novel estimation method that employs an instrumental variable approach with an artificially constructed instrument with a data-driven combination of a (possibly oscillating) near-stationary and near-explosive root. The resulting IV estimators for the AR parameter in the autoregression and the slope parameter in the predictive regression framework are both shown to have a mixed-Gaussian limit distribution under all persistence regimes, and independently of the distribution of the innovations and the initial condition. Consequently, the t-statistic based on the new estimators is asymptotically standard normal with uniform size over arbitrary closed subintervals of $(-\infty, \infty)$ and gives rise to asymptotically correctly-sized CIs. To our knowledge, this is the first method that delivers central limit theory and, consequently, general distribution-free asymptotic inference with a regressor with autoregressive root in $(-\infty, -1) \cup (1, \infty)$ and achieves uniform asymptotic inference over the entire autoregressive range $(-\infty, \infty)$.

We demonstrate that our inference procedure exhibits very good finite sample properties in an extensive Monte Carlo study and compares favourably to existing procedures for inference in both autoregressions (Andrews and Guggenberger (2014)) and predictive regressions (Elliott et al. (2015)) in their parametric validity range $(-1, 1]$ while providing correct inference on $(-\infty, -1] \cup (1, \infty)$, where no existing alternative approach has general asymptotic validity.

Finally, we show that the basic SIR model for modelling epidemics' dynamics, upon linearisation around DFE, reveals that the number of active infections evolves as a first order autoregressive process with an explosive root whenever the basic reproduction number is above unity. We employ our procedure to model early dynamics of the Covid-19 epidemic across countries and construct CIs for the model's parameters without restricting the parameter space, i.e. without *a priori* knowledge of whether the epidemic is in a controllable or uncontrollable stage.

Appendix A

This Appendix contains the mathematical proofs of Theorems 1-3 of the paper. Some auxiliary results, as well as the proofs of Lemmata 1-6 and Corollary 1, can be found in Appendix B.

Proof of Theorem 3. Under C_+(i)-C_+(ii) of Assumption 1b,

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left(\tilde{\rho}_{1n}-\rho_n\right)=\frac{n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{1/2}\left(\sum_{t=1}^n \tilde{z}_{1t-1}u_t-n\bar{z}_{1n-1}\bar{u}_n\right)}{n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{1t-1}}$$

with Lemma 3(i) and $\bar{u}_n=O_p\left(n^{-1/2}\right)$ implying that

$$n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{1/2}n\bar{z}_{1n-1}\bar{u}_n=O_p\left(n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\right)+O_p\left(n^{-1}\left(1-\varphi_{1n}^2\right)^{-1}\right)=o_p\left(1\right)$$

and, similarly for $\tilde{\beta}_{1n}$, $n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{1/2}n\bar{z}_{1n-1}\bar{\varepsilon}_n=o_p\left(1\right)$. By Lemma 3(ii), the common denominator of $\pi_n\left(\tilde{\rho}_{1n}-\rho_n\right)$ and $\pi_n\left(\tilde{\beta}_{1n}-\beta\right)$ is asymptotically equivalent to $\tilde{\Psi}_n$ in (36) we obtain, under C_+(i)-C_+(ii),

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left[\tilde{\rho}_{1n}-\rho_n,\tilde{\beta}_{1n}-\beta\right]=\left[1+o_p\left(1\right)\right]\tilde{\Psi}_n^{-1}\left[\tilde{U}_n\left(1\right),\tilde{U}_n^{\varepsilon}\left(1\right)\right] \tag{A.1}$$

where $\tilde{U}_n\left(\cdot\right)$ is defined as $U_n\left(\cdot\right)$ in Lemma 5 with z_{1t-1} replaced by $\tilde{z}_{1t-1}$ (and $e_t=u_t$ under Assumption 2) and $\tilde{U}_n^{\varepsilon}\left(\cdot\right)$ as $\tilde{U}_n\left(\cdot\right)$ with e_t replaced by ε_t.

Under C_+(i) and Assumption 2, $u_t=e_t$ and $\Gamma=0$ so $\tilde{\Psi}\left(c\right)=\sigma^2$ and $\tilde{U}_n\left(1\right)\to_d\mathcal{N}\left(0,\sigma^4\right)$ by Lemma 3(iii), so substituting into (A.1) yields

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left(\tilde{\rho}_{1n}-\rho_n\right)\to_d\mathcal{N}\left(0,1\right). \tag{A.2}$$

For $\tilde{\beta}_n$ under C_+(i) and Assumption 4, $\tilde{\Psi}\left(c\right)=\sigma^2+2\rho\Gamma$ by Lemma 3(i) and $\tilde{U}_n^{\varepsilon}\left(1\right)\to_d\mathcal{N}\left(0,v_{\varepsilon}^2\right)$ with $v_{\varepsilon}^2=\left(\sigma^2+2\rho\Gamma\right)\sigma_{\varepsilon}^2$ by Lemma 3(iii) with the martingale difference e_t replaced by ε_t, giving

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left(\tilde{\beta}_{1n}-\beta\right)\to_d\mathcal{N}\left(0,\sigma_{\varepsilon}^2/\left(\sigma^2+2\rho\Gamma\right)\right). \tag{A.3}$$

Under C_-(i) and Assumption 4, $\tilde{\Psi}_n$ in (36) satisfies $\tilde{\Psi}_n=\tilde{\Psi}_n^+ +o_p\left(1\right)$, where

$$\tilde{\Psi}_n^+=\frac{1}{n}\left(1-\rho_n^2\left|\varphi_{1n}^-\right|^2\right)\sum_{t=1}^n x_{t-1}^+\tilde{z}_{0t-1}^+=\sigma^2+2\left|\rho_n\right|\Gamma_n^+ +o_p\left(1\right)$$

by Lemma 3(i) and (B.12), where, by Lemma 2.2(i) of MP(2020),

$$\Gamma_n^+=n^{-1}\sum_{t=1}^n x_{t-1}^+\left(-1\right)^t u_t\to_p\sum_{k=1}^{\infty}\left|\rho\right|^{k-1}\left(-1\right)^k\gamma_u\left(k\right)=-\Gamma \tag{A.4}$$

since $\left|\rho\right|=-\rho$; we conclude that $\tilde{\Psi}_n^+\to_p\sigma^2-2\left|\rho\right|\Gamma=\sigma^2+2\rho\Gamma$. We now prove part (i) of the theorem for $\tilde{\rho}_n$ under C(i): by Lemma 2(i),

$$\pi_n\left(\tilde{\rho}_n-\rho_n\right)=\pi_n\left(\tilde{\rho}_{1n}-\rho_n\right)\mathbf{1}_{F_n^+}+\pi_n\left(\tilde{\rho}_{1n}^- -\rho_n\right)\mathbf{1}_{F_n^-}+o_p\left(1\right) \tag{A.5}$$

under Assumption 4. When $\rho_n\to1$, Lemma 2(iii) implies that $\pi_n\left(\tilde{\rho}_n-\rho_n\right)=\pi_n\left(\tilde{\rho}_{1n}-\rho_n\right)+o_p\left(1\right)\to_d\mathcal{N}\left(0,1\right)$ by (A.2). When $\rho_n\to-1$, Lemma 2(iv) and Lemma 6 imply that

$$\pi_n\left(\tilde{\rho}_n-\rho_n\right)=\pi_n\left(\tilde{\rho}_{1n}^- -\rho_n\right)\mathbf{1}_{F_n^-}+o_p\left(1\right)=-\pi_n\left(\tilde{\rho}_{1n}^+ -\left|\rho_n\right|\right)+o_p\left(1\right)$$

since $\hat{\rho}_n^+\to_p\left|\rho_n\right|=1$ implies that $\mathbf{1}_{F_n^{++}}\to_p1$. From its definition in Lemma 6, $\tilde{\rho}_{1n}^+$ is an IV estimator generated by the (regular) C_+(i) autoregression (41) and the regular near-stationary instrument $\tilde{z}_{0t}^{+}$ in in (42); hence, (A.2) implies that $\pi_n\left(\tilde{\rho}_{1n}^{+}-\rho_n\right) \to_d \mathcal{N}(0,1)$, showing that $\pi_n(\tilde{\rho}_n-\rho_n) \to_d \mathcal{N}(0,1)$ when $\rho_n \to -1$. To complete the proof of part (i) for $\tilde{\rho}_n$, it remains to deal with the stationary case $\rho_n \to \rho \in (-1,1)$, where under Assumption 2, $\pi_n \sim n^{1/2}\left(1-\rho_n^2\right)^{-1/2}$, $\mathbf{1}_{F_n^{+}} \to_p \mathbf{1}_{\{\rho\geq 0\}}$ and $\mathbf{1}_{F_n^{-}} = \mathbf{1}_{F_n^{++}} \to_p \mathbf{1}_{\{\rho<0\}}$ by Lemma 2 and the consistency of $\hat{\rho}_n$; (A.5) and Lemma 6 then yield

$$\pi_n(\tilde{\rho}_n-\rho_n) = \pi_n(\tilde{\rho}_{1n}-\rho_n)\mathbf{1}_{\{\rho\geq 0\}} - \pi_n\left(\tilde{\rho}_{1n}^{+}-|\rho_n|\right)\mathbf{1}_{\{\rho<0\}} + o_p(1) \to_d \mathcal{N}(0,1)$$

since $\pi_n(\tilde{\rho}_{1n}-\rho_n) \to_d \mathcal{N}(0,1)$ and $\pi_n\left(\tilde{\rho}_{1n}^{+}-|\rho_n|\right) \to_d \mathcal{N}(0,1)$ as established above for the regular C_{+}(i) autoregression case. This completes the proof of part (i) for $\tilde{\rho}_n$. The proof of part (i) for $\tilde{\beta}_n$ follows a similar argument: since $\tilde{\Psi}_n^{+} \to_p \sigma^2+2\rho\Gamma$, (A.3) holds with $\left(\varphi_{1n},\tilde{\beta}_{1n}\right)$ replaced by $\left(\left|\varphi_{1n}^{-}\right|,\tilde{\beta}_{1n}^{+}\right)$ and the same argument then applies by replacing $\left(\tilde{\rho}_n,\tilde{\rho}_{1n},\tilde{\rho}_{1n}^{-},\tilde{\rho}_{1n}^{+}\right)$ by $\left(\tilde{\beta}_n,\tilde{\beta}_{1n},\tilde{\beta}_{1n}^{-},\tilde{\beta}_{1n}^{+}\right)$. When $\rho_n \to \rho \in (-1,0)$ Lemma B1(iv) implies that $\hat{\rho}_n-\rho = -\left(\hat{\rho}_n^{+}-|\rho_n|\right) + o_p(1) = \Gamma/\left(\sigma^2+2\rho\Gamma\right)$, i.e. $\hat{\rho}_n \to_p \rho_\Gamma := \rho+\Gamma/\left(\sigma^2+2\rho\Gamma\right)$ under Assumption 4, which implies that $\mathbf{1}_{F_n^{+}} \to_p \mathbf{1}_{\{\rho_\Gamma\geq 0\}}$ and $\mathbf{1}_{F_n^{-}} = \mathbf{1}_{F_n^{++}} \to_p \mathbf{1}_{\{\rho_\Gamma<0\}}$. Hence,

$$\pi_n\left(\tilde{\beta}_n-\beta\right) = \pi_n\left(\tilde{\beta}_{1n}-\beta\right)\mathbf{1}_{\{\rho_\Gamma\geq 0\}} - \pi_n\left(\tilde{\beta}_{1n}^{+}-\beta\right)\mathbf{1}_{\{\rho_\Gamma<0\}} + o_p(1) \to_d \mathcal{N}\left(0,\sigma_\varepsilon^2/\left(\sigma^2+2\rho\Gamma\right)\right)$$

since both $\pi_n\left(\tilde{\beta}_{1n}-\beta\right)$ and $\pi_n\left(\tilde{\beta}_{1n}^{+}-\beta\right)$ have the same limit distribution given in (A.3). This completes the proof of part (i) of the theorem.

We proceed with the proof under C_{+}(ii)-C_{+}(iii). $R_{1n} = o_p(1)$ in Lemma B2(iv) implies that $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left|n\bar{z}_{2n-1}\bar{u}_n - \bar{u}_n\sum_{t=1}^n z_{2t-1}\right| = o_p(1)$. Since $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}n^{-1/2}\sum_{t=1}^n z_{2t-1} = O_p(n^{-1/2}\left(\varphi_{2n}^2-1\right)^{-1/2})$, we conclude that $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left(n\bar{z}_{2n-1}\bar{u}_n\right) = o_p(1)$. By a similar argument for $\tilde{\beta}_{2n}$: $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left(n\bar{z}_{2n-1}\bar{\varepsilon}_n\right) = o_p(1)$. The above and Lemma 4(i) imply that the numerators of $\pi_n(\tilde{\rho}_{2n}-\rho_n)$ and $\pi_n(\tilde{\beta}_{2n}-\beta)$ are asymptotically equivalent to

$$\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left[\sum_{t=1}^n \tilde{z}_{2t-1}u_t, \sum_{t=1}^n \tilde{z}_{2t-1}\varepsilon_t\right] = \left[Y_nZ_n, Y_n^{\varepsilon}Z_n\right] + o_p(1). \tag{A.6}$$

The approximation for R_{1n} in Lemma B2(iv) and Lemma B1(i) give

$$n\left(\rho_n\varphi_{2n}-1\right)\nu_n^{-1}\nu_{n,z}^{-1}\bar{x}_{n-1}\bar{z}_{2n-1} = \left[1+o_p(1)\right]\frac{\kappa_n}{n}\nu_{n,z}^{-1}\left(\rho_n\varphi_{2n}-1\right)\sum_{t=1}^n z_{2t-1}\frac{\nu_n^{-1}}{\kappa_n}\sum_{j=1}^n x_{j-1} \tag{A.7}$$

which is $o_p(1)$ under C_{+}(iii): $O_p(\kappa_n/n)$ if $\left(\rho_n-1\right)/\left(\varphi_{2n}-1\right) \to 0$ and $O_p((\varphi_{2n}-1)^{-1}/n)$ if $\left(\varphi_{2n}-1\right)/\left(\rho_n-1\right) = O(1)$. Under C_{+}(ii), (A.7) becomes $Z_nn^{-3/2}\sum_{j=1}^n x_{j-1} + o_p(1)$ by (B.41), showing that (A.7) contributes asymptotically under C_{+}(ii). Combining the above with the approximation of $s_n^{-1}\sum_{t=1}^n x_{t-1}z_{2t-1}$ in Lemma 4(i), we obtain that the common denominator of $\pi_n(\tilde{\rho}_{2n}-\rho_n)$ and $\pi_n\left(\tilde{\beta}_{2n}-\beta\right)$ satisfies

$$s_n^{-1}\sum_{t=1}^n \underline{x}_{t-1}z_{2t-1} = Z_n\underline{X}_n + o_p(1), \quad \underline{X}_n := X_n - n^{-3/2}\sum_{j=1}^n x_{j-1} \tag{A.8}$$

under C_{+}(ii)-C_{+}(iii) and Assumption 4, where Z_n and X_n are defined in (39) and (40). Recalling the definition of s_n in (38) and noting that $\rho_n\varphi_{2n}-1 \sim \varphi_{2n}-1$ under C_{+}(ii), the normalisation under C$_+$(ii)-C$_+$(iii) becomes $s_n/((\varphi_{2n}^2-1)^{-1}\varphi_{2n}^n)=\pi_n$ in (43). Combining (A.8) and (A.6),

$$\frac{(\varphi_{2n}^2-1)^{1/2}\nu_n}{\rho_n\varphi_{2n}-1}\left[\tilde{\rho}_{2n}-\rho_n,\tilde{\beta}_{2n}-\beta\right]=\frac{1}{\underline{X}_n}\left[Y_n,Y_n^{\varepsilon}\right]+o_p(1) \tag{A.9}$$

under C$_+$(ii)-C$_+$(iii) and Assumption 4. We now prove part (iii) of Theorem 3: under C$_+$(iii), $\underline{X}_n=X_n+o_p(1)$ and applying parts (ii) and (iii) of Lemma 4 and the continuous mapping theorem to (A.9) we obtain

$$\pi_n(\tilde{\rho}_{2n}-\rho_n)\to_d Y/X \text{ and } \pi_n\left(\tilde{\beta}_{2n}-\beta\right)\to_d Y^{\varepsilon}/X \tag{A.10}$$

where $X=_d\mathcal{N}(0,\omega^2)$ when $\rho_n\to 1$ and $X=X_\infty$ when $\rho_n\to\rho>1$, so that $X\neq 0$ *a.s.* under C$_+$(iii), X is independent of (Y,Y^{ε}) and $Y=_d\mathcal{N}(0,\sigma^2)$, $Y^{\varepsilon}=_d\mathcal{N}(0,\sigma_{\varepsilon}^2)$ by Lemma 4. Under Assumption 2, $\omega^2=\sigma^2$, so $Y/X=_d\mathcal{MN}(0,\sigma^2/X^2)$; under Assumption 4, $Y^{\varepsilon}/X=_d\mathcal{MN}(0,\sigma_{\varepsilon}^2/X^2)$. Thus, (A.10) gives the correct limit distributions for C$_+$(iii), the theorem under C$_+$(iii) follows from the asymptotic equivalences $\pi_n(\tilde{\rho}_n-\tilde{\rho}_{2n})=o_p(1)$ and $\pi_n\left(\tilde{\beta}_n-\tilde{\beta}_{2n}\right)=o_p(1)$ by applying parts (ii) and (iii) of Lemma 2 to (21) and (22). Under C$_-$(iii), parts (ii) and (iv) of Lemma 2 imply that

$$\begin{aligned}\pi_n\left[(\tilde{\rho}_n-\rho_n),\left(\tilde{\beta}_n-\beta\right)\right] &= \pi_n\left[\left(\tilde{\rho}_{2n}^{-}-\rho_n\right),\left(\tilde{\beta}_{2n}^{-}-\beta\right)\right]\mathbf{1}_{\bar{F}_n^{-}}+o_p(1)\\ &= -\pi_n\left[\left(\tilde{\rho}_{2n}^{+}-|\rho_n|\right),\left(\tilde{\beta}_{2n}^{+}-\beta\right)\right]\mathbf{1}_{\bar{F}_n^{++}}+o_p(1)\end{aligned}$$

by Lemma 6. Since $\mathbf{1}_{\bar{F}_n^{++}}\to_p 1$ by Lemma 2, and $\tilde{\rho}_{2n}^{+}$ is generated by the regular C$_+$(iii) autoregression (41) and the regular mildly explosive instrument z_{2t}^{+} in (42), (A.10) implies that $\pi_n(\tilde{\rho}_n-\rho_n)\to_d -Y/X=_d Y/X$ and $\pi_n\left(\tilde{\beta}_n-\beta\right)\to_d -Y^{\varepsilon}/X=_d Y^{\varepsilon}/X$ by the symmetry of $\mathcal{MN}(0,\sigma^2/X^2)$ and $\mathcal{MN}(0,\sigma_{\varepsilon}^2/X^2)$ around 0. This proves part (iii) of the theorem.

We proceed to prove part (ii) of the theorem under Assumption C$_+$(ii). In the notation of (A.1) and Lemma 5, $\left|\tilde{U}_n(1)-U_n(1)\right|=o_p(1)$ by the approximation for r_{1n} of Lemma B2(ii) and Lemma 3.2(i) of Magdalinos and Phillips (2020). Using Lemma 2(iii) and combining (21), (A.1) and (A.9) and recalling (43) and the above approximation for $\tilde{U}_n(1)$, we obtain

$$\begin{aligned}\pi_n(\tilde{\rho}_n-\rho_n) &= n^{1/2}\left(1-\varphi_{1n}^2\right)^{-1/2}(\tilde{\rho}_{1n}-\rho_n)\mathbf{1}_{F_n^{+}}+2n^{1/2}\left(\varphi_{2n}^2-1\right)^{-1/2}(\tilde{\rho}_{2n}-\rho_n)\mathbf{1}_{\bar{F}_n^{+}}+o_p(1)\\ &= \frac{U_n(1)}{\tilde{\Psi}_n}\mathbf{1}_{F_n}+\frac{Y_n(1)}{\underline{X}_n}\mathbf{1}_{\bar{F}_n}+o_p(1),\end{aligned} \tag{A.11}$$

where $U_n(\cdot)$ and $Y_n(\cdot)$ are defined in Lemma 5 (with $u_t=e_t$ under Assumption 2) and the last line follows since $\left|\mathbf{1}_{F_n^{+}}-\mathbf{1}_{F_n}\right|\to_p 0$ and $\left|\mathbf{1}_{\bar{F}_n^{+}}-\mathbf{1}_{\bar{F}_n}\right|\to_p 0$ by Lemma 2(iii). $\tilde{\Psi}_n$ in (36), $n^{-1/2}x_n$, $\mathbf{1}_{F_n}$ and $\mathbf{1}_{\bar{F}_n}$ are functionals of $B_n(s)=n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor}u_t$, on $D[0,1]$, so the functional CLT of Lemma 5 on $[U_n(s),B_n(s),Y_n(s)]$ and the continuous mapping theorem imply that

$$\frac{U_n(1)}{\tilde{\Psi}_n}\mathbf{1}_{F_n}+\frac{Y_n(1)}{\underline{X}_n}\mathbf{1}_{\bar{F}_n}\to_d\frac{U(1)}{\omega^2\Psi_-(c)}\mathbf{1}_{F_c}+\frac{Y}{\omega\Psi_+(c)}\mathbf{1}_{\bar{F}_c} \tag{A.12}$$

since, by Lemma 3(i), $\tilde{\Psi}_n\to_d\tilde{\Psi}(c)$ with $\sigma^2+2\rho\Gamma=\omega^2$ under C$_+$(ii), $\tilde{\Psi}(c)=\omega^2\Psi_-(c)$ on the event F_c and $2\left(J_c(1)-\int_0^1 J_c(r)\,dr\right)=2\omega\Psi_+(c)$ on the event $\bar{F}_c$. The continuous mapping theorem is applicable to (A.12) because $x = 0$ is the only discontinuity point of the function $x \mapsto \mathbf{1}_{(-\infty,0]}(x)$ and $\mathbb{P}(K_c + c = 0) = 0$ since K_c in (35) is a continuously distributed random variable for all $c \in \mathbb{R}$. Denoting $\zeta := [\sigma^{-2}U(1), \sigma^{-1}Y]'$, Lemma 5 implies that ζ is independent of $\mathcal{F}_B = \sigma(B(s) : s \in [0,1])$ and $\zeta =_d \mathcal{N}(0, I_2)$. Since the random variables $J_c(1)$, $\Psi(c)$, $\mathbf{1}_{F_c}$ and $\mathbf{1}_{\bar{F}_c}$ are $\mathcal{F}_B$-measurable (as non-stochastic functionals of $B(r)$ on $D[0,1]$) the independence of ζ and $\mathcal{F}_B$ implies the independence of the random vectors ζ and $\left[J_c(1), \tilde{\Psi}(c), \mathbf{1}_{F_c}, \mathbf{1}_{\bar{F}_c}\right]'$. Under Assumption 2, $\omega^2 = \sigma^2$ and we conclude that the limit in (A.12) is given by $\left[\frac{1}{\Psi_-(c)}\mathbf{1}_{F_c}, \frac{1}{\Psi_-(c)}\mathbf{1}_{\bar{F}_c}\right]\zeta$ has a $\mathcal{MN}\left(0, \frac{1}{\Psi_-^2(c)}\mathbf{1}_{F_c} + \frac{1}{\Psi_+^2(c)}\mathbf{1}_{\bar{F}_c}\right)$ distribution as required by the theorem for $\pi_n(\tilde{\rho}_n - \rho_n)$. For $\pi_n\left(\tilde{\beta}_n - \beta\right)$, the same argument applies with $\tilde{U}_n(s)$ and Y_n replaced by $\tilde{U}_n^{\varepsilon}(s)$ and Y_n^{ε} in (A.11); defining $\tilde{U}_n(s)$ and $Y_n^{\varepsilon}(s)$ as $U_n(s)$ and $Y_n(s)$ with e_t replaced by ε_t, Lemma 5 implies that

$$\pi_n\left(\tilde{\beta}_n - \beta\right) \to_d \frac{\tilde{U}(1)}{\omega^2\Psi_-(c)}\mathbf{1}_{F_c} + \frac{Y^{\varepsilon}}{\omega\Psi_+(c)}\mathbf{1}_{\bar{F}_c} = \frac{\sigma_\varepsilon}{\omega}\left[\frac{1}{\Psi_-(c)}\mathbf{1}_{F_c}, \frac{1}{\Psi_+(c)}\mathbf{1}_{\bar{F}_c}\right]\tilde{\zeta}$$

where $\tilde{\zeta} =_d \mathcal{N}(0, I_2)$, which yields the limit distribution under C_+(ii).

Under C_-(ii), Lemma 2 (iv) and Lemma 6 imply that

$$\begin{aligned}\pi_n(\tilde{\rho}_n - \rho_n) &= \pi_n\left(\tilde{\rho}_{1n}^- - \rho_n\right)\mathbf{1}_{F_n^-} + \pi_n\left(\tilde{\rho}_{2n}^- - \rho_n\right)\mathbf{1}_{\bar{F}_n^-} + o_p(1) \\ &= -\left[\pi_n\left(\tilde{\rho}_{1n}^+ - |\rho_n|\right)\mathbf{1}_{F_n^{++}} + \pi_n\left(\tilde{\rho}_{2n}^+ - |\rho_n|\right)\mathbf{1}_{\bar{F}_n^{++}}\right] + o_p(1) \qquad \text{(A.13)}\end{aligned}$$

Now $\tilde{\rho}_{1n}^+ = \sum_{t=1}^n x_t^+ \tilde{z}_{0t-1}^+ / \sum_{t=1}^n x_{t-1}^+ \tilde{z}_{0t-1}^+$ is an IV estimator generated by the (regular) $C_+(ii)$ autoregression (41) with $\mu = 0$ and the regular instrument $\tilde{z}_{0t}^+$ in (42) with denominator satisfying

$$\begin{aligned}\tilde{\Psi}_n^+ &= \pi_n^{-2}\textstyle\sum_{t=1}^n x_{0t-1}\tilde{z}_{0t-1} + o_p(1) = 2\left(1 - |\rho_n|\,|\varphi_{1n}^-|\right)\frac{1}{n}\sum_{t=1}^n x_{t-1}^+\tilde{z}_{0t-1}^+ + o_p(1) \\ &= \sigma^2 + \frac{1}{n}x_{0n}^{+2} + 2\frac{1}{n}\textstyle\sum_{t=1}^n z_{1t-1}^+(-1)^t u_t + o_p(1) \qquad \text{(A.14)}\end{aligned}$$

with $z_{1t}^+ = |\varphi_{1n}^-|\, z_{1t-1}^+ + (-1)^t u_t$, by combining (B.9) and (B.12) in the proof of Lemma 3 applied to the recursions for x_t^+ and $\tilde{z}_{0t}^+$ in (41) and (42). Hence, (A.11) and (A.12) imply that

$$\frac{n^{1/2}\left(\tilde{\rho}_{1n}^+ - |\rho_n|\right)}{(1 - |\varphi_{1n}^-|^2)^{1/2}}\mathbf{1}_{F_n^{++}} + \frac{2n^{1/2}\left(\tilde{\rho}_{2n} - |\rho_n|\right)}{(|\varphi_{2n}^-|^2 - 1)^{1/2}}\mathbf{1}_{\bar{F}_n^{++}} = \frac{U_n^+(1)}{\tilde{\Psi}_n^+}\mathbf{1}_{F_n} + \frac{Y_n^+(1)}{n^{-1/2}x_n^+}\mathbf{1}_{\bar{F}_n} + o_p(1) \qquad \text{(A.15)}$$

$U_n^+(1) = n^{1/2}(1 - |\varphi_{1n}^-|^2)^{1/2}\sum_{t=1}^n z_{1t-1}^+(-1)^t u_t$, $Y_n^+(1) = (|\varphi_{2n}^-|^2 - 1)^{1/2}\sum_{t=1}^n |\varphi_{2n}^-|^{-(n-t)-1}(-1)^t u_t$, the simplification of the indicator functions on the right of (A.15) follows from $\left|\mathbf{1}_{F_n^{++}} - \mathbf{1}_{F_n}\right| \to_p 0$ and $\left|\mathbf{1}_{\bar{F}_n^{++}} - \mathbf{1}_{\bar{F}_n}\right| \to_p 0$ in view of Lemma 2(iii). By a Lindeberg-type FCLT (e.g. Theorem 3.33 of Jacod and Shiryaev (2003)) $n^{-1/2}\sum_{j=1}^{\lfloor nt \rfloor}(-1)^j u_j = C(1)\, n^{-1/2}\sum_{j=1}^{\lfloor nt \rfloor}(-1)^j e_j + o_p(1) \Rightarrow B_+(t)$ where $B_+(\cdot)$ is a Brownian motion with variance ω^2 under Assumption 4 and, since $n(1 - |\rho_n|) = n(1 + \rho_n) \to -c$ by Assumption 1b, $n^{-1/2}x_{\lfloor nt \rfloor}^+ \Rightarrow \int_0^t e^{c(t-s)}dB_+(s)$. Since $\{B_+(t) =_d B(t) : t \in [0,1]\}$ and $n^{-1}\sum_{t=1}^n z_{1t-1}^+(-1)^t u_t \to_p -\Gamma$ by (A.4), we conclude by (A.14) that $\tilde{\Psi}_n^+ \to_d \sigma^2 - 2\Gamma + J_c(1)^2$. Under Assumption 2, $\Gamma = 0$ and applying Lemma 5 to the right side of (A.15) and substituting into (A.13) we conclude that

$$\pi_n(\tilde{\rho}_n-\rho_n)\to_d-\left(\frac{\sigma^{-2}U(1)}{1+W_c(1)^2}\mathbf{1}_{F_c}+\frac{\sigma^{-1}Y(1)}{W_c(1)}\mathbf{1}_{\bar{F}_c}\right)=_d\mathcal{MN}\left(0,\left(1+W_c(1)^2\right)^{-2}\mathbf{1}_{F_c}+W_c(1)^{-2}\mathbf{1}_{\bar{F}_c}\right)$$

by symmetry of a centred mixed Gaussian distribution around 0. The proof for $\tilde{\beta}_n$ under Assumption 4 follows a similar argument with $\Psi_-(c)=\sigma^2-2\Gamma+J_c(1)^2$.

It remains to prove that $\pi_n\left(\tilde{\beta}_n-\beta_n^*\right)\to_p 0$; since $\beta_n^*=\tilde{\beta}_n$ on the event $\bar{F}_n\cup\{\hat{\rho}_n<0\}$ by construction, it is enough to show the result under C$_+$(i)-C$_+$(ii): $\pi_n\left(\tilde{\beta}_{1n}-\beta_{1n}^*\right)\to_p 0$ with $\pi_n=n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}$. From the definitions in (22) and (27)

$$\pi_n\left(\tilde{\beta}_{1n}-\beta_{1n}^*\right)=\pi_n^{-1}x_n\bar{z}_{1n-1}\left(\pi_n^{-2}\sum_{t=1}^n\underline{x}_{t-1}\tilde{z}_{1t-1}\right)^{-1}\hat{\rho}_{\varepsilon u}\hat{\sigma}_\varepsilon/\hat{\omega}_u,$$

so it is enough to show that $\pi_n^{-1}x_n\bar{z}_{1n-1}\to_p 0$. By Lemma B2(iii) and $x_n=O_p(\kappa_n^{1/2})$, $x_n\bar{z}_{1n-1}=O_p((1-\rho_n^2\varphi_{1n}^2)^{-1})$; $\pi_n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1}=n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\to 0$ completes the proof.

Proof of Theorem 1 and Theorem 2. For $i\in\{1,2\}$, denote $\Psi_{in}=\pi_n^{-2}\sum_{t=1}^n\underline{x}_{t-1}\tilde{z}_{it-1}$ and $\zeta_{in}=\left(\pi_n^{-2}\sum_{t=1}^n\tilde{z}_{it-1}^2\right)^{-1/2}\pi_n^{-1}\sum_{t=1}^n\tilde{z}_{it-1}v_t$, and by ζ_{in}^- and Ψ_{in}^- the oscillating counterparts of ζ_{in} and Ψ_{in} with $\tilde{z}_{it-1}$ replaced by $\tilde{z}_{it-1}^-$, where $v_t:=u_t/\sigma$ for $\tilde{T}_n(\rho_n)$ and $v_t:=\varepsilon_t/\sigma_\varepsilon$ for $\tilde{T}_n(\beta_n)$. By using (18) and $\hat{\sigma}_n\to_p\sigma$, $\hat{\sigma}_\varepsilon\to_p\sigma_\varepsilon$ we obtain that $\tilde{T}_n(\rho_n)=[1+o_p(1)]\,T_n$, and $\tilde{T}_n(\beta_n)=[1+o_p(1)]\,T_n$ where

$$T_n=T_{1n}\mathbf{1}_{F_n^+}+T_{1n}^-\mathbf{1}_{F_n^-}+T_{2n}\mathbf{1}_{\bar{F}_n^+}+T_{2n}^-\mathbf{1}_{\bar{F}_n^-},\quad T_{in}=\frac{|\Psi_{in}|}{\Psi_{in}}\zeta_{in}\;T_{in}^-=\frac{\left|\Psi_{in}^-\right|}{\Psi_{in}^-}\zeta_{in}^-\quad i\in\{1,2\}\,. \tag{A.16}$$

Proving the more general result $T_n\to_d\mathcal{N}(0,1)$ for any innovation sequence (v_t) satisfying Assumption 2 with $\mathbb{E}_{\mathcal{F}_{t-1}}(v_t^2)=1$ *a.s.* and x_t generated by (1) with innovations (u_t) satisfying Assumption 4 will establish the $\mathcal{N}(0,1)$ asymptotic distribution of both $\tilde{T}_n(\rho_n)$ under Assumption 2 and $\tilde{T}_n(\beta_n)$ under Assumption 4.

We first prove that $T_n\to_d\mathcal{N}(0,1)$ under the stronger Assumption 1b and then we employ Lemma 1 to extend the validity of the theorem under Assumption 1a. Under C(i), both Ψ_{1n} and $\Psi_{1n}^-=\tilde{\Psi}_n^++o_p(1)$ converge in probability to $\sigma^2+2\rho\Gamma$ (Lemma 3(i) for Ψ_{1n} and (A.4) for Ψ_{1n}^-), and both ζ_{1n} and ζ_{1n}^- converge in distribution to a $\mathcal{N}(0,1)$ (by Lemma 3 (ii) and (iii) with $\sigma_e^2=\mathbb{E}_{\mathcal{F}_{t-1}}(v_t^2)=1$ for ζ_{1n} and by (B.23), Lemma 6 and Lemma 3(ii) since $r_{2n}^-=o_p(1)$ in Lemma B2(ii) and $\sum_{t=1}^n\tilde{z}_{0t-1}^{-2}=\sum_{t=1}^n\tilde{z}_{0t-1}^{+2}$ for ζ_{1n}^-). Lemma 2 implies that: under C(i) with $\rho=1$, $T_n=T_{1n}+o_p(1)=(1+o_p(1))\,\zeta_{1n}\to_d\zeta_1$ and under C(i) with $\rho=-1$, $T_n=T_{1n}^-+o_p(1)=(1+o_p(1))\,\zeta_{1n}^-\to_d\zeta_1$, with $\zeta_1=_d\mathcal{N}(0,1)$ as required. To complete the proof under C(i), we need to show that $T_n\to_d\mathcal{N}(0,1)$ when $\rho_n\to\rho\in(-1,1)$, in which case $\hat{\rho}_n\to_p\rho_\Gamma:=\rho+\Gamma/(\sigma^2+2\rho\Gamma)$ under Assumption 4 and Lemma 2 implies that

$$\begin{aligned}T_n &= T_{1n}\mathbf{1}_{F_n^+}+T_{1n}^-\mathbf{1}_{F_n^-}+o_p(1)=(1+o_p(1))\left(\zeta_{1n}\mathbf{1}\{\rho_\Gamma\geq 0\}+\zeta_{1n}^-\mathbf{1}\{\rho_\Gamma<0\}\right)\\ &\to_d \left(\mathbf{1}\{\rho_\Gamma\geq 0\}+\mathbf{1}\{\rho_\Gamma<0\}\right)\zeta_1=_d\mathcal{N}(0,1)\,.\end{aligned}$$

Under C(iii), Lemma 2 implies that $T_n = T_{2n} + o_p(1)$ when $\rho \geq 1$ and $T_n = T_{2n}^{-} + o_p(1)$ when $\rho \leq -1$; By Lemma 4, $T_{2n} = (1 + o_p(1))(|X_n| / X_n) Y_n(v)$ with

$$Y_n(v) = \left(\varphi_{2n}^2 - 1\right)^{1/2} \sum_{t=1}^{n} \varphi_{2n}^{-(n-t)-1} v_t \to_d \zeta_2 =_d \mathcal{N}(0,1) \tag{A.17}$$

from the convergence $Y_n \to_d Y$ with $\omega^2 = 1$. Since $[X_n, Y_n(v)] \to_d [X, \zeta_2]$ where $X \neq 0$ *a.s.* and X is independent of ζ_2, $T_{2n} \to_d sign(X) \zeta_2 =_d \mathcal{N}(0,1)$. By $R_{4n}^{-} = o_p(1)$ in Lemma B2(iv), (B.23), (B.24) and $\sum_{t=1}^{n} z_{2t-1}^{-2} = \sum_{t=1}^{n} z_{2t-1}^{+2}$ we obtain

$$T_{2n}^{-} = -\frac{\left|\sum_{t=1}^{n} x_{t-1}^{+} z_{2t-1}^{+}\right|}{\sum_{t=1}^{n} x_{t-1}^{+} z_{2t-1}^{+}} \left(\sum_{t=1}^{n} z_{2t-1}^{+2}\right)^{-1/2} \sum_{t=1}^{n} z_{2t-1}^{+} (-1)^t v_t + o_p(1) = -\frac{|X_n^{+}|}{X_n^{+}} Y_n^{+}(v) + o_p(1)$$

with X_n^{+} defined as X_n in (40) with (ρ_n, u_t) replaced by $\left(|\rho_n|, (-1)^t u_t\right)$ and $Y_n^{+}(v)$ defined as $Y_n(v)$ in (A.17) with (φ_{2n}, v_t) replaced by $\left(\left|\varphi_{2n}^{-}\right|, (-1)^t v_t\right)$ satisfying $Y_n^{+}(v) \to_d \zeta_2^{+} =_d \mathcal{N}(0,1)$ by (A.17). Using Lemma 4 as in the C_+(iii) case yields $T_{2n}^{-} \to_d sign(X) \zeta_2^{+} =_d \mathcal{N}(0,1)$.

Under C_+(ii), defining $\hat{U}_n(\cdot)$ and $\hat{Y}_n(\cdot)$ in the same way as $U_n(\cdot)$ and $Y_n(\cdot)$ in Lemma 5 with u_t replaced by v_t, Lemmata 2, 3, 4 and 5 give

$$T_n = \omega^{-1} \left(\left|\tilde{\Psi}_n\right| / \tilde{\Psi}_n\right) \hat{U}_n(1) \mathbf{1}_{F_n^{+}} + \left(|\underline{X}_n| / \underline{X}_n\right) \hat{Y}_n(1) \mathbf{1}_{\bar{F}_n^{+}} + o_p(1) \to_d T_c \tag{A.18}$$

where $T_c := sign(\Psi_1) \zeta_1 \mathbf{1}_{F_c} + sign(\Psi_2) \zeta_2 \mathbf{1}_{\bar{F}_c}$, $\Psi_1 = \omega^2 + J_c(1)^2 - 2 J_c(1) \int_0^1 J_c(r) dr$, $\Psi_2 = J_c(1) - \int_0^1 J_c(r) dr$ and $\zeta_1, \zeta_2 =_d \mathcal{N}(0,1)$ with ζ_1 independent of (Ψ_1, F_c) and ζ_2 independent of $\left(\Psi_1, \bar{F}_c\right)$. Since Ψ_1 and Ψ_2 are continuously distributed $\Psi_1 \Psi_2 \neq 0$ *a.s.*. By independence of ζ_1 and (Ψ_1, F_c) and the fact that $-\zeta_1 =_d \mathcal{N}(0,1)$ we obtain

$$\begin{aligned} \mathbb{P}(\zeta_1 sign(\Psi_1) \leq x, F_c) &= \mathbb{P}(\zeta_1 \leq x, F_c, \Psi_1 > 0) + \mathbb{P}(-\zeta_1 \leq x, F_c, \Psi_1 < 0) \\ &= \mathbb{P}(\zeta_1 \leq x) \mathbb{P}(F_c, \Psi_1 > 0) + \mathbb{P}(-\zeta_1 \leq x) \mathbb{P}(F_c, \Psi_1 < 0) \\ &= \Phi(x) [\mathbb{P}(F_c, \Psi_1 > 0) + \mathbb{P}(F_c, \Psi_1 < 0)] = \Phi(x) \mathbb{P}(F_c). \end{aligned}$$

The above argument also gives $\mathbb{P}\left(\zeta_2 sign(\Psi_2) \leq x, \bar{F}_c\right) = \Phi(x) \mathbb{P}\left(\bar{F}_c\right)$, so the distribution function of the limit T_c in (A.18) is given by

$$\mathbb{P}(T_c \leq x) = \mathbb{P}(\zeta_1 sign(\Psi_1) \leq x, F_c) + \mathbb{P}\left(\zeta_2 sign(\Psi_2) \leq x, \bar{F}_c\right) = \Phi(x) \left[\mathbb{P}(F_c) + \mathbb{P}\left(\bar{F}_c\right)\right] = \Phi(x).$$

Under C_-(ii), (A.18) continues to hold with $\underline{X}_n, \tilde{\Psi}_n, \hat{U}_n(1), \hat{Y}_n(1), F_n^{+}$ replaced by $n^{-1/2} x_n^{+}$, $\tilde{\Psi}_n^{+}$, $\hat{U}_n^{+}(1)$, $\hat{Y}_n^{+}(1)$, F_n^{-}, defined in (41), (A.14) and $\hat{U}_n^{+}(1)$, $\hat{Y}_n^{+}(1)$ are defined in (A.15) with u_t replaced by v_t. Since $\mathbf{1}_{F_n^{-}} \to_d \mathbf{1}_{F_c}$ by Lemma 2(iv), the argument employed under C_+(ii) continues to apply with Ψ_1 and Ψ_2 replaced by $\Psi_1^{+} = \sigma^2 - 2\Gamma + J_c(1)^2$ and $\Psi_2^{+} = J_c(1)$ and shows that $T_c^{+} := sign\left(\Psi_1^{+}\right) \zeta_1 \mathbf{1}_{F_c} + sign\left(\Psi_2^{+}\right) \zeta_2 \mathbf{1}_{\bar{F}_c} =_d \mathcal{N}(0,1)$.

The above argument proves that T_n in (A.16) satisfies $T_n \to_d \mathcal{N}(0,1)$ for any (v_t) satisfying Assumption 2 with $\mathbb{E}_{\mathcal{F}_{t-1}}(v_t^2) = 1$ *a.s.* under Assumption 1b, when $(\rho_n)_{n \in \mathbb{N}}$ in (1) belongs to one of the autoregressive classes C(i)-C(iii) of Assumption 1b.

Now suppose that $(\rho_n)_{n \in \mathbb{N}}$ in (1) satisfies Assumption 1a and consider an arbitrary subsequence

$\left(\rho_{k_n}\right)_{n\in\mathbb{N}}$ of $(\rho_n)_{n\in\mathbb{N}}$ and $(T_{k_n})_{n\in\mathbb{N}}$ of $(T_n)_{n\in\mathbb{N}}$. By Lemma 1, there exists a further subsequence $\left(\rho_{m_n}\right)_{n\in\mathbb{N}}$ of $\left(\rho_{k_n}\right)_{n\in\mathbb{N}}$ satisfying Assumption 1b; as a result, $\left(\rho_{m_n}\right)_{n\in\mathbb{N}}$ belongs to one of the classes C(i)-C(iii) and the preceding argument shows that $T_{m_n} \to_d \mathcal{N}(0,1)$. We conclude that for any subsequence $(T_{k_n})_{n\in\mathbb{N}}$ of $(T_n)_{n\in\mathbb{N}}$ there exists a further subsequence $(T_{m_n})_{n\in\mathbb{N}}$ of $(T_{k_n})_{n\in\mathbb{N}}$ such that $T_{m_n} \to_d \mathcal{N}(0,1)$; but this implies that the entire sequence $(T_n)_{n\in\mathbb{N}}$ satisfies $T_n \to_d \mathcal{N}(0,1)$.

Using the argument following (A.16), we conclude that $\tilde{T}_n(\rho_n) \to_d \mathcal{N}(0,1)$ under Assumptions 1a and 2-3 and $\tilde{T}_n(\beta_n) \to_d \mathcal{N}(0,1)$ under Assumptions 1a and 3-4. To prove $T_n^*(\beta_n) \to_d \mathcal{N}(0,1)$, we first show that $\tilde{T}_n(\beta_n) - T_n^*(\beta_n) \to_p 0$ under Assumption 1b. Since $T_n^*(\beta_n) = \tilde{T}_n(\beta_n)$ on the event $\bar{F}_n \cup \{\hat{\rho}_n < 0\}$, it suffices to show that $\tilde{T}_{1n} - T_{1n}^* \to_p 0$ (in the notation of (A.16)) under C(i)-C(ii) of Assumption 1b. Since $n\bar{z}_{1n}^2 = o_p\left(n\left(1-\rho_n\varphi_{1n}\right)^{-1}\right) = o_p\left(\pi_n^2\right)$ by Lemma B2(iii), the denominator of T_{1n}^* in (28), $v_{1n}^2 := \sum_{t=1}^n \tilde{z}_{t-1}^2 - n\bar{z}_{1,n-1}^2\left(1-\hat{\rho}_{\varepsilon u}^2\right)\mathbf{1}_{F_n}$, satisfies $\pi_n^{-2}\left(\sum_{t=1}^n \tilde{z}_{1t-1}^2 - v_{1n}^2\right) \to_p 0$.

$$\left|\tilde{T}_{1n} - T_{1n}^*\right| \leq \frac{1}{\hat{\sigma}_\varepsilon}\left|\pi_n^{-2}\textstyle\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{1t-1}\right|\left[\pi_n\left|\tilde{\beta}_{1n} - \beta_{1n}^*\right|\left(\pi_n^{-2}\textstyle\sum_{t=1}^n \tilde{z}_{1t-1}^2\right)^{-1/2} + o_p(1)\right] = o_p(1)$$

since $\pi_n\left|\tilde{\beta}_{1n} - \beta_{1n}^*\right| \to_p 0$ by Theorem 3 and the other sample moments are $O_p(1)$ by Lemma 3. This proves that $T_n^*(\beta_n) \to_d \mathcal{N}(0,1)$ under Assumption 1b and the subsequential argument employed on T_n shows that $T_n^*(\beta_n) \to_d \mathcal{N}(0,1)$ under Assumption 1a.

Next, we prove part (ii) of Theorem 1 on the uniform asymptotic size and asymptotic similarity of the CR $\mathcal{R}_n$ of Theorem 1 by verifying Assumptions A1 and S of Andrews, Cheng and Guggenberger (2020), (henceforth ACG (2020)). Let $(\rho_n)_{n\in\mathbb{N}} \subset \Theta = [-M, M]$; any subsequence $\left(\rho_{w_n}\right)_{n\in\mathbb{N}}$ of $(\rho_n)_{n\in\mathbb{N}}$ is bounded so it has a convergent subsequence $\left(\rho_{k_n}\right)_{n\in\mathbb{N}}$ in $[-M, M]$; since $\rho_{k_n} \to \rho \in [-M, M]$, $\left(\rho_{k_n}\right)_{n\in\mathbb{N}}$ satisfies Assumption 1a, so the asymptotic distribution of $\tilde{T}_{\rho_n}(\rho_n)$ under Assumptions 1a, 2 and 3 implies that

$$\lim_{n\to\infty}\mathbb{P}_{\rho_{k_n}}\left(\mathcal{R}_{k_n}\right) = \lim_{n\to\infty}\mathbb{P}_{\rho_{k_n}}\left(\left|\tilde{T}_{k_n}\left(\rho_{k_n}\right)\right| > \Phi^{-1}(1-\alpha/2)\right) = \alpha. \tag{A.19}$$

Convergence in (A.19) proves simultaneously the validity of Assumptions A1 and S of ACG(2020) over Θ, so part (ii) of Theorem 1 follows from Theorem 2.1(e) of ACG(2020). The uniformity of the coverage of the CI $I_n(\tilde{\rho}_n, \alpha)$ of part (iii) of Theorem 1 may be proved along similar lines: since $\mathbb{P}_\rho\left[\rho \in I_n(\tilde{\rho}_n, \alpha)\right] = \mathbb{P}_\rho\left(\left|T_n(\tilde{\rho}_n)\right| \leq \Phi^{-1}(1-\alpha/2)\right)$, we may employ the above argument by replacing $\mathcal{R}_n$ by $\left\{\left|T_n(\tilde{\rho}_n)\right| \leq \Phi^{-1}(1-\alpha/2)\right\}$ so that (A.19) holds with α replaced by $1-\alpha$, verifying Assumptions A1 and S of ACG(2020). This completes the proof of Theorem 1.

For Theorem 2, the asymptotic similarity of the critical regions $\tilde{\mathcal{R}}_n$ and $\mathcal{R}_n^*$ over the parameter space $\tilde{\Theta}$ may be proved along the same lines as the asymptotic similarity of $\mathcal{R}_n$ over Θ in Theorem 1: since $T_n^*(\beta_n)$ and $\tilde{T}_n(\beta_n)$ are asymptotically $\mathcal{N}(0,1)$ for all drifting sequences (ρ_n) satisfying Assumption 1a from the first part of Theorem 2, the argument for $\mathcal{R}_n$ may be employed. The proof for the CIs $I_n\left(\tilde{\beta}_n, \alpha\right)$ and $I_n(\beta_n^*, \alpha)$ is similar to that for $I_n(\tilde{\rho}_n, \alpha)$.

References

Allen, L.J.S. & Van den Driessche, P. (2008) The basic reproduction number in some discrete-time epidemic models, *Journal of Difference Equations and Applications* **14**: 1127-1147

Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* **30**: 676-687

Andrews, D.W.K. (1993) Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models. *Econometrica,* **61**(1): 139-165

Andrews, D.W.K. and Guggenberger, P. (2009) Hybrid and size-corrected subsampling methods. *Econometrica* **77**(3): 721-762

Andrews, D.W.K. and Guggenberger, P. (2012) Asymptotics for LS, GLS, and feasible GLS statistics in an AR(1) model with conditional heteroskedasticity. *Journal of Econometrics* **169**(2): 196-210

Andrews, D.W.K. and Guggenberger, P. (2014) A conditional-heteroskedasticity-robust confidence interval for the AR parameter. *Review of Economics and Statistics* **96**(2): 376-381

Andrews, D.W.K., Cheng, X. and Guggenberger, P. (2020) Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics* **218**(2): 496-531

Arvanitis, S. and Magdalinos, T. (2019) Mildly explosive autoregression under stationary conditional heteroscedasticity. *Journal of Time Series Analysis* **39**(6): 892-908

Breitung, J. and Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics* **187**: 358-375

Campbell, J. Y. and Yogo, M. (1996) Efficient tests of stock return predictability. *Journal of Financial Economics* **81**(1): 27-60

Cavaliere, G. and Georgiev, I. (2020) Inference Under Random Limit Bootstrap Measures, *Econometrica* **88** (6): 2547-2574

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

Demeterscu, M. Georgiev, I., Rodrigues P.M.M. and Taylor, A.M.R. (2022) Extensions to IVX methods of inference for return predictability. *Journal of Econometrics*, Forthcoming

Duffy, J.A. and Kasparis, I. (2020) Estimation and Inference in the presence of fractional $d = 1/2$ and weakly nonstationary processes, *Annals of Statistics* **49**(2): 1195-1217

Elliott, G., Müller, U. K. and Watson, M. W. (2015) Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis, *Econometrica* **83**(2): 771-811

Fei, Y. (2018) Limit theory for mildly integrated process with intercept. *Economics Letters* **163**: 98-101

Giraitis, L. and Phillips, P.C.B. (2006) Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis* **27**(1): 51-60

Hansen, B.E. (1999) The grid bootstrap and the autoregressive model. *Review of Economics and Statistics* **81**(4): 594-607

Harvey, D., Leybourne, S. and Taylor, R. (2021) Simple Tests for Stock Return Predictability with Good size and Power Properties. *Journal of Econometrics* **224**: 198-214

Hosseinkouchack, M. and Demetrescu, M. (2021) Finite-sample size control of IVX-based tests in predictive regressions. *Econometric Theory*, Forthcoming

Hu, Z., Kasparis, I. and Wang, Q. (2021) Chronological Trimming Methods for Nonlinear Predictive Regressions with Persistent Data. *Working Paper*

Jansson, M. and Moreira, M. J. (2006) Optimal Inference in Regression Models with Nearly Integrated Regressors, *Econometrica,* **74**(3): 681-714

Kermack, W. O. and McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society A* **115**: 700-721

Kostakis, A., Magdalinos, T. and Stamatogiannis, M.P. (2015) Robust Econometric Inference for Stock Return Predictability, *Review of Financial Studies* **28**(5):1506-1553

Lai, T.L. and Wei, C.Z. (1983) A note on martingale difference sequences satisfying the local Marcinkiewicz-Zygmund condition. *Bulletin of the Institute of Mathematics Academia Sinica* **11**: 1-13

Li, S. and Linton, O. (2021) When will the Covid-19 pandemic peak? *Journal of Econometrics* **220**: 130-157

Liu, L., Moon, H. R. and Schorfheide, F. (2021) Panel forecasts of country-level Covid-19 infections. *Journal of Econometrics* **220**: 2–22

Magdalinos, T. (2012) Mildly explosive autoregression under weak and strong dependence. *Journal of Econometrics* **169**(2): 179-187

Magdalinos, T. and Phillips, P.C.B. (2020) Econometric inference in matrix vicinities of unity and stationarity, *Working paper*

Mikusheva, A. (2007) Uniform inference in autoregressive models. *Econometrica* **75**(5): 1411-1452

Mann, H. and Wald, A. (1943) On the Statistical Treatment of Linear Stochastic Difference Equations. *Econometrica* **11**: 173-200

Phillips, P.C.B. (1987a) Time series regression with a unit root. *Econometrica* **55**(2): 277-301

Phillips, P.C.B. (1987b) Towards a unified asymptotic theory for autoregression. *Biometrika* **74**: 535-547

Phillips, P.C.B and Hansen, B. E. (1990) Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* **57**(1): 99-125

Phillips, P.C.B. and Magdalinos, T. (2007) Limit theory for Moderate deviations from a unit root. *Journal of Econometrics* **136**: 115-130

Phillips, P.C.B. and Magdalinos, T. (2009) Representation and inference in VAR processes with mixed integration orders, *Working Paper*

Phillips P.C B., Wu, Y. and Yu, J. (2011) Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values, *International Economic Review* **52**(1): 201-226

Phillips P. C. B. and Yu, J. (2011) Dating the Timeline of Financial Bubbles during the Subprime Crisis. *Quantitative Economics* **2**(3): 455-491

Romano, J.P. and Wolf, M. (2001) Subsampling intervals in autoregressive models with linear time trend. *Econometrica* **69**(5): 1283–1314

Stock, J. (1991) Confidence intervals for the largest autoregressive root in US macroeconomic time series. *Journal of Monetary Economics* **28**: 435-459

Wang, X. and Yu, J. (2015) Limit theory for an explosive autoregressive process *Economics Letters* **126**: 176-180

Werker, B.J.M. and Zhou, B. (2021) Semiparametric Testing with Highly Persistent Predictors. *Journal of Econometrics*, Forthcoming

White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* **29**: 1188–1197

Yang, B., Long, W., Peng, L. and Cai, Z. (2020) Testing the predictability of U.S. housing price index returns based on an IVX-AR model. *Journal of the American Statistical Association* **115**: 1598-1619

Zhou, B., van den Akker, R. and Werker, B.J.M. (2019) Semiparametrically point-optimal hybrid rank tests for unit roots, *Annals of Statistics* **47**(5): 2601-2638

Supplementary Online Appendix B

This online Appendix contains: (i) two auxiliary results (Lemma B2 and B2) in Section 1.1, (ii) the proofs of Lemmata 1-6 and Corollary 1 of the main paper, as well as the proofs of Lemmata B1 and B2 in Section 1.2, and (iii) some additional simulation results in Section 1.3 below.

1.1 Auxiliary mathematical results

Lemma B1 below is concerned with the limit distribution of the normalised and centred OLS estimator $\hat{\rho}_n$ in (14) obtained from the autoregression (5)/(13) under weakly dependent errors. Define the normalisation sequence

$$c_n = \begin{cases} n^{1/2}\left(1-\rho_n^2\right)^{-1/2}, & \text{under C(i) and Assumption 2} \\ \left(1-\rho_n^2\right)^{-1}, & \text{under C(i) and Assumption 4} \\ n, & \text{under C(ii) and Assumption 4} \\ \left(\rho_n^2-1\right)^{-1}|\rho_n|^n, & \text{under C(iii) and Assumption 4.} \end{cases}$$

Lemma B1. *Consider the autoregressions x_t in (5)/(13) and x_{0t} in (13) and the stochastic sequences X_n in (40) and $\Upsilon_n = \left(\rho_n^2-1\right)^{1/2}\sum_{t=1}^n \rho_n^{-(n-t+1)}u_t$. Under Assumptions 1b, 3 and 4, the following hold:*

(i) *$\bar{x}_{n-1} = [1+o_p(1)](\mu+\bar{x}_{0n-1})$ under C_+(i); $\bar{x}_{n-1} = [1+o_p(1)]\bar{x}_{0n-1} = O_p\left(n^{-1}|\rho_n|^n \kappa_n^{3/2}\right)$ under C_+(ii)-C_+(iii); $\bar{x}_{n-1} = \mu + o_p(1)$ under C_-(i)-C_-(ii); $\bar{x}_{n-1} = O_p\left(n^{-1}|\rho_n|^n\left(\rho_n^2-1\right)^{-1/2}\right)$ under C_-(iii). Under C(i)-C(iii) and Assumption 2 or C(ii)-C(iii) and Assumption 4*

$$c_n^{-1}\sum_{t=1}^n \underline{x}_{t-1}\underline{u}_t = c_n^{-1}\sum_{t=1}^n x_{0t-1}u_t - \mathbf{1}_{\{c\in\mathbb{R}\}}\bar{x}_{0n-1}\bar{u}_n + o_p(1) \tag{B.1}$$

$$c_n^{-2}\sum_{t=1}^n \underline{x}_{t-1}^2 = c_n^{-2}\sum_{t=1}^n x_{0t-1}^2 - \mathbf{1}_{\{c\in\mathbb{R}\}}n^{-1}\bar{x}_{0n-1}^2 + o_p(1). \tag{B.2}$$

Under C(i) and Assumption 4, *(B.2) continues to hold, and (B.1) holds with c_n^{-1} replaced by n^{-1}.*

(ii) *Under C_+(i), $c_n(\hat{\rho}_n-\rho_n) \to_d \xi$ with $\xi = \Gamma/\left(\sigma^2+2\rho\Gamma\right)$ under Assumption 4 and $\xi =_d \mathcal{N}(0,1)$ under Assumption 2.*

(iii) *Under C_+(iii), $c_n(\hat{\rho}_n-\rho_n) = \Upsilon_n/X_n + o_p(1)$ and $|\Upsilon_n/X_n| = O_p(1)$. When $\rho_n \to 1$, $\Upsilon_n/X_n \to_d \mathcal{C}$ (standard Cauchy distribution); when $\rho_n \to \rho > 1$ $\Upsilon_n/X_n \to_d \Upsilon_\infty/X_\infty$ where $\Upsilon_\infty =_d X_\infty$ and the random variables X_∞ and Y_∞ are independent.*

(iv) *The OLS estimator from an oscillating autoregression x_t in C_-(i)-C_-(iii) satisfies*

$$c_n(\hat{\rho}_n-\rho_n) = -c_n\left(\hat{\rho}_n^+ - |\rho_n|\right) + o_p(1)$$

where $\hat{\rho}_n^+$ is the OLS estimator of the autoregression (41) generated by $x_t^+ = (-1)^t x_{0t}$.

As a consequence of Lemma B1(iv), $\hat{\rho}_n - \rho_n$ requires the same normalisation when $\rho \geq 0$ and $\rho < 0$ across the autoregressive classes C(i)-C(iii) of Assumption 1b. Also, Lemma B1 and Phillips (1987b) imply that the following orders of magnitude apply under C(ii)-C(iii):

$$\sum_{t=1}^n x_{t-1}u_t = O_p\left(\kappa_n^{1/2}\nu_n\right), \ \sum_{t=1}^n x_{t-1}^2 = O_p\left(\kappa_n\nu_n^2\right) \text{ and } |\hat{\rho}_n-\rho_n| = O_p\left(\kappa_n^{-1/2}\nu_n^{-1}\right). \tag{B.3}$$

For the next result, Let x_{0t} denote the autoregression in (6) when $\mu = x_0 = 0$ and $\tilde{z}_{0t} = \sum_{j=1}^t \varphi_{1n}^{t-j}\Delta x_{0j}$ and $\tilde{z}_{0t}^- = \sum_{j=1}^t \left(\varphi_{1n}^-\right)^{t-j}\nabla x_{0j}$ be the instruments in (19) generated by x_{0t}.

Lemma B2. *Consider the instruments in (19)-(20) and the processes z_{1t}, and z_{2t} in (32) and (33). Under Assumptions 1b, 3 and 4, the following hold:*

(i) $[n(1-\varphi_{1n})]^p \varphi_{1n}^n \to 0$, $\sum_{t=1}^n t^p \varphi_{1n}^t \sim (1-\varphi_{1n})^{-p-1}\Gamma(p+1)$ *for any* $p \geq 0$ *and any sequence* $(\varphi_{1n})_{n\in\mathbb{N}}$ *in C(i);* $[n(\varphi_{2n}-1)]^p \varphi_{2n}^{-n} \to 0$, $\sum_{t=1}^n t^p \varphi_{2n}^{-t} \sim (\varphi_{2n}-1)^{-p-1}\Gamma(p+1)$ *for any* $p \geq 0$ *and any sequence* $(\varphi_{2n})_{n\in\mathbb{N}}$ *in C(iii), where* $\Gamma(\cdot)$ *denotes the gamma function.*

(ii) *Let with* $\pi_n = n^{1/2}\left(1-\rho_n^2\phi_{1n}^2\right)^{-1/2}$ *with* ϕ_{1n} *defined in (43). Under C(i)-C(ii), consider the sequences* $r_{1n} = \pi_n^{-1}\sum_{t=1}^n (\tilde{z}_{1t-1}-\tilde{z}_{0t-1})u_t$, $r_{2n} = \pi_n^{-2}\sum_{t=1}^n (\tilde{z}_{1t}^2-\tilde{z}_{0t}^2)$, $r_{3n} = \pi_n^{-2}\sum_{t=1}^n (\tilde{z}_{1t}x_t - \tilde{z}_{0t}x_{0t})$ *and* $r_{1n}^-, r_{2n}^-, r_{3n}^-$ *with* r_{jn}^- *defined as* r_{jn} *with* $(\tilde{z}_{1t}, \tilde{z}_{0t})$ *replaced by* $\left(\tilde{z}_{1t}^-, \tilde{z}_{0t}^-\right)$*. Then* $r_{jn} = o_p(1)$ *and* $r_{jn}^- = o_p(1)$ *for all* $j \in \{1,2,3\}$*.*

(iii) *Under C(i)-C(ii)* $(1-\rho_n\varphi_{1n})\sum_{t=1}^n \tilde{z}_{1t} = O_p\left(n^{1/2}\right) + o_p\left(n\kappa_n^{-1/2}\right)$ *where* (κ_n) *is defined in (4); under C(ii),* $(1-\varphi_{1n})n^{-1/2}\sum_{t=1}^n \tilde{z}_{1t} = n^{-1/2}x_{0n-1} + o_p(1)$*.*

(iv) *Under C(ii)-C(iii),* $\tilde{z}_{2t} = z_{2t} - r_{nt}$ *and* $\tilde{z}_{2t}^- = z_{2t}^- - r_{nt}^-$*, where*

$$r_{nt} = \frac{\hat{\rho}_n - \rho_n}{\phi_{2n} - \rho_n}\left(\phi_{2n}z_{2t-1} - \rho_n x_{t-1}\right)\mathbf{1}\left\{n\left|\phi_{2n}-\rho_n\right| \to \infty\right\} + \phi_{2n}^t g_n \tag{B.4}$$

$g_n = o_p\left(n^{-1/2}(\phi_{2n}-1)^{-1}\right)$ *independently of* t *and* r_{nt}^- *is given by (B.4) with* z_{2t-1} *replaced by* z_{2t-1}^-*. Consider* $R_{1n} = (\phi_{2n}-1)\nu_{n,z}^{-1}\left(\sum_{t=1}^n \tilde{z}_{2t} - \sum_{t=1}^n z_{2t}\right)$, $R_{2n} = \left(\phi_{2n}^2-1\right)|\phi_{2n}|^{-n}\sum_{t=1}^n (\tilde{z}_{2t-1}-z_{2t-1})u_t$, $R_{3n} = s_n^{-1}\sum_{t=1}^n (\tilde{z}_{2t}x_t - z_{2t}x_{0t})$, $R_{4n} = \left(\phi_{2n}^2-1\right)^2|\phi_{2n}|^{-2n}\left(\sum_{t=1}^n \tilde{z}_{2t}^2 - \sum_{t=1}^n z_{2t}^2\right)$ *and* R_{1n}^-, R_{2n}^-, R_{3n}^-, R_{4n}^- *with* R_{jn}^- *defined as* R_{jn} *with* $(\tilde{z}_{2t}, z_{2t})$ *replaced by* $\left(\tilde{z}_{2t}^-, z_{2t}^-\right)$*. Then* $R_{jn} = o_p(1)$ *and* $R_{jn}^- = o_p(1)$ *for all* $j \in \{1,2,3,4\}$*.*

(v) *Define* $\left(\check{Y}_n, \check{Z}_n\right)$ *by replacing* u_j *by* $C(1)e_j$ *in the expressions for* (Y_n, Z_n) *in (39). The following approximation holds:* $\left(\check{Y}_n, \check{Z}_n\right) - (Y_n, Z_n) \to_p 0$*.*

1.2 Mathematical Proofs

Proof of Lemma 1. Convergence of $(\rho_n)_{n\in\mathbb{N}}$ to $\rho \in \mathbb{R}$ ensures that Assumption 1b holds for the entire sequence $(\rho_n)_{n\in\mathbb{N}}$ when $\rho \neq 1$, so it is enough to show the result for $\rho = 1$. Denote $(c_n)_{n\in\mathbb{N}} := \{n(|\rho_n|-1) : n \in \mathbb{N}\}$. Given an arbitrary subsequence $\left(\rho_{m_n}\right)_{n\in\mathbb{N}}$ of $(\rho_n)_{n\in\mathbb{N}}$, $(c_{m_n})_{n\in\mathbb{N}}$ has a monotone subsequence $(c_{s_n})_{n\in\mathbb{N}}$ (by the monotone subsequence theorem for real sequences). By monotonicity, $(c_{s_n})_{n\in\mathbb{N}}$ converges to $c_\infty \in \mathbb{R}\cup\{-\infty,\infty\}$; hence: $\left(\rho_{s_n}\right)_{n\in\mathbb{N}}$ belongs to C(i) if $c_\infty = -\infty$, or $\left(\rho_{s_n}\right)_{n\in\mathbb{N}}$ belongs to C(ii) if $c_\infty \in \mathbb{R}$, or $\left(\rho_{s_n}\right)_{n\in\mathbb{N}}$ belongs to C(iii) if $c_\infty = \infty$.

Proof of Lemma 2. Writing $n(|\hat{\rho}_n|-1) = n(|\hat{\rho}_n|-|\rho_n|) + n(|\rho_n|-1)$ we obtain the identity

$$n(|\hat{\rho}_n|-1) = n(|\rho_n|-1)(1-\epsilon_n), \quad \epsilon_n = \frac{|\hat{\rho}_n|-|\rho_n|}{1-|\rho_n|} \tag{B.5}$$

We first show that, under C(i) and C(iii),

$$\limsup_{n\to\infty} \mathbb{P}(\epsilon_n > 1-\eta) = 0 \text{ for some } \eta \in (0,1). \tag{B.6}$$

The inequality $||x|-|y|| \leq |x-y|$ implies that $|\epsilon_n| \leq |\hat{\rho}_n - \rho_n|\,|1-|\rho_n||^{-1} \to_p 0$ under C(iii) and Assumption 4 ($|\hat{\rho}_n-\rho_n|(|\rho_n|-1)^{-1} = O_p\left(|\rho_n|^{-n}\right)$ by Lemma B1 (iii) and (iv)) and under C(i) and Assumption 2 ($|\hat{\rho}_n-\rho_n|(1-|\rho_n|)^{-1} = O_p(n^{-1/2}(1-|\rho_n|)^{-1/2})$ by Lemma B1(ii) and (iv)), showing (B.6) for the above cases. It remains to prove (B.6) under C(i) and Assumption 4: writing

$$\begin{aligned}\epsilon_n &= \frac{1}{|\hat{\rho}_n|+|\rho_n|}\frac{\hat{\rho}_n^2-\rho_n^2}{1-|\rho_n|} = \frac{\hat{\rho}_n+\rho_n}{|\hat{\rho}_n|+|\rho_n|}(1+|\rho_n|)\left(1-\rho_n^2\right)^{-1}(\hat{\rho}_n-\rho_n) \\ &\leq (1+|\rho_n|)\left(1-\rho_n^2\right)^{-1}(\hat{\rho}_n-\rho_n)\end{aligned} \tag{B.7}$$

because $\left|\frac{\hat{\rho}_n+\rho_n}{|\hat{\rho}_n|+|\rho_n|}\right| \leq 1$. Lemma B1 (ii) and (iv) imply that

$$(1+|\rho_n|)\left(1-\rho_n^2\right)^{-1}(\hat{\rho}_n-\rho_n) \to_p \epsilon(\rho) = \epsilon_+(\rho)\mathbf{1}_{[0,1]}(\rho) + \epsilon_-(\rho)\mathbf{1}_{[-1,0]}(\rho) \tag{B.8}$$

where $\epsilon_+(\rho) = \frac{(1+\rho)\Gamma}{\sigma^2+2\rho\Gamma}$ and $\epsilon_-(\rho) = -\frac{(1-\rho)\Gamma}{\sigma^2+2\rho\Gamma}$. We will prove (B.6) from (B.7) and (B.8) by showing that $\sup_{\rho\in[0,1]}\epsilon_+(\rho) < 1$ and $\sup_{\rho\in[-1,0]}\epsilon_-(\rho) < 1$. We may assume that $\Gamma > 0$ when $\rho \in [0,1]$ and $\Gamma < 0$ when $\rho \in [-1,0]$ (otherwise $\epsilon_+(\rho) \leq 0$ and $\epsilon_-(\rho) \leq 0$ and there is nothing to prove). Differentiating, $\epsilon_+'(\rho) = \Gamma(\sigma^2-2\Gamma)(\sigma^2+2\rho\Gamma)^{-2}$ and $\epsilon_-'(\rho) = \Gamma(\sigma^2+2\Gamma)(\sigma^2+2\rho\Gamma)^{-2}$, so $\epsilon_+'(\rho)$ is increasing on $[0,1]$ if and only if $\sigma^2 > 2\Gamma$ and $\epsilon_-'(\rho)$ is increasing on $[-1,0]$ if and only if $\sigma^2 < -2\Gamma$. Hence, when $\sigma^2 > 2\Gamma$, $\sup_{\rho\in[0,1]}\epsilon_+(\rho) \leq \epsilon_+(1) = 2\Gamma/(\sigma^2+2\Gamma) < 1$; when $\sigma^2 < 2\Gamma$, $\sup_{\rho\in[0,1]}\epsilon_+(\rho) \leq \epsilon_+(0) = 0$ (since $\Gamma = 0$ when $\rho = 0$ by (34)); when $\sigma^2 = 2\Gamma$, $\epsilon_+(\rho) = 1/2$, showing that $\sup_{\rho\in[0,1]}\epsilon_+(\rho) < 1$. Similarly, when $\sigma^2 < -2\Gamma$, $\sup_{\rho\in[-1,0]}\epsilon_-(\rho) \leq \epsilon_-(0) = 0$; when $\sigma^2 > -2\Gamma$, $\sup_{\rho\in[-1,0]}\epsilon_-(\rho) \leq \epsilon_-(-1) = -2\Gamma/(\sigma^2-2\Gamma) < 1$ (since $\Gamma < 0$); when $\sigma^2 = -2\Gamma$, $\epsilon_-(\rho) = 1/2$, showing that $\sup_{\rho\in[-1,0]}\epsilon_-(\rho) < 1$. We conclude that the limit in (B.8) satisfies $\sup_{\rho\in[-1,1]}\epsilon(\rho) < 1$, so the inequality (B.7) implies that (B.6) is satisfied under C(i) and Assumption 4 with $\eta \in \left(0, 1-\sup_{\rho\in[-1,1]}\epsilon(\rho)\right)$. This completes the proof of (B.6).

For part (i), $\mathbf{1}_{\bar{F}_n^+} \leq \mathbf{1}_{\bar{F}_n}$ and $\mathbf{1}_{\bar{F}_n^-} \leq \mathbf{1}_{\bar{F}_n}$ so it is sufficient to show that $m_n\mathbf{1}_{\bar{F}_n} \to_p 0$. For some $\eta \in (0,1)$ that satisfies (B.6) and using (B.5), we obtain for any $\delta > 0$

$$\begin{aligned}
\mathbb{P}(m_n\mathbf{1}_{\bar{F}_n} > \delta) &\leq \mathbb{P}(m_n\mathbf{1}_{\bar{F}_n} > \delta, \epsilon_n \leq 1-\eta) + \mathbb{P}(\epsilon_n > 1-\eta) \\
&= \mathbb{P}(m_n\mathbf{1}\{n(|\rho_n|-1)(1-\epsilon_n) > 0\} > \delta, \epsilon_n \leq 1-\eta) + \mathbb{P}(\epsilon_n > 1-\eta) \\
&\leq \mathbb{P}(m_n\mathbf{1}\{n(|\rho_n|-1)\eta > 0\} > \delta) + \mathbb{P}(\epsilon_n > 1-\eta) \\
&\leq \mathbb{P}(\epsilon_n > 1-\eta)
\end{aligned}$$

for all $n \geq n_0(\delta)$ because $n(|\rho_n|-1)\eta \to -\infty$, so $\mathbf{1}\{n(|\rho_n|-1)\eta > 0\} = 0$ for all but finitely many n; since η satisfies (B.6), part (i) follows.

For part (ii), for some η satisfying (B.6) and any $\delta > 0$ we may write

$$\begin{aligned}
\mathbb{P}(m_n\mathbf{1}_{F_n} > \delta) &= \mathbb{P}(m_n\mathbf{1}\{n(|\rho_n|-1)(1-\epsilon_n) \leq 0\} > \delta, \epsilon_n \leq 1-\eta) + \mathbb{P}(\epsilon_n > 1-\eta) \\
&\leq \mathbb{P}(m_n\mathbf{1}\{n(|\rho_n|-1)\eta \leq 0\} > \delta) + \mathbb{P}(\epsilon_n > 1-\eta) \\
&= \mathbb{P}(\epsilon_n > 1-\eta)
\end{aligned}$$

for all $n \geq n_0(\delta)$ because $n(|\rho_n|-1)\eta \to \infty$, so $\mathbf{1}\{n(|\rho_n|-1)\eta \leq 0\} = 0$ for all but finitely many n and part (ii) follows since η satisfies (B.6) and $\max\left(\mathbf{1}_{F_n^+}, \mathbf{1}_{F_n^-}\right) \leq \mathbf{1}_{F_n}$.

For part (iii), $m_n\max\left(\mathbf{1}_{\bar{F}_n^-}, \mathbf{1}_{F_n^-}\right) \leq m_n\mathbf{1}\{\hat{\rho}_n < 0\}$; for arbitrary $\delta > 0$

$$\begin{aligned}
\mathbb{P}(m_n\mathbf{1}\{\hat{\rho}_n < 0\} > \delta) &= \mathbb{P}(m_n\mathbf{1}\{\hat{\rho}_n < 0\} > \delta, |\hat{\rho}_n-\rho_n| < 1/2) + \mathbb{P}(|\hat{\rho}_n-\rho_n| \geq 1/2) \\
&\leq \mathbb{P}(m_n\mathbf{1}\{\hat{\rho}_n < 0\} > \delta, \rho_n - 1/2 < \hat{\rho}_n) + \mathbb{P}(|\hat{\rho}_n-\rho_n| \geq 1/2) \\
&\leq \mathbb{P}(m_n\mathbf{1}\{\rho_n < 1/2\} > \delta) + \mathbb{P}(|\hat{\rho}_n-\rho_n| \geq 1/2) \\
&= \mathbb{P}(|\hat{\rho}_n-\rho_n| \geq 1/2)
\end{aligned}$$

for all $n \geq n_0(\delta)$, since $\rho_n \to \rho \geq 1$ so $\mathbf{1}\{\rho_n < 1/2\} = 0$ for all but finitely many n. Part (iii) follows since $|\hat{\rho}_n-\rho_n| \to_p 0$ under Assumption 4 when $\rho_n \to \rho \geq 1$. For part (iv), $m_n\max\left(\mathbf{1}_{\bar{F}_n^+}, \mathbf{1}_{F_n^+}\right) \leq$

$m_n \mathbf{1}\{\hat{\rho}_n \geq 0\}$ and the argument of part (iii) gives

$$\begin{aligned}
\mathbb{P}(m_n \mathbf{1}\{\hat{\rho}_n \geq 0\} > \delta) &\leq \mathbb{P}(m_n \mathbf{1}\{\hat{\rho}_n \geq 0\} > \delta, \hat{\rho}_n < \rho_n + 1/2) + \mathbb{P}(|\hat{\rho}_n - \rho_n| \geq 1/2) \\
&\leq \mathbb{P}(m_n \mathbf{1}\{\rho_n > -1/2\} > \delta) + \mathbb{P}(|\hat{\rho}_n - \rho_n| \geq 1/2) \\
&= \mathbb{P}(|\hat{\rho}_n - \rho_n| \geq 1/2)
\end{aligned}$$

for all $n \geq n_0(\delta)$, since $\rho_n \to \rho \leq -1$ so $\mathbf{1}\{\rho_n > -1/2\} = 0$ for all but finitely many n; part (iv) follows since $|\hat{\rho}_n - \rho_n| \to_p 0$ under Assumption 4 when $\rho_n \to \rho \leq -1$.

Proof of Lemma 3. Using the approximation for r_{3n} in Lemma B2(ii), we may write

$$\begin{aligned}
(1-\rho_n\varphi_{1n})\frac{1}{n}\sum_{t=1}^n x_{t-1}\tilde{z}_{t-1} &= (1-\rho_n\varphi_{1n})\frac{1}{n}\sum_{t=1}^n x_{0t-1}\tilde{z}_{0t-1} + o_p(1) \\
&= \sigma^2 + \frac{1}{n}\sum_{t=1}^n \tilde{z}_{0t-1}u_t + \frac{1}{n}(2\rho_n - 1)\sum_{t=1}^n x_{0t-1}u_t \\
&\quad + \rho_n(\rho_n - 1)\frac{1}{n}\sum_{t=1}^n x_{0t-1}^2 + o_p(1) \qquad \text{(B.9)}
\end{aligned}$$

where the last asymptotic equivalence follows by equations (66)-(68) of Magdalinos and Phillips (2020) (henceforth MP(2020)). Under Assumption 4 on (u_t), $n^{-1}\sum_{t=1}^n \tilde{z}_{0t-1}u_t = \Gamma_n + o_p(1)$ by Lemma 3.1(ii) of MP(2020). Also, under C_+(i), Lemma B2(iii) and $\bar{x}_{0n-1} = O_p(n^{-1/2}\kappa_n)$ give $n^{-1}(1-\rho_n\varphi_{1n})n\bar{z}_{1n-1}\bar{x}_{n-1} = O_p(\kappa_n/n) + O_p\left(n^{-1}(1-\varphi_{1n})^{-1}\right) = o_p(1)$, since $\kappa_n/n \to 0$ under C_+(i). Under C_+(ii), Lemma B2(iii) yields

$$n^{-1}(1-\rho_n\varphi_{1n})n\bar{z}_{1n-1}\bar{x}_{n-1} = \frac{x_{0n}}{n^{1/2}}\frac{1}{n^{3/2}}\sum_{j=1}^n x_{0j-1} + o_p(1). \qquad \text{(B.10)}$$

Combining (B.9)-(B.10) and using $(1-\rho_n^2\varphi_{1n}^2)/(1-\rho_n\varphi_{1n}) \sim 1+\rho_n$, we obtain that

$$\left(1-\rho_n^2\varphi_{1n}^2\right)\frac{1}{n}\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{t-1} = \tilde{\Psi}_n + o_p(1) \qquad \text{(B.11)}$$

with $\tilde{\Psi}_n$ defined in (36) under C_+(i)-C_+(ii), with the term in (B.10) being $o_p(1)$ under C_+(i). Under C_+(i), $n^{-1}\sum_{t=1}^n x_{0t-1}u_t \to_p \Gamma$ by Lemma 2.2(i) of MP(2020), so the identity (obtained from the recursion for x_{0t})

$$\frac{1}{n}\left(1-\rho_n^2\right)\sum_{t=1}^n x_{0t-1}^2 = \frac{1}{n}\sum_{t=1}^n u_t^2 + 2\rho_n\frac{1}{n}\sum_{t=1}^n x_{0t-1}u_t - \frac{1}{n}x_{0n}^2 \to_p \sigma^2 + 2\rho\Gamma \qquad \text{(B.12)}$$

implies that $\tilde{\Psi}_n \to_p \sigma^2 + 2\rho\Gamma$ under C_+(i). Under C_+(ii), $\Gamma_n \to \lambda$ and standard local to unit asymptotics, e.g. Phillips (1987b), yield

$$\begin{aligned}
\tilde{\Psi}_n &= 2\left(\omega^2 + \frac{1}{n}\sum_{t=1}^n x_{0t-1}u_t - \lambda + c\frac{1}{n^2}\sum_{t=1}^n x_{0t-1}^2 - \frac{x_{0n}}{n^{1/2}}\frac{\sum_{j=1}^n x_{0j-1}}{n^{3/2}}\right) + o_p(1) \\
&\to_d 2\left(\omega^2 + \int_0^1 J_c(r)\,dB(r) + c\int_0^1 J_c(r)^2\,dr - J_c(1)\int_0^1 J_c(r)\,dr\right) \\
&= \omega^2 + J_c(1)^2 - 2J_c(1)\int_0^1 J_c(r)\,dr
\end{aligned}$$

where the last equality holds by applying the integration by parts formula to the stochastic integral $\int_0^1 J_c(r)\,dB(r)$; see equation (79) of MP(2020). The expression for the weak limit $\tilde{\Psi}_c$ in the lemma follows since $\sigma^2 + 2\rho\Gamma = \omega^2$ under C_+(ii), completing the proof of part (i). For part (ii), by the approximation for r_{2n} in Lemma B2(ii), it is enough to show that $\tilde{v}_{0n} := (1-\rho_n^2\varphi_{1n}^2)n^{-1}\sum_{t=1}^n \tilde{z}_{0t}^2 \to_p \sigma^2 + 2\rho\Gamma$. The proof of Lemma 3.1(iv) of MP(2020) shows that $\tilde{v}_{0n} = (1-\varphi_{1n}^2)n^{-1}\sum_{t=1}^n z_{1t}^2 + o_p(1) \to_p \omega^2$ when $(1-\varphi_{1n})\kappa_n \to \infty$ and $\tilde{v}_{0n} = (1-\rho_n^2)n^{-1}\sum_{t=1}^n x_{0t}^2 + o_p(1)$ when $(1-\varphi_{1n})\kappa_n \to 0$. In both cases, $\tilde{v}_{0n} \to_p \sigma^2 + 2\rho\Gamma$; by (B.12) when $(1-\varphi_{1n})\kappa_n \to 0$ and the fact that $(1-\varphi_{1n})\kappa_n \to \infty$ implies that $\rho = 1$ and $\sigma^2 + 2\rho\Gamma = \omega^2$. It remains to show that that $(1-\rho_n^2\varphi_{1n}^2)\,n^{-1}\sum_{t=1}^n \tilde{z}_{0t}^2 \to_p \omega^2$ when $(1-\varphi_{1n})\,/\,(1-\rho_n) \to \phi \in (0,\infty)$: in this case, (ρ_n) belongs to C$_+$(i) and equations (74) and (75) of MP(2020) imply that

$$\frac{1}{n}\left(1-\rho_n^2\varphi_{1n}^2\right)\sum_{t=1}^n \tilde{z}_{0t-1}^2 = \frac{1-\rho_n^2\varphi_{1n}^2}{1-\varphi_{1n}^2}\left(\omega^2 - 2\frac{1-\rho_n}{1-\rho_n^2\varphi_{1n}^2}\left(1-\rho_n^2\varphi_{1n}^2\right)\frac{1}{n}\sum_{t=1}^n x_{0t-1}\tilde{z}_{0t-1}\right) + o_p(1).$$

Since $n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)\sum_{t=1}^n x_{0t-1}\tilde{z}_{0t-1} \to_p \omega^2$, the result follows from

$$\frac{1-\rho_n^2\varphi_{1n}^2}{1-\varphi_{1n}^2}\left(1-\frac{2(1-\rho_n)}{1-\rho_n^2\varphi_{1n}^2}\right) \sim \frac{2\rho_n}{1-\varphi_{1n}^2}(1-\varphi_{1n}) \to 1$$

since $\varphi_{1n}\to 1$ and $\rho_n \to 1$. For part (iii), in view of the approximation for r_{1n} in Lemma B2(ii), it is sufficient to show the result for $\sum_{t=1}^n \xi_{nt}$ with $\xi_{nt} := n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{1/2}\tilde{z}_{0t-1}e_t$. Since ξ_{nt} is an $\mathcal{F}_t$-martingale array under Assumption 4 that satisfies the Lindeberg condition by Lemma 3.2 of MP(2020) and

$$\sum_{t=1}^n \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\xi_{nt}^2\right) = \sigma_e^2\left(1-\rho_n^2\varphi_{1n}^2\right)\frac{1}{n}\sum_{t=1}^n \tilde{z}_{0t-1}^2 \to_p \sigma_e^2\left(\sigma^2+2\rho\Gamma\right)$$

by part (ii) of the lemma, the result follows by a standard martingale central limit theorem (e.g. Corollary 3.1 of Hall and Heyde (1980)).

Proof of Lemma 4. The statement for $\left[\sum_{t=1}^n z_{2t-1}u_t, \sum_{t=1}^n z_{2t-1}^2\right]$ and $\left[\check{Y}_n, \check{Z}_n\right] \to_d [Y,Z]$ follow by Lemma 5 and Lemma 2 of Magdalinos (2012). Hence $[Y_n, Z_n] \to_d [Y,Z]$ of part (i) follows from the martingale approximation of Lemma B2(v). The only statement of part (i) that requires proof is for $s_n^{-1}\sum_{t=1}^n x_{t-1}z_{2t-1}$. The recursions for x_t and z_t in (5) and (33) give

$$\begin{aligned}(\rho_n\varphi_{2n}-1)\sum_{t=1}^n x_{t-1}z_{2t-1} &= x_nz_{2n} - \varphi_{2n}\sum_{t=1}^n z_{2t-1}u_t - \rho_n\sum_{t=1}^n x_{t-1}u_t - \sum_{t=1}^n u_t^2 \\ &\quad +\varphi_{2n}\mu(\rho_n-1)\sum_{t=1}^n z_{2t-1} + \mu(\rho_n-1)\sum_{t=1}^n u_t \\ &= x_nz_{2n} + o_p(\nu_n\nu_{n,z}) \end{aligned} \tag{B.13}$$

where the order of magnitude follows from: $\nu_n^{-1}(\rho_n-1)\sum_{t=1}^n u_t$ is of order $O_p(\rho_n^{-n}(n(\rho_n-1))^{1/2}) = o_p(1)$ by Lemma B2(i) under C$_+$(iii) and $O_p(n^{-1})$ under C$_+$(ii); $\nu_n^{-1}\nu_{n,z}^{-1}\sum_{t=1}^n z_{2t-1}u_t$ is of order $O_p[\nu_n^{-1}(\varphi_{2n}-1)^{-1/2}] = o_p(\rho_n^{-n}[n(\rho_n-1)]^{1/2}) = o_p(1)$ under C$_+$(iii) (by Lemma B2(i)) and $O_p[(n(\varphi_{2n}-1))^{-1/2}]$ under C$_+$(ii); by (B.3) $\nu_n^{-1}\nu_{n,z}^{-1}\sum_{t=1}^n x_{t-1}u_t = O_p(\varphi_{2n}^{-n}[n(\varphi_{2n}-1)]^{1/2}) = o_p(1)$ by Lemma B2(i); finally, the recursion (33) gives

$$\nu_n^{-1}\nu_{n,z}^{-1}(\rho_n-1)\sum_{t=1}^n z_{2t-1} = \nu_n^{-1}\nu_{n,z}^{-1}(\rho_n-1)(\varphi_{2n}-1)^{-1}\left(z_{2n}-\sum_{t=1}^n u_t\right) = o_p\left(n^{-1/2}\right) + o_p\left(\varphi_{2n}^{-n}\right)$$

since $z_{2n} = O_p(\nu_{n,z})$ and $\sum_{t=1}^n u_t = O_p\left(n^{1/2}\right)$. The proof of (B.13) follows by Lemma B2(i) and the fact that $n(\varphi_{2n}-1)\to\infty$. By (B.13), we conclude that $s_n^{-1}\sum_{t=1}^n x_{t-1}z_{2t-1} = \frac{x_n}{\nu_n}\frac{z_{2n}}{\nu_{n,z}} + o_p(1)$ and the result follows from the definitions of Z_n and X_n in (39) and (40).

For part (ii), the martingale approximation of Lemma B2(v) implies that

$$[Y_n, X_n]' = C(1)\sum_{j=1}^n c_{nj}e_j + o_p(1) \quad \text{with } c_{nj} = \left[\left(\varphi_{2n}^2-1\right)^{1/2}\varphi_{2n}^{-(n-j)-1}, \left(\rho_n^2-1\right)^{1/2}\rho_n^{-j}\right]' \tag{B.14}$$

(for X_n we may use the part of Lemma B2(v) corresponding to Z_n replacing φ_{2n} with the mildly explosive root ρ_n). We apply a standard martingale central limit theorem, e.g. Corollary 3.1 of Hall and Heyde (1980), to the martingale array in (B.14): the conditional variance matrix $V_n = \sum_{j=1}^n c_{nj}c_{nj}'\mathbb{E}_{\mathcal{F}_{j-1}}\left(u_j^2\right)$ has typical elements: $V_{11}^{(n)} = \omega^2(\varphi_{2n}^2-1)\sum_{j=1}^n \varphi_{2n}^{-2j} \to \omega^2$; $V_{22}^{(n)} = \omega^2(\rho_n^2-1)\sum_{j=1}^n \rho_n^{-2j} \to \omega^2$; $V_{12}^{(n)} = (\rho_n^2-1)^{1/2}(\varphi_{2n}^2-1)^{1/2}\varphi_{2n}^{-n-1}\sum_{j=1}^n\left(\frac{\varphi_{2n}}{\rho_n}\right)^j$. When $n|\rho_n-\varphi_{2n}|\to\infty$, evaluating the geometric progression yields $V_{12}^{(n)} = O\left(\varphi_{2n}^{-n}\right) + O\left(\rho_n^{-n}\right)$; when $|\rho_n-\varphi_{2n}| = O(n^{-1})$, $\sum_{j=1}^n(\varphi_{2n}/\rho_n)^j \le bn$ for some $b>0$ and $\left|V_{12}^{(n)}\right| \le bn\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n-1} = o(1)$

by Lemma B2(i). In both cases $V_{12}^{(n)} \to 0$ so $V_n \to \omega^2 I_2$ as required for the covariance matrix of a random vector $[Y, X]'$ consisting of independent $\mathcal{N}(0, \omega^2)$ variates. For the Lindeberg condition associated with (B.14), the bound $\max_{j\le n} \|c_{nj}\|^2 \le 2\lambda_n$ with $\lambda_n = (\varphi_{2n}^2 - 1) \vee (\rho_n^2 - 1)$ yields

$$\sum_{j=1}^n \|c_{nj}\|^2 \mathbb{E}\left(e_t^2 \mathbf{1}\left\{\|c_{nj}\|^2 e_j^2 > \delta\right\}\right) \le \max_{j\le n} \mathbb{E}\left(e_j^2 \mathbf{1}\left\{e_j^2 > \lambda_n^{-1}\delta/2\right\}\right) \sum_{j=1}^n \|c_{nj}\|^2 \to 0$$

by uniform integrability of $\left(e_j^2\right)_{j\in\mathbb{N}}$, since $\lambda_n^{-1} \to \infty$ when $\rho_n \to 1$ and $\sum_{t=1}^n \|c_{nt}\|^2 = O(1)$.

For part (iv), $\rho_n \to \rho > 1$; X_∞ in (7) is well-defined $a.s.$ because $\eta_n := \sum_{j=1}^n \rho^{-j} u_j$ converges $a.s.$ under Assumption 4: $\mathbb{P}\left(\sup_{k\ge1} \|\eta_{n+k} - \eta_n\|_{L_1} > \delta\right) \le \delta^{-1}\mathbb{E}\left(\sup_{k\ge1} \|\eta_{n+k} - \eta_n\|_{L_1}\right)$ for any $\delta > 0$ and $\mathbb{E}\left(\sup_{k\ge1} \|\eta_{n+k} - \eta_n\|_{L_1}\right) \le \sup_{k\ge1} \|u_k\|_{L_1} \sum_{j=n+1}^\infty |\rho|^{-j} \to 0$. By the above convergence of X_∞ and since $X_0(n) \to_p X_0$ when $\rho_n \to \rho > 1$ by Assumption 3, $X_n \to_p X_\infty$ will follow from showing that $\sum_{j=1}^n (\rho_n^{-j} - \rho^{-j}) u_j \to_{L_1} 0$, which, in turn, will follow from showing that $\sum_{j=1}^n |\rho_n^{-j} - \rho^{-j}| \to 0$. To prove the last statement, we apply the mean value theorem to the function $x \mapsto x^{-j}$: $\rho_n^{-j} - \rho^{-j} = -(\rho_n - \rho) j\phi_n^{-j-1}$ for some $\phi_n \to \rho$; hence, we may choose $\delta \in (0, \rho-1)$ and $n_0(\delta) \in \mathbb{N}$ such that for all $n \ge n_0(\delta)$: $\phi_n > \rho - \delta$ which implies that $\sum_{j=1}^n |\rho_n^{-j} - \rho^{-j}| = |\rho_n - \rho| \sum_{j=1}^n j\phi_n^{-j-1} \le |\rho_n - \rho| \sum_{j=1}^\infty j(\rho-\delta)^{-j-1} \to 0$ since $\rho - \delta > 1$ from the choice $\delta \in (0, \rho-1)$. Next we show that $X_\infty \ne 0$ $a.s.$ under Assumption 4. Writing $X_\infty = \lim_{n\to\infty} \sum_{i=1}^n \pi_i e_i + Y_0$ $a.s.$, where $\pi_i = (\rho^2-1)^{-1/2}\left(\sum_{j=0}^\infty \rho^{-j} c_j\right)\rho^{-i}$ and $Y_0 = (\rho^2-1)^{-1/2}\left(\sum_{i=0}^\infty \left(\sum_{j=1}^\infty \rho^{-j} c_{j+i}\right) e_{-i} + X_0 - \mu\right)$ satisfy $\sum_{i=1}^\infty |\pi_i| < \infty$ and $\pi_i \ne 0$ for all i by Assumption 4 and is an $\mathcal{F}_0$-measurable random variable by Assumptions 3 and 4 (under Assumption 2, $\pi_i = (\rho^2-1)^{-1/2}\rho^{-i}$ and $Y_0 = (\rho^2-1)^{-1/2}(X_0 - \mu)$). By $\mathbb{E}_{\mathcal{F}_{t-1}}(e_t^2) = \sigma^2$ and $\liminf_{t\to\infty} \mathbb{E}_{\mathcal{F}_{t-1}} |e_t| > 0$ $a.s.$, the martingale difference sequence $(e_t, \mathcal{F}_t)_{t\in\mathbb{N}}$ satisfies the local Marcinkiewicz-Zygmund conditions (equation (1.1) of Lai and Wei (1983)), so applying Corollary 2 of Lai and Wei (1983) to X_∞ yields $\mathbb{P}(X_\infty = 0) = \mathbb{P}(\lim_{n\to\infty} \sum_{i=1}^n \pi_i e_i = -Y_0) = 0$.

We turn to the limit distribution of $g(X_n) Y_n$. Let $(k_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ be an increasing sequence satisfying $k_n/n \to 0$ and $k_n/(\varphi_{2n}^2 - 1)^{-1} \to \infty$ and let $Y_n' = (\varphi_{2n}^2 - 1)^{1/2} C(1) \sum_{t=k_n}^n \varphi_{2n}^{-(n-t)-1} e_t$. It is easy to see that $\left\|\check{Y}_n - Y_n'\right\|_{L_2} = O\left(\varphi_{2n}^{-k_n}\right) = o(1)$ so Lemma B2(v) implies that $\|Y_n - Y_n'\| = o_p(1)$. Also,

$$|X_n - X_{k_n-1}| \le \left(\rho_n^2 - 1\right)^{1/2}\left(\sum_{j=k_n}^n \rho_n^{-j} u_j + X_0(n) - X_0(k_n - 1)\right) \to_p 0$$

by Assumption 3. Using the fact that $X_\infty \ne 0$ $a.s.$ and the continuity of g away from zero, $|g(X_n) - g(X_{k_n-1})| \to_p |g(X_\infty) - g(X_\infty)| = 0$, so we conclude that

$$g(X_n) Y_n = g(X_{k_n-1}) Y_n' + o_p(1) = \sum_{t=0}^{n-k_n} \xi_{n,t} + o_p(1) \tag{B.15}$$

where $\xi_{n,t} = C(1)\zeta_{nt} e_{t+k_n}$ and $\zeta_{nt} = (\varphi_{2n}^2 - 1)^{1/2} \varphi_{2n}^{-(n-k_n-t)-1} g(X_{k_n-1})$. Since ζ_{nt} is an $\mathcal{F}_{k_n-1}$-measurable sequence for all n, t, $\left\{\left(\xi_{n,t}, \mathcal{F}_{n,t}\right) : 0 \le t \le n - k_n\right\}$ with $\mathcal{F}_{n,t} = \mathcal{F}_{t+k_n}$ is a martingale difference array with $\mathcal{F}_{n,t} \subseteq \mathcal{F}_{n+1,t}$ since the sequence $(k_n)_{n\in\mathbb{N}}$ was chosen to be increasing. We apply a martingale central limit theorem (Corollary 3.1 of Hall and Heyde (1980)) to a mixed Gaussian distribution. The conditional variance of the martingale array in (B.15) is given by

$$\begin{aligned} \sum_{t=0}^{n-k_n} \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\xi_{n,t}^2\right) &= \omega^2 \sum_{t=0}^{n-k_n} \zeta_{nt}^2 = \omega^2 g^2(X_{k_n-1})\left(\varphi_{2n}^2 - 1\right) \sum_{t=0}^{n-k_n} \varphi_{2n}^{-2(n-k_n-t+1)} \\ &\to_p \omega^2 g(X_\infty)^2 . \end{aligned} \tag{B.16}$$

For the Lindeberg condition, $L_n(\delta) := \sum_{t=0}^{n-k_n} \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\xi_{n,t}^2 \mathbf{1}\left\{\xi_{n,t}^2 > \delta\right\}\right) \to_p 0$ for all $\delta > 0$, let

$\lambda_n(\delta) := C(1)^{-1}(\varphi_{2n}^2 - 1)^{-1/2}\delta^{1/2}$ and note that $\lambda_n(\delta) \to \infty$ for any $\delta > 0$. The estimation

$$\mathbf{1}\left\{\xi_{n,t}^2 > \delta\right\} \leq \mathbf{1}\left\{g^2(X_{k_n-1})e_{t+k_n}^2 > \lambda_n(\delta)^2\right\} \leq \mathbf{1}\left\{g^2(X_{k_n-1}) > \lambda_n(\delta)\right\} + \mathbf{1}\left\{e_{t+k_n}^2 > \lambda_n(\delta)\right\}$$

and $\mathcal{F}_{k_n-1}$-measurability of X_{k_n-1} imply that $L_n(\delta) \leq L_{1n}(\delta) + g^2(X_{k_n-1})L_{2n}(\delta)$, where

$$L_{2n}(\delta) = C(1)^2\left(\varphi_{2n}^2 - 1\right)\sum_{t=k_n}^n \varphi_{2n}^{-2(n-t+1)}\mathbb{E}_{\mathcal{F}_{t-1}}\left(e_t^2\mathbf{1}\left\{e_t^2 > \lambda_n(\delta)\right\}\right) \to_{L_1} 0$$

since $\|L_{2n}(\delta)\|_{L_1} \leq O(1)\max_{t\leq n}\mathbb{E}\left(e_t^2\mathbf{1}\left\{e_t^2 > \lambda_n(\delta)\right\}\right) \to 0$ by UI of (e_t^2) and

$$L_{1n}(\delta) = \omega^2\mathbf{1}\left\{g^2(X_{k_n-1}) > \lambda_n(\delta)\right\}\sum_{t=0}^{n-k_n}\zeta_{nt}^2 \to_p 0$$

since both $g^2(X_{k_n-1})$ and $\sum_{t=0}^{n-k_n}\zeta_{nt}^2$ converge in probability to $g^2(X_\infty) < \infty$ $a.s.$ and $\lambda_n(\delta) \to \infty$. We conclude that, for any $\delta > 0$, $L_n(\delta) \leq o_p(1) + g^2(X_{k_n-1})o_p(1) = o_p(1)$ proving the Lindeberg condition. In view of (B.16), the martingale central limit theorem applied to $\sum_{t=0}^{n-k_n}\xi_{n,t}$ in (B.15) then implies that $g(X_n)Y_n \to_d \psi$ where ψ has characteristic function $\phi_\psi(x) = \exp\left\{-\frac{1}{2}t^2\sigma^2 g(X_\infty)^2\right\}$ i.e. $\psi =_d \mathcal{MN}\left(0, \sigma^2 g(X_\infty)^2\right)$. The statement for $g(X_n)Y_n^\varepsilon$ follows by an identical argument by replacing $C(1)e_t$ by ε_t in $\check{Y}_n$.

Proof of Lemma 5. Denote $\xi_{nt} = \left[\xi_{1,nt}, \xi_{2,nt}, \xi_{3,nt}\right]'$ with $\xi_{1,nt} = \left(n\left(1-\varphi_{1n}^2\right)^{-1}\right)^{-1/2} z_{1t-1}e_t$, $\xi_{2,nt} = C(1)n^{-1/2}e_t$ and $\xi_{3,nt} = C(1)\left(\varphi_{2n}^2 - 1\right)^{1/2}\varphi_{2n}^{-(\lfloor ns\rfloor - t)-1}e_t$. The martingale approximation of Lemma B2(v) for $Y_n(s)$ and a standard approximation for $B_n(s)$ give

$$[U_n(s), B_n(s), Y_n(s)]' = \sum_{t=1}^{\lfloor ns\rfloor}\xi_{nt} + o_p(1). \tag{B.17}$$

Since z_{1t-1} is $\mathcal{F}_{t-1}$-measurable, ξ_{nt} is a $\mathcal{F}_t$-martingale difference array and we may apply a Lindeberg-type functional CLT for vector-valued martingale difference arrays to (B.17): see Theorem 3.33 (pp. 478) of Jacod and Shiryaev (2003). The conditional Lindeberg condition on $\|\xi_{nt}\|^2$ (3.31 in Jacod and Shiryaev (2003)) is implied by the stronger unconditional Lindeberg condition (LC) on $\|\xi_{nt}\|^2$ which, in turn, is implied by establishing the LC on each of $\xi_{1,nt}^2$, $\xi_{2,nt}^2$ and $\xi_{3,nt}^2$. The LC for $\xi_{1,nt}^2$ is established by Proposition A1 and Lemma 3.3 of MP(2020). The LC for $\xi_{2,nt}^2$ follows from the bound $\sum_{t=1}^{\lfloor ns\rfloor}\mathbb{E}\left(\xi_{2,nt}^2\mathbf{1}\left\{\xi_{2,nt}^2 > \delta\right\}\right) \leq C(1)^2\max_{t\leq n}\mathbb{E}\left(e_t^2\mathbf{1}\left\{e_t^2 > n\delta C(1)^{-2}\right\}\right)$ and uniform integrability of $(e_t^2)_{t\in\mathbb{N}}$. For the LC for $\xi_{3,nt}^2$, $\varphi_{2n}^{-2(\lfloor ns\rfloor - t+1)} \leq 1$ for all $t \leq \lfloor ns\rfloor$ and $s \in [0,1]$ implies that

$$\sum_{t=1}^{\lfloor ns\rfloor}\mathbb{E}\left(\xi_{3,nt}^2\mathbf{1}\left\{\xi_{3,nt}^2 > \delta\right\}\right) \leq C(1)^2\max_{t\leq n}\mathbb{E}\left(e_t^2\mathbf{1}\left\{e_t^2 > \lambda_n(\delta)^2\right\}\right)\left(\varphi_{2n}^2 - 1\right)\sum_{t=1}^n\varphi_{2n}^{-2t} \tag{B.18}$$

where $\lambda_n(\delta) = C(1)^{-1}\left(\varphi_{2n}^2 - 1\right)^{-1/2}\delta^{1/2} \to \infty$ for any $\delta > 0$. Since $\left(\varphi_{2n}^2 - 1\right)\sum_{t=1}^n\varphi_{2n}^{-2t} = O(1)$, $(e_t^2)_{t\in\mathbb{N}}$ is a UI sequence and $\lambda_n(\delta)^2 \to \infty$, the right side of (B.18) is $o(1)$. This establishes the LC for the martingale difference array ξ_{nt} in (B.17). The conditional variance matrix of the array in (B.17) is given by $V^{(n)} := \sum_{t=1}^{\lfloor ns\rfloor}\mathbb{E}_{\mathcal{F}_{t-1}}\left(\xi_{nt}\xi_{nt}'\right)$; denoting the typical elements of $V^{(n)}$ by $\left[V_{ij}^{(n)}\right]_{i,j=1}^3$: $V_{11}^{(n)} = \sigma_\varepsilon^2\frac{1}{n}\left(1-\varphi_{1n}^2\right)\sum_{t=1}^{\lfloor ns\rfloor}z_{1t-1}^2 \to_p \sigma_e^2\omega^2 s$, by Lemma 3.1(iv) of MP(2020); $V_{22}^{(n)} = \omega^2\lfloor ns\rfloor/n \to \omega^2 s$; $V_{33}^{(n)} = \omega^2\left(\varphi_{2n}^2 - 1\right)\sum_{t=1}^{\lfloor ns\rfloor}\varphi_{2n}^{-2t} \to \omega^2$ for all $s > 0$; $V_{23}^{(n)} = \omega^2 n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{1/2}\sum_{t=1}^{\lfloor ns\rfloor}\varphi_{2n}^{-t} = O\left(\left[n\left(1-\varphi_{2n}\right)\right]^{-1/2}\right) = o(1)$; since $\sum_{t=1}^{\lfloor ns\rfloor}z_{1t-1} = O_p\left(n^{1/2}\left(1-\varphi_{1n}^2\right)^{-1}\right)$, $V_{12}^{(n)} = \omega^2\frac{1}{n}\left(1-\varphi_{1n}^2\right)^{1/2}\sum_{t=1}^{\lfloor ns\rfloor}z_{1t-1} = O_p\left(\left(n\left(1-\varphi_{1n}^2\right)\right)^{-1/2}\right) = o_p(1)$; $V_{13}^{(n)} = \omega^2\left(\varphi_{2n}^2 - 1\right)^{1/2}\left(1-\varphi_{1n}^2\right)^{1/2}n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor}\varphi_{2n}^{-(\lfloor ns\rfloor - t+1)}z_{1t-1}$ satisfies

$$\left\|V_{13}^{(n)}\right\|_{L_1} \leq C(1)\sigma_e^2\max_{t\leq n}\left\|\left(1-\varphi_{1n}^2\right)^{1/2}z_{1t}\right\|_{L_2}\left(\varphi_{2n}^2 - 1\right)^{1/2}n^{-1/2}\sum_{t=1}^n\varphi_{2n}^{-t} = O\left(\left(n\left(\varphi_{2n} - 1\right)\right)^{-1/2}\right).$$

We conclude that $V^{(n)} \to_p diag\left(\sigma_e^2\omega^2 s, \omega^2 s, \omega^2\right)$ for $s \in [0,1]$, and applying Theorem 3.33 of Jacod and Shiryaev (2003) to (B.17), $\sum_{t=1}^{\lfloor ns \rfloor} \xi_{nt} \Rightarrow \xi(s)$ where $\xi(s)$ is a continuous Gaussian martingale with quadratic variation $\langle \xi \rangle_s = diag\left(\sigma_e^2\omega^2 s, \omega^2 s, \omega^2\right)$. By Levy's characterisation (e.g. Theorem 4.4 II of Jacod and Shiryaev (2003), $\xi(s)$ is characterised by its quadratic variation process, $\xi(s) =_d [U(s), B(s), Y]'$ with the right side defined in the statement of the lemma and independence between the components of $\xi(s)$ implied by the diagonality of the quadratic variation matrix $\langle \xi \rangle_s$.

Proof of Lemma 6. First, we prove (41) and (42). Since $\rho_n, \varphi_{1n}^-, \varphi_{2n}^- < 0$

$$\begin{aligned} x_t^+ &= (-1)^t x_{0t} = -\rho_n (-1)^{t-1} x_{0t-1} + (-1)^t u_t = |\rho_n| x_{t-1}^+ + (-1)^t u_t \\ z_{2t}^+ &= (-1)^t z_{2t}^- = \left|\varphi_{2n}^-\right| z_{2t-1}^+ + (-1)^t u_t \end{aligned}$$

and

$$\tilde{z}_{1t}^+ = (-1)^t \tilde{z}_{1t}^- = \left|\varphi_{1n}^-\right| (-1)^t \tilde{z}_{1t-1}^- + (-1)^t (x_t + x_{t-1}) = \left|\varphi_{1n}^-\right| \tilde{z}_{1t-1}^+ + \Delta x_t^+$$

so the recursions for x_t^+ in (41) and $\tilde{z}_{1t}^+, z_{2t}^+$ are satisfied. The identity $\hat{\rho}_n^+ = -\hat{\rho}_n$ implies that

$$\begin{aligned} F_n^- &= \{-n(\hat{\rho}_n + 1) \leq 0\} \cap \{\hat{\rho}_n < 0\} = \left\{-n\left(-\hat{\rho}_n^+ + 1\right) \leq 0\right\} \cap \left\{\hat{\rho}_n^+ > 0\right\} \\ &= \left\{n\left(\hat{\rho}_n^+ - 1\right) \leq 0\right\} \cap \left\{\hat{\rho}_n^+ > 0\right\} = F_n^{++} \end{aligned}$$

and $\bar{F}_n^- = \left\{n\left(\hat{\rho}_n^+ - 1\right) > 0\right\} \cap \left\{\hat{\rho}_n^+ > 0\right\} = \bar{F}_n^{++}$. By summing the recursions for $\tilde{z}_{1t}^-$ and $\tilde{z}_{2t}^-$ in (19)-(20), we obtain the identities

$$\bar{z}_{1n-1}^- = \left(1 - \varphi_{1n}^-\right)^{-1} \frac{1}{n} \left(\sum_{t=1}^n \nabla \underline{x}_t - \tilde{z}_{1n}^-\right) = -\left(\frac{1}{2} + o(1)\right) \frac{1}{n} \tilde{z}_{1n}^- \tag{B.19}$$

$$\bar{z}_{2n-1}^- = \left(1 - \varphi_{2n}^-\right)^{-1} \frac{1}{n} \left(\sum_{t=1}^n \hat{u}_t - \tilde{z}_{2n}^-\right) = -\left(\frac{1}{2} + o(1)\right) \frac{1}{n} \tilde{z}_{2n}^- \tag{B.20}$$

because $\sum_{t=1}^n \nabla \underline{x}_t = 0$, $\sum_{t=1}^n \hat{u}_t = 0$ and $1 - \varphi_{1n}^-$ and $1 - \varphi_{2n}^-$ converge to 2. Write

$$\pi_n^{-2} \sum_{t=1}^n \underline{x}_{t-1} \tilde{z}_{1t-1}^- = \pi_n^{-2} \sum_{t=1}^n x_{0t-1} \tilde{z}_{0t-1}^- - n\pi_n^{-2} \bar{x}_{n-1} \bar{z}_{1n-1}^- + r_{3n}^- \tag{B.21}$$

where r_{3n}^-, defined in Lemma B1(ii), satisfies $r_{3n}^- = o_p(1)$ under C_(i)-C_(ii).Together with part (i) of Lemma B1, (B.19) implies that the second term of (B.21) satisfies: $n\pi_n^{-2}\bar{x}_{n-1}\bar{z}_{1n-1}^- \mathbf{1}_{F_n^-} = O_p\left(\pi_n^{-2}\left|\tilde{z}_{1n}^-\right|\right) \mathbf{1}_{F_n^-} = O_p\left(n^{-1/2}\right)$ under C_(i)-C_(ii) and $\pi_n^{-2}\left\|\left|\tilde{z}_{1n}^-\right|\right\|_{L_1} \mathbf{1}_{F_n^-} \to_p 0$ under C_(iii) by Lemma 2, so that

$$\pi_n^{-2} \left| \sum_{t=1}^n \underline{x}_{t-1} \tilde{z}_{1t-1}^- - \sum_{t=1}^n x_{0t-1} \tilde{z}_{0t-1}^- \right| \mathbf{1}_{F_n^-} \to_p 0 \tag{B.22}$$

C_(i)-C_(iii). By Lemma B1(ii), $r_{1n}^- = \pi_n^{-1} \sum_{t=1}^n \left(\tilde{z}_{1t-1}^- - \tilde{z}_{0t-1}^-\right) u_t = o_p(1)$ and $\pi_n^{-1} n\bar{u}_n \bar{z}_{1n-1}^- = o_p\left(\pi_n^{-1}\right)$ by (B.19) under C_(i)-C_(ii); hence

$$\pi_n^{-1} \left\| \sum_{t=1}^n \tilde{z}_{1t-1}^- u_t - n\bar{u}_n \bar{z}_{1n-1}^- - \sum_{t=1}^n \tilde{z}_{0t-1}^- u_t \right\| \mathbf{1}_{F_n^-} \to_p 0 \tag{B.23}$$

under C_(i)-C_(iii), with the claim under C_(iii) following by Lemma 2(i). Combining (B.22) and (B.23), we obtain under C_(i)-C_(ii),

$$\begin{aligned} \pi_n \left(\tilde{\rho}_{1n}^- - \rho_n\right) \mathbf{1}_{F_n^-} &= \frac{\pi_n^{-1} \sum_{t=1}^n \tilde{z}_{0t-1}^- u_t}{\pi_n^{-2} \sum_{t=1}^n x_{0t-1} \tilde{z}_{0t-1}^-} \mathbf{1}_{F_n^-} + o_p(1) = \frac{-\pi_n^{-1} \sum_{t=1}^n \tilde{z}_{0t-1}^+ (-1)^t u_t}{\pi_n^{-2} \sum_{t=1}^n x_{t-1}^+ \tilde{z}_{0t-1}^+} \mathbf{1}_{F_n^{++}} + o_p(1) \\ &= -\pi_n \left(\tilde{\rho}_{1n}^+ - |\rho_n|\right) \mathbf{1}_{F_n^{++}} + o_p(1). \end{aligned} \tag{B.24}$$

Since, by Lemma 2, both sides of (B.24) are $o_p(1)$ under C_(iii), (B.24) holds under C_(i)-C_(iii).

Lemma B1(ii) and (B.20) imply that $ns_n^{-1}\bar{x}_{n-1}\bar{z}_{2n-1}^{-}\mathbf{1}_{\bar{F}_n^-} = o_p\left(s_n^{-1}\left|\bar{z}_{2n}^{+}\right|\right)\mathbf{1}_{\bar{F}_n^-} = o_p(1)$ since $s_n^{-1}\bar{z}_{2n}^{+} = o_p\left(n^{-1/2}\right)$ under C$_-$(ii), $s_n^{-1}\bar{z}_{2n}^{+} = O_p\left(|\rho_n|^{-n}\right)$ under C$_-$(iii) and $s_n^{-1}\left\|\bar{z}_{2n}^{+}\right\|_{L_1}\mathbf{1}_{\bar{F}_n^-} \rightarrow_p 0$ under C$_-$(i) by Lemma 2. Using the above and $R_{3n}^{-} = o_p(1)$ by Lemma B2(iv), we obtain

$$s_n^{-1}\sum_{t=1}^{n}\underline{x}_{t-1}\tilde{z}_{2t-1}^{-}\mathbf{1}_{\bar{F}_n^-} = s_n^{-1}\sum_{t=1}^{n}x_{0t-1}z_{2t-1}^{-}\mathbf{1}_{\bar{F}_n^-} + o_p(1) \tag{B.25}$$

under C$_-$(i)-C$_-$(iii). By (B.20), $n(\left|\varphi_{2n}^{-}\right|^2-1)\left|\varphi_{2n}^{-}\right|^{-n}\bar{u}_n\bar{z}_{2n-1}^{-}\mathbf{1}_{\bar{F}_n^-} = O_p(n^{-1/2}((\left|\varphi_{2n}^{-}\right|^2-1)\left|\varphi_{2n}^{-}\right|^{-n}\bar{z}_{2n-1}^{-})\mathbf{1}_{\bar{F}_n^-}$ is $o_p\left(n^{-1/2}\right)$ under C_(ii)-C_(iii) and $o_p(1)$ under C_(i) since $\left\|\bar{z}_{2n-1}^{-}\right\|_{L_1}\mathbf{1}_{\bar{F}_n^-} \rightarrow_p 0$ by Lemma 2; the above and $R_{2n}^{-} = o_p(1)$ by Lemma B2(iv) imply that, under C_(i)-C_(iii),

$$\left(\varphi_{2n}^2 - 1\right)\varphi_{2n}^{-n}\left\|\sum_{t=1}^{n}\tilde{z}_{2t-1}^{-}u_t - n\bar{u}_n\bar{z}_{2n-1}^{-} - \sum_{t=1}^{n}z_{2t-1}^{-}u_t\right\|\mathbf{1}_{\bar{F}_n^-} \rightarrow_p 0. \tag{B.26}$$

Combining (B.25) and (B.26), since $(\left|\varphi_{2n}^{-}\right|^2 - 1)\left|\varphi_{2n}^{-}\right|^{-n}/s_n^{-1} = \pi_n$ under C_(ii)-C_(iii),

$$\begin{aligned}
\pi_n\left(\tilde{\rho}_{2n}^{-} - \rho_n\right)\mathbf{1}_{\bar{F}_n^-} &= \frac{\left(\varphi_{2n}^2 - 1\right)\varphi_{2n}^{-n}\sum_{t=1}^{n}z_{2t-1}^{-}u_t}{s_n^{-1}\sum_{t=1}^{n}x_{0t-1}z_{2t-1}^{-}}\mathbf{1}_{\bar{F}_n^-} + o_p(1)\\
&= \frac{-\left(\varphi_{2n}^2 - 1\right)\varphi_{2n}^{-n}\sum_{t=1}^{n}z_{2t-1}^{+}(-1)^t u_t}{s_n^{-1}\sum_{t=1}^{n}x_{t-1}^{+}z_{2t-1}^{+}}\mathbf{1}_{\bar{F}_n^{++}} + o_p(1)\\
&= -\pi_n\left(\tilde{\rho}_{2n}^{+} - |\rho_n|\right)\mathbf{1}_{\bar{F}_n^{++}} + o_p(1).
\end{aligned} \tag{B.27}$$

Since, by Lemma 2, both sides of (B.27) are $o_p(1)$ under C_(i), (B.27) holds under C_(i)-C_(iii).

Since $\tilde{\beta}_{jn}^{-} - \beta$ has the same denominator as $\tilde{\rho}_{jn}^{-} - \rho_n$ for $j \in \{1,2\}$, it is enough to examine the numerators $\sum_{t=1}^{n}\tilde{z}_{jt-1}^{-}\varepsilon_t - n\bar{\varepsilon}_n\bar{z}_{jn-1}^{-}$: these can immediately be seen to satisfy (B.23) and (B.26) with $(u_t, \bar{u}_n)$ replaced by $(\varepsilon_t, \bar{\varepsilon}_n)$. Since, under the restrictions $\gamma = \mu = 0$,

$$y_t^{+} = (-1)^{t-1} = \beta(-1)^{t-1}x_{0t-1} + (-1)^{t-1}\varepsilon_t = \beta x_{t-1}^{+} - (-1)^t\varepsilon_t$$

so that $\tilde{\beta}_{1n}^{+} = \beta - \sum_{t=1}^{n}(-1)^t\varepsilon_t\tilde{z}_{0t-1}^{+}/\sum_{t=1}^{n}x_{t-1}^{+}\tilde{z}_{0t-1}^{+}$ and $\tilde{\beta}_{1n}^{+} = \beta - \sum_{t=1}^{n}(-1)^t\varepsilon_t z_{2t-1}^{+}/\sum_{t=1}^{n}x_{t-1}^{+}z_{2t-1}^{+}$, the result for $\pi_n(\tilde{\beta}_{jn}^{-} - \beta)$ follows from (B.24) and (B.27) with u_t replaced by ε_t.

Proof of Lemma B1. Denote $x_0 = X_0(n)$ for brevity. By employing (6), we obtain

$$\bar{x}_{n-1} = \mu + \bar{x}_{0n-1} + (x_0 - \mu)\frac{1}{n}\sum_{t=1}^{n}\rho_n^t. \tag{B.28}$$

$\bar{x}_{0n-1}$ has exact rate $O_p(n^{-1/2}(1-\rho_n)^{-1})$ under C$_+$(i), $O_p\left(n^{1/2}\right)$ under C$_+$(ii), $O_p(n^{-1}\rho_n^n(\rho_n - 1)^{-3/2})$ under C$_+$(iii) all of which dominate the $o_p\left(n^{-1}\kappa_n^{3/2}(\rho_n^n + 1)\right)$ order of the last term of (B.28). $\bar{x}_{0n-1} = O_p\left(n^{-1}\rho_n^n\kappa_n^{3/2}\right)$ dominates μ under C$_+$(ii)-C$_+$(iii) and also under C$_+$(i) provided that $n^{-1/2}(1-\rho_n)^{-1} \rightarrow \infty$; when $n^{-1/2}(1-\rho_n)^{-1} \rightarrow 0$ (the half of the C(i) region closer to stationarity), μ is the dominant term in (B.28). Combining the above yields the result for $\bar{x}_{n-1}$ for C$_+$(i)-C$_+$(iii). When $\rho_n \rightarrow \rho \leq -1$, $\bar{x}_{n-1}$ looses rate: $\sum_{t=1}^{n}x_{0t-1} = (1-\rho_n)^{-1}\left(\sum_{j=1}^{n}u_j - x_{0n}\right)$ and $\lim_{n\rightarrow\infty}(1-\rho_n) \geq 2$, imply that $\bar{x}_{0n-1} = O_p\left(n^{-1/2}\right)$ under C$_-$(i)-C$_-$(ii) and $\bar{x}_{0n-1} = O_p\left(x_n^{+}/n\right)$ under C$_-$(iii). Similarly, $\sum_{t=1}^{n}\rho_n^t = O\left(1 + |\rho_n|^n\right)$. Substituting the above rates into (B.28) yields the result for $\bar{x}_{n-1}$ for C$_-$(i)-C$_-$(iii). For the proof of (B.1) and (B.2), (6) implies the following identities for $\sum_{t=1}^{n}\underline{x}_{t-1}\underline{u}_t = \sum_{t=1}^{n}x_{t-1}u_t - n\bar{x}_{n-1}\bar{u}_n$ and $\sum_{t=1}^{n}\underline{x}_{t-1}^2 = \sum_{t=1}^{n}x_{t-1}^2 - n\bar{x}_{n-1}^2$:

$$\sum_{t=1}^{n}\underline{x}_{t-1}\underline{u}_t = \sum_{t=1}^{n}x_{0t-1}u_t + (X_0(n) - \mu)\sum_{t=1}^{n}\rho_n^t u_t - n(\bar{x}_{n-1} - \mu)\bar{u}_n \tag{B.29}$$

$$\begin{aligned}
\sum_{t=1}^{n}\underline{x}_{t-1}^2 &= \sum_{t=1}^{n}x_{0t-1}^2 - n\left(\bar{x}_{n-1}^2 - \mu^2\right) + 2\mu n\bar{x}_{0t-1} + 2(x_0 - \mu)\sum_{t=1}^{n}\rho_n^t x_{0t-1}\\
&\quad + 2\mu(x_0 - \mu)\sum_{t=1}^{n}\rho_n^t + (x_0 - \mu)^2\sum_{t=1}^{n}\rho_n^{2t}.
\end{aligned} \tag{B.30}$$

Now $c_n^{-1}(X_0(n)-\mu)\sum_{t=1}^n \rho_n^t u_t = o_p(1)\left((1-\rho_n^2)^{-1/2}\wedge n^{1/2}\right)c_n^{-1}\sum_{t=1}^n \rho_n^t u_t = o_p(1)$ under C(i)-C(iii), and (B.28) gives

$$c_n^{-1}n(\bar{x}_{n-1}-\mu)\bar{u}_n = c_n^{-1}\left(\bar{x}_{0n-1}+(x_0-\mu)\frac{1}{n}\frac{1-\rho_n^n}{1-\rho_n}\right)\sum_{t=1}^n u_t = nc_n^{-1}\bar{x}_{0n-1}\bar{u}_n + O_p\left(n^{-1/2}x_0\right)$$

under C(i)-C(iii). Under Assumption 2, the leading term is $O_p\left(n^{-1/2}\left|1-\rho_n^2\right|^{-1/2}\right)$ under C(i) and C(iii), $O_p(n^{-1})$ under C_-(ii), whereas under C_+(ii)

$$nc_n^{-1}\bar{x}_{0n-1}\bar{u}_n = \bar{x}_{0n-1}\bar{u}_n = n^{-3/2}\sum_{t=1}^n x_{0t-1}n^{-1/2}\sum_{j=1}^n u_j \to_d \int_0^1 J_c(t)\,dtB(1).$$

Substituting the above into to (B.29) proves (B.1) under Assumption 2. Under Assumption 4, the same orders apply to $nc_n^{-1}\bar{x}_{0n-1}\bar{u}_n$ under C(ii)-C(iii) and (B.1) continues to hold. Under C(i), $c_n = (1-\rho_n^2)^{-1}$ gives $nc_n^{-1}\bar{x}_{0n-1}\bar{u}_n = O_p(1)$; since $c_n = o(n)$, (B.1) continues to hold with c_n^{-1} replaced by n^{-1}. Turning to the right side of (B.30), the last term $c_n^{-2}(x_0-\mu)^2\sum_{t=1}^n \rho_n^{2t}$ dominates the penultimate term and is $O_p\left(x_0^2\left(\rho_n^2-1\right)\right) = o_p(1)$ by Assumption 3 under C(i)-C(iii). For the fourth term of (B.30),

$$c_n^{-2}|x_0|\left\|\sum_{t=1}^n x_{0t-1}\rho_n^t\right\|_{L_1} \le c_n^{-2}|x_0|\max_{t\le n}\|x_{0t}\|_{L_2}\sum_{j=1}^n|\rho_n|^j = |x_0|\,O\left(\left(\rho_n^2-1\right)^{1/2}\right) = o_p(1)$$

by Assumption 3 under C(i)-C(iii). For the third term of (B.30), $c_n^{-2}n\bar{x}_{0t-1} = c_n^{-2}\sum_{t=1}^n x_{0t-1}$ is $O_p\left(n^{-1/2}\right)$ under C(i)-C(ii) and $O_p\left(\left(\rho_n^2-1\right)^{1/2}\rho_n^{-n}\right)$ under C(iii). For the second term of (B.30),

$$\frac{n}{c_n^2}\left(\bar{x}_{n-1}^2-\mu^2\right) = \frac{n}{c_n^2}(\bar{x}_{n-1}-\mu)(\bar{x}_{n-1}+\mu) = [1+o_p(1)]\frac{n}{c_n^2}\bar{x}_{0n-1}(\bar{x}_{0n-1}+2\mu)$$

which is: $O_p\left(n^{-1}(1-\rho_n)^{-1}\right)$ under C(i), $O_p(\kappa_n/n)$ under C(iii), $o_p(n^{-1})$ under C_-(ii) and

$$\frac{n}{c_n^2}\left(\bar{x}_{n-1}^2-\mu^2\right) = \frac{1}{n}\bar{x}_{0n-1}^2 + o_p(1) \to_d \left(\int_0^1 J_c(t)\,dt\right)^2$$

under C_+(ii). Substituting the above to (B.30) proves (B.2).

For part (ii), $n^{-1}\sum_{t=1}^n x_{0t-1}u_t \to_p \Gamma$ under C_+(i) and Assumption 4 by Lemma 2.2(i) of MP(2020). Using the recursion for x_{0t}, we obtain the identity

$$n^{-1}\left(1-\rho_n^2\right)\sum_{t=1}^n x_{0t-1}^2 = n^{-1}\sum_{t=1}^n u_t^2 + 2\rho_n n^{-1}\sum_{t=1}^n x_{0t-1}u_t - n^{-1}x_{0t-1}^2 \to_p \sigma^2 + 2\rho\Gamma.$$

Hence, using (B.1) with c_n^{-1} replaced by n^{-1} and (B.2) we may write

$$\left(1-\rho_n^2\right)^{-1}(\hat{\rho}_n-\rho_n) = \frac{\frac{1}{n}\sum_{t=1}^n x_{0t-1}u_t}{\left(1-\rho_n^2\right)\frac{1}{n}\sum_{t=1}^n x_{0t-1}^2} + o_p(1) \to_p \frac{\Gamma}{\sigma^2+2\rho\Gamma}.$$

Under Assumption 2, using (B.1) and (B.2) we may write

$$\left(n\left(1-\rho_n^2\right)^{-1}\right)^{1/2}(\hat{\rho}_n-\rho_n) = \frac{n^{-1/2}\left(1-\rho_n^2\right)^{1/2}\sum_{t=1}^n x_{0t-1}u_t}{n^{-1}\left(1-\rho_n^2\right)\sum_{t=1}^n x_{0t-1}^2} + o_p(1)$$

and the last term converges in distribution to $\mathcal{N}(0,1)$ under Assumption 2 by Giraitis and Phillips (2006). For part (iii), using (B.1) and (B.2),

$$\begin{aligned}\left(\rho_n^2-1\right)^{-1}\rho_n^n(\hat{\rho}_n-\rho_n) &= \frac{\left(\rho_n^2-1\right)\rho_n^{-n}\sum_{t=1}^n x_{0t-1}u_t}{\left(\rho_n^2-1\right)^2\rho_n^{-2n}\sum_{t=1}^n x_{0t-1}^2} + o_p(1)\\ &= \frac{\Upsilon_n X_n}{X_n^2} + o_p(1) = \frac{\Upsilon_n}{X_n} + o_p(1)\end{aligned}$$

form the approximations $\left(\rho_n^2-1\right)^2\rho_n^{-2n}\left|\sum_{t=1}^n x_{0t-1}^2 - x_{0n}^2\right| = o_p(1)$ with $\rho_n^{-n}x_{0n} = \sum_{j=1}^n \rho_n^{-j}u_j$ and

$$\left(\rho_n^2-1\right)\rho_n^{-n}\sum_{t=1}^n x_{0t-1}u_t = \left(\rho_n^2-1\right)\left(\sum_{j=1}^n \rho_n^{-j}u_j\right)\sum_{t=1}^n \rho_n^{-(n-t+1)}u_t + o_p(1)$$

established in Phillips and Magdalinos (2007). When $\rho_n \to 1$, Magdalinos (2012) shows that $[X_n, \Upsilon_n] \to_d \mathcal{N}\left(0, \sigma^2 I_2\right)$ implying that $\Upsilon_n/X_n \to_d \mathcal{C}$; when $\rho_n \to \rho > 1$, Lemma 4(iii) shows that

$X_n \to_p X_\infty \neq 0\ a.s.$, and $\mathbb{E}\Upsilon_n^2 \to \sigma^2$, so in both cases $|\Upsilon_n/X_n| = O_p(1)$ and $(\rho_n^2-1)^{-1}\rho_n^n(\hat{\rho}_n-\rho_n) = O_p(1)$ over the C_+(iii) range.

For part (iv) $\rho_n < 0$ and $\bar{x}_{0n-1} = O_p\left(n^{-1/2}\right)$ under C_-(ii); hence, using (B.1) and (B.2) and recalling the notation $x_t^+ = (-1)^t x_{0t}$ of Lemma 6 we obtain

$$\begin{aligned} c_n(\hat{\rho}_n-\rho_n) &= \frac{c_n^{-1}\sum_{t=1}^n x_{0t-1}u_t}{c_n^{-2}\sum_{t=1}^n x_{0t-1}^2} + o_p(1) = \frac{-c_n^{-1}\sum_{t=1}^n x_{t-1}^+(-1)^t u_t}{c_n^{-2}\sum_{t=1}^n x_{t-1}^{+2}} + o_p(1) \\ &= -c_n\left(\hat{\rho}_n^+ - |\rho_n|\right) + o_p(1) \end{aligned}$$

as required. This completes the proof of Lemma B1.

Proof of Lemma B2. For part (i), write $\varphi_{1n}^n = e^{n\log[1-(1-\varphi_{1n}))]} = e^{-n(1-\varphi_{1n})(1+o(1))}$ since $\log(1-x) = -x + O(x^2)$ as $x \to 0$; hence $[n(1-\varphi_{1n})]^p\varphi_{1n}^n = [n(1-\varphi_{1n})]^p O\left(e^{-n(1-\varphi_{1n})}\right) \to 0$ for any $p \geq 0$ since $n(1-\varphi_{1n}) \to \infty$ under C(i). Under C(iii), $n(\varphi_{2n}-1) \to \infty$ and $\varphi_{2n}^{-n} = e^{-n\log[1+(\varphi_{2n}-1)]} = O\left(e^{-n(\varphi_{2n}-1)}\right)$ shows that $[n(\varphi_{2n}-1)]^p\varphi_{2n}^{-n} \to 0$ for any $p \geq 0$. The orders of $\sum_{t=1}^n t^p\varphi_{1n}^t$ and $\sum_{t=1}^n t^p\varphi_{2n}^t$ for $p=0$ are trivial (geometric progression). For $p>0$, employing an Euler summation argument and the change of variables $s = (1-\varphi_{1n})t$

$$\begin{aligned} \textstyle\sum_{t=1}^n t^p\varphi_{1n}^t &= \textstyle\int_1^{n+1}\lfloor t\rfloor^p\varphi_{1n}^{\lfloor t\rfloor}dt \\ &= (1-\varphi_{1n})^{-1-p}\int_{1-\varphi_{1n}}^{(n+1)(1-\varphi_{1n})}\left(\frac{\lfloor(1-\varphi_{1n})^{-1}s\rfloor}{(1-\varphi_{1n})^{-1}}\right)^p\varphi_{1n}^{\lfloor(1-\varphi_{1n})^{-1}s\rfloor}ds. \end{aligned} \tag{B.31}$$

Since $1-\varphi_{1n} \to 0$, $n(1-\varphi_{1n}) \to \infty$ and

$$\begin{aligned} \varphi_{1n}^{\lfloor(1-\varphi_{1n})^{-1}s\rfloor} &= (1-(1-\varphi_{1n}))^{\lfloor(1-\varphi_{1n})^{-1}s\rfloor} = \exp\left\{\lfloor(1-\varphi_{1n})^{-1}s\rfloor\log(1-(1-\varphi_{1n}))\right\} \\ &= \exp\left\{-\lfloor(1-\varphi_{1n})^{-1}s\rfloor(1-\varphi_{1n}) + O((1-\varphi_{1n}))\right\} \to e^{-s} \end{aligned}$$

the dominated convergence theorem implies that the integral on the right side of (B.31) converges to $\int_0^\infty s^p e^{-s}ds = \Gamma(p+1)$, and the claim for $\sum_{t=1}^n t^p\varphi_{1n}^t$ follows from (B.31). The result for $\sum_{t=1}^n t^p\varphi_{2n}^{-t}$ can be derived in the same way by interchanging the roles of $1-\varphi_{1n}$ and $\varphi_{2n}-1$.

For part (ii), applying (6) to the instrument $\tilde{z}_{1t} = \sum_{j=1}^t \varphi_{1n}^{t-j}\Delta x_j$ in (19), we obtain

$$\tilde{z}_{1t} - \tilde{z}_{0t} = \textstyle\sum_{j=1}^t \varphi_{1n}^{t-j}(\Delta x_j - \Delta x_{0j}) = (X_0(n)-\mu)(\rho_n-1)\varphi_{1n}^{t-1}\sum_{j=0}^{t-1}(\rho_n/\varphi_{1n})^j.$$

Evaluating the geometric progression when $n|\varphi_{1n}-\rho_n| \to \infty$ and noting that $\left|\sum_{j=0}^{t-1}(\rho_n/\varphi_{1n})^j\right| \leq bt$ for all t and some $b \in (0,\infty)$, we obtain the following decomposition for $\tilde{z}_{1t}$:

$$\tilde{z}_{1t} = \tilde{z}_{0t} + (X_0(n)-\mu)q_{nt}, \quad \tilde{z}_{0t} = \textstyle\sum_{j=1}^t \varphi_{1n}^{t-j}\Delta x_{0j}, \tag{B.32}$$

where $q_{nt} = \frac{1-\rho_n}{\varphi_{1n}-\rho_n}(\rho_n^t - \varphi_{1n}^t)$ when $n|\varphi_{1n}-\rho_n| \to \infty$ and $|q_{nt}| \leq bt(1-\varphi_{1n})\varphi_{1n}^t$ for all $t \leq n$ and some $b \in (0,\infty)$ when $|\varphi_{1n}-\rho_n| = O(n^{-1})$. Similarly, applying (6) to the instrument $\tilde{z}_{1t}^- = \sum_{j=1}^t \left(\varphi_{1n}^-\right)^{t-j}\nabla\underline{x}_j$ in (19), we obtain

$$\tilde{z}_{1t}^- - \tilde{z}_{0t}^- = (X_0(n)-\mu)(1+\rho_n)\left(\varphi_{1n}^-\right)^{t-1}\textstyle\sum_{j=0}^{t-1}\left(\rho_n/\varphi_{1n}^-\right)^j - \eta_{nt}$$

where

$$\eta_{nt} = 2\left[(X_0(n)-\mu)\frac{1}{n}\rho_n\frac{1-\rho_n^n}{1-\rho_n} + \bar{x}_{0n}\right]\frac{1-\left(\varphi_{1n}^-\right)^t}{1-\varphi_{1n}^-} = O_p\left(n^{-1/2}\right)\left(1-\left(\varphi_{1n}^-\right)^t\right) \tag{B.33}$$

satisfies $\max_{t\leq n}|\eta_{nt}| = O_p\left(n^{-1/2}\right)$ because $1-\varphi_{1n}^- \to 2$ and, when $\rho_n \to \rho \in [-1,1)$: $|\bar{x}_{0n}| = O_p\left(n^{-1/2}\right)$ and $\sup_{n\geq 1}(1-\rho_n)^{-1} < \infty$. The above yields the following decomposition for $\tilde{z}_{1t}^-$:

$$\tilde{z}_{1t}^- = \tilde{z}_{0t}^- - (X_0(n)-\mu)q_{nt} - \eta_{nt}, \quad \tilde{z}_{0t}^- = \textstyle\sum_{j=1}^t\left(\varphi_{1n}^-\right)^{t-j}\nabla x_{0j} \tag{B.34}$$

where $\max_{t\leq n}|\eta_{nt}| = O_p\left(n^{-1/2}\right)$, $q_{nt} = \frac{1+\rho_n}{\varphi_{1n}-\rho_n}\left(\rho_n^t - \left(\varphi_{1n}^-\right)^t\right)$ when $n\left|\varphi_{1n}^- - \rho_n\right| \to \infty$ and $|q_{nt}| \leq$

$bt\left(1+\varphi_{1n}^{-}\right)\left|\varphi_{1n}^{-}\right|^{t}$ for all $t\leq n$ and some $b\in(0,\infty)$ when $\left|\varphi_{1n}^{-}-\rho_{n}\right|=O\left(n^{-1}\right)$. We may use a common q_{nt} in the decompositions (B.32) and (B.34) by defining

$$q_{nt}=\frac{1-|\rho_{n}|}{\phi_{1n}-\rho_{n}}\left(\rho_{n}^{t}-\phi_{1n}^{t}\right)\text{ when }n\left|\phi_{1n}-\rho_{n}\right|\rightarrow\infty\tag{B.35}$$

(with ϕ_{1n} given in (43)) and $|q_{nt}|\leq bt\left(1-|\phi_{1n}|\right)|\phi_{1n}|^{t}$ for all $t\leq n$ and some $b\in(0,\infty)$ when $|\phi_{1n}-\rho_{n}|=O\left(n^{-1}\right)$.

We first show that q_{nt} and κ_{n} in (B.35) and (4) satisfy

$$[\epsilon_{1n},\epsilon_{2n}]=n^{-1/2}\left[\left(1-\rho_{n}^{2}\phi_{1n}^{2}\right)^{1/2}\kappa_{n}^{1/2},\left(1-\rho_{n}^{2}\phi_{1n}^{2}\right)o\left(\kappa_{n}\right)\right]\left(\textstyle\sum_{t=1}^{n}q_{nt}^{2}\right)^{1/2}\rightarrow0.\tag{B.36}$$

When $n\left|\phi_{1n}-\rho_{n}\right|\rightarrow\infty$, $\epsilon_{1n}\leq\sqrt{2}\left(\kappa_{n}\left|\phi_{1n}-\rho_{n}\right|\right)^{-1/2}\left(n^{-1}\sum_{t=1}^{n}\left(\rho_{n}^{2t}+\phi_{1n}^{2t}\right)\right)^{1/2}\rightarrow0$ and $\epsilon_{2n}=o\left(1\right)\left(n^{-1}\sum_{t=1}^{n}\left(\rho_{n}^{2t}+\varphi_{1n}^{2t}\right)\right)^{1/2}\rightarrow0$. When $\left|\phi_{1n}-\rho_{n}\right|=O\left(n^{-1}\right)$, $\sum_{t=1}^{n}t^{2}\left|\phi_{1n}\right|^{2t}=O\left(\left(1-\left|\phi_{1n}\right|\right)^{-3}\right)$ by part (i) and $\kappa_{n}^{-1}=O\left(1-\left|\phi_{1n}\right|\right)$ imply that both ϵ_{1n} and ϵ_{1n} are $O\left(n^{-1/2}\left(1-\left|\phi_{1n}\right|\right)^{-1/2}\right)$.

Now (B.32) and Assumption 3 give $r_{1n}=n^{-1/2}\left(1-\rho_{n}^{2}\phi_{1n}^{2}\right)^{1/2}o_{p}\left(\kappa_{n}^{1/2}\right)\sum_{t=1}^{n}q_{nt}u_{t}$. Since

$$\begin{aligned}\left\|\textstyle\sum_{t=1}^{n}q_{nt}u_{t}\right\|_{L_{2}}^{2}&\leq2\textstyle\sum_{s=1}^{n}\sum_{t=s}^{n}\left|q_{nt}\right|\left|q_{ns}\right|\left|\gamma_{u}\left(t-s\right)\right|=2\sum_{t=0}^{n}\left|\gamma_{u}\left(t\right)\right|\sum_{s=1}^{n-t}\left|q_{nt+s}\right|\left|q_{ns}\right|\\&\leq2\textstyle\sum_{t=0}^{n}\left|\gamma_{u}\left(t\right)\right|\left(\sum_{s=1}^{n-t}q_{ns}^{2}\right)^{1/2}\left(\sum_{s=1}^{n-t}q_{n,t+s}^{2}\right)^{1/2}\leq2\sum_{s=1}^{n}q_{ns}^{2}\sum_{t=0}^{\infty}\left|\gamma_{u}\left(t\right)\right|\end{aligned}$$

and $\sum_{t=0}^{\infty}\left|\gamma_{u}\left(t\right)\right|<\infty$ by Assumption 4, $r_{1n}\rightarrow_{p}0$ follows from the fact that $\epsilon_{1n}\rightarrow0$ in (B.36). By (B.34),

$$r_{1n}^{-}=r_{1n}+\pi_{n}^{-1}\textstyle\sum_{t=1}^{n}\eta_{nt}u_{t}=r_{1n}+O_{p}\left(n^{-1/2}\right)\pi_{n}^{-1}\sum_{t=1}^{n}\left(1-\left(\varphi_{1n}^{-}\right)^{t}\right)u_{t}$$

by (B.33); since the second term on the right is $O_{p}\left(\pi_{n}^{-1}\right)$, $r_{1n}\rightarrow_{p}0$ implies that $r_{1n}^{-}\rightarrow_{p}0$. For r_{2n}, (B.32) gives $\sum_{t=1}^{n}\left(\tilde{z}_{1t}^{2}-\tilde{z}_{0t}^{2}\right)=\left(X_{0}\left(n\right)-\mu\right)^{2}\sum_{t=1}^{n}q_{nt}^{2}+2\left(X_{0}\left(n\right)-\mu\right)\sum_{t=1}^{n}\tilde{z}_{0t}q_{nt}$ with $\left(X_{0}\left(n\right)-\mu\right)\left|\sum_{t=1}^{n}\tilde{z}_{0t}q_{nt}\right|\leq\left(\sum_{t=1}^{n}\tilde{z}_{0t}^{2}\right)^{1/2}\left(\left(X_{0}\left(n\right)-\mu\right)^{2}\sum_{t=1}^{n}q_{nt}^{2}\right)^{1/2}$ by the Cauchy-Schwarz inequality. We conclude that

$$\left|r_{2n}\right|\leq\epsilon_{1n}^{2}+2\left(n^{-1}\left(1-\rho_{n}^{2}\varphi_{1n}^{2}\right)\textstyle\sum_{t=1}^{n}\tilde{z}_{0t}^{2}\right)^{1/2}\left(\epsilon_{1n}^{2}\right)^{1/2}=o_{p}\left(1\right)$$

by (B.36) since $n^{-1}\left(1-\rho_{n}^{2}\varphi_{1n}^{2}\right)\sum_{t=1}^{n}\tilde{z}_{0t}^{2}=O_{p}\left(1\right)$ by Lemma 3.1 in MP(2020). By (B.34),

$$\begin{aligned}\left|r_{2n}^{-}\right|&\leq2\pi_{n}^{-2}\left(X_{0}\left(n\right)-\mu\right)^{2}\textstyle\sum_{t=1}^{n}q_{nt}^{2}+2\pi_{n}^{-2}\sum_{t=1}^{n}\left(\eta_{nt}^{2}+\tilde{z}_{0t}^{-}\eta_{nt}\right)+2\left|X_{0}\left(n\right)-\mu\right|\pi_{n}^{-2}\sum_{t=1}^{n}\tilde{z}_{0t}^{-}q_{nt}\\&=2\pi_{n}^{-2}\textstyle\sum_{t=1}^{n}\eta_{nt}^{2}+2\pi_{n}^{-2}\sum_{t=1}^{n}\tilde{z}_{0t}^{-}\eta_{nt}+o_{p}\left(1\right)\end{aligned}$$

from the estimation used to show that $r_{2n}\rightarrow_{p}0$; since $\max_{t\leq n}\left|\eta_{nt}\right|=O_{p}\left(n^{-1/2}\right)$ the first term on the right is $O_{p}\left(\pi_{n}^{-2}\right)$ and the second term is $O_{p}\left(n^{1/2}\pi_{n}^{-2}\max_{t\leq n}\left\|\tilde{z}_{0t}^{-}\right\|_{L_{2}}\right)=o_{p}\left(n\pi_{n}^{-2}\right)=o_{p}\left(1\right)$, showing that $r_{2n}^{-}\rightarrow_{p}0$.

For $r_{3n}=r_{3n}^{\prime}+r_{3n}^{\prime\prime}$, with

$$\left[r_{3n}^{\prime},r_{3n}^{\prime\prime}\right]=n^{-1}\left(1-\rho_{n}^{2}\varphi_{1n}^{2}\right)\left[\textstyle\sum_{t=1}^{n}\left(\tilde{z}_{1t}-\tilde{z}_{0t}\right)x_{t},\sum_{t=1}^{n}\left(x_{t}-x_{0t}\right)\tilde{z}_{0t}\right],$$

the Cauchy-Schwarz inequality and (B.32) imply that $r_{3n}^{\prime}\leq O_{p}\left(1\right)\epsilon_{2n}\left(n^{-1}\kappa_{n}^{-1}\sum_{t=1}^{n}x_{t}^{2}\right)^{1/2}=o_{p}\left(1\right)$ by (B.36) and $n^{-1}\kappa_{n}^{-1}\sum_{t=1}^{n}x_{t}^{2}=O_{p}\left(1\right)$. For $r_{3n}^{\prime\prime}$, (6) and $\sum_{t=1}^{n}\tilde{z}_{0t}=O_{p}\left(n^{1/2}\left(1-\rho_{n}^{2}\varphi_{1n}^{2}\right)^{-1}\right)$ imply that

$$r_{3n}^{\prime\prime}=o_{p}\left(1\right)r_{3n}^{\prime\prime\prime}+O_{p}\left(n^{-1/2}\right),\quad r_{3n}^{\prime\prime\prime}=n^{-1}\left(1-\rho_{n}^{2}\varphi_{1n}^{2}\right)\kappa_{n}^{1/2}\textstyle\sum_{t=1}^{n}\tilde{z}_{0t}\rho_{n}^{t}.\tag{B.37}$$

When $\left|\varphi_{1n}-\rho_{n}\right|=O\left(n^{-1}\right)$, the Cauchy-Schwarz inequality and $n^{-1}\left(1-\rho_{n}^{2}\varphi_{1n}^{2}\right)\sum_{t=1}^{n}\tilde{z}_{0t}^{2}=O_{p}\left(1\right)$ imply that $r_{3n}^{\prime\prime\prime}\leq O_{p}\left(1\right)\left(n^{-1}\left(1-\rho_{n}^{2}\varphi_{1n}^{2}\right)\kappa_{n}^{1/2}\sum_{t=1}^{n}\rho_{n}^{2t}\right)^{1/2}=O_{p}\left(1\right)\left(n^{-1}\left(1-\varphi_{1n}^{2}\right)^{-1/2}\right)^{1/2}=o_{p}\left(1\right)$. When $n\left|\varphi_{1n}-\rho_{n}\right|\rightarrow\infty$, the definition of $\tilde{z}_{0t}$, the summation by parts formula and the Cauchy-

Schwarz inequality give

$$\begin{aligned}\sum_{t=1}^{n}\tilde{z}_{0t}\rho_n^t &= \sum_{t=1}^{n}\sum_{j=1}^{t}\varphi_{1n}^{t-j}\Delta x_{0j}\rho_n^t = \sum_{j=1}^{n}\rho_n^j\Delta x_{0j}\sum_{t=0}^{n-j}(\varphi_{1n}\rho_n)^t \\ &= (1-\varphi_{1n}\rho_n)^{-1}\left(\sum_{j=1}^{n}\rho_n^j\Delta x_{0j}-\varphi_{1n}\rho_n^n\tilde{z}_{0n}\right) \\ &= (1-\varphi_{1n}\rho_n)^{-1}\left[(1-\rho_n)\sum_{j=1}^{n}\rho_n^j x_{0j}+\rho_n^n(x_{0n}-\varphi_{1n}\tilde{z}_{0n})\right] \\ &\leq (1-\varphi_{1n}\rho_n)^{-1}\left[(1-\rho_n)\left(\sum_{j=1}^{n}\rho_n^{2j}\right)^{1/2}\left(\sum_{j=1}^{n}x_{0j}^2\right)^{1/2}+O_p\left(\rho_n^n\kappa_n^{1/2}\right)\right].\end{aligned}$$

Since $(1-\rho_n)\left(\sum_{j=1}^{n}\rho_n^{2j}\right)^{1/2}=O\left(\kappa_n^{-1/2}\right)$, $|r_{3n}'''|\leq O(1)\left(n^{-2}\sum_{j=1}^{n}x_{0j}^2\right)^{1/2}+O_p(\rho_n^n)=O_p(1)$ so $r_{3n}=o_p(1)$ follows from (B.37). For r_{3n}^-, write $r_{3n}^-=r_{3n}'+r_{3n}''$, where r_{3n}',r_{3n}'' are defined as above with $(\tilde{z}_{1t},\tilde{z}_{0t},\varphi_{1n})$ replaced by $\left(\tilde{z}_{1t}^-,\tilde{z}_{0t}^-,\varphi_{1n}^-\right)$, with (B.34) and the Cauchy-Schwarz inequality on $\sum_{t=1}^{n}q_{nt}x_t$ and $\sum_{t=1}^{n}\eta_{nt}x_t$ giving $|r_{3n}'|\leq\left[O_p(1)\epsilon_{2n}+\pi_n^{-2}n^{1/2}\kappa_n^{1/2}\left(\sum_{t=1}^{n}\eta_{nt}^2\right)^{1/2}\right]\left(n^{-1}\kappa_n^{-1}\sum_{t=1}^{n}x_t^2\right)^{1/2}$ which is $o_p(1)$ since $\epsilon_{2n}=o_p(1)$ and $n^{-1}\kappa_n^{-1}\sum_{t=1}^{n}x_t^2$ and $\sum_{t=1}^{n}\eta_{nt}^2$ are $O_p(1)$. Hence, (B.37) continues to hold with $(\tilde{z}_{0t},\varphi_{1n})$ replaced by $\left(\tilde{z}_{0t}^-,\varphi_{1n}^-\right)$, and recalling the notation $\tilde{z}_{0t}^+=(-1)^t\tilde{z}_{0t}^-$,

$$r_{3n}'''=\pi_n^{-2}\kappa_n^{1/2}\sum_{t=1}^{n}\tilde{z}_{0t}^-\rho_n^t=\pi_n^{-2}\kappa_n^{1/2}\sum_{t=1}^{n}\tilde{z}_{0t}^+|\rho_n|^t=o_p(1)$$

from the computation used for r_{3n}, completing the proof of $r_{3n}^-=o_p(1)$.

For part (iii), (B.32) gives $(1-\rho_n\varphi_{1n})\sum_{t=1}^{n}q_{nt}\sim(1-\varphi_{1n})^2\sum_{t=1}^{n}t\varphi_{1n}^t=O(1)$ by part (i) when $|\varphi_{1n}-\rho_n|=O(n^{-1})$ and $(1-\rho_n\varphi_{1n})\sum_{t=1}^{n}q_{nt}=O(\kappa_n^{-1})\sum_{t=1}^{n}(\rho_n^t-\varphi_{1n}^t)=O(1)+O((\kappa_n(1-\varphi_{1n}))^{-1})$ when $n|\varphi_{1n}-\rho_n|\to\infty$. Substituting into (B.32), $(1-\rho_n\varphi_{1n})\sum_{t=1}^{n}(\tilde{z}_{1t}-\tilde{z}_{0t})=o_p(\kappa_n^{1/2})+o_p(\kappa_n^{-1/2}(1-\varphi_{1n})^{-1})$, and $(1-\rho_n\varphi_{1n})\sum_{t=1}^{n}\tilde{z}_{0t}=O_p\left(n^{1/2}\right)$ established in MP(2020) proves the order of magnitude of part (iii). Under C(ii), $n^{-1/2}(1-\varphi_{1n})\sum_{t=1}^{n}(\tilde{z}_{1t}-\tilde{z}_{0t})=o_p(1)$. The recursion $\tilde{z}_{0t}=\varphi_{1n}\tilde{z}_{0t-1}+\Delta x_{0t}$ implies that $(1-\varphi_{1n})n^{-1/2}\sum_{t=1}^{n}\tilde{z}_{0t-1}=n^{-1/2}(x_{0n}-\tilde{z}_{0n})$ and part (iii) follows from the fact that $n^{-1/2}\tilde{z}_{0n}\to_p 0$.

For part (iv), applying the identity $\hat{u}_t=\underline{u}_t-(\hat{\rho}_n-\rho_n)\underline{x}_{t-1}$ to $\tilde{z}_{2t}=\sum_{j=1}^{t}\varphi_{2n}^{t-j}\hat{u}_j$ and $\tilde{z}_{2t}^-=\sum_{j=1}^{t}\left(\varphi_{2n}^-\right)^{t-j}\hat{u}_j$ in (20), we obtain the decomposition

$$\tilde{z}_{2t}=z_{2t}-(\hat{\rho}_n-\rho_n)\psi_{nt-1}+\phi_{2n}^t g_{n,t}\text{ and }\tilde{z}_{2t}^-=z_{2t}^--(\hat{\rho}_n-\rho_n)\psi_{nt-1}+\phi_{2n}^t g_{n,t}\qquad\text{(B.38)}$$

where $\phi_{2n}=\varphi_{2n}$ for $\tilde{z}_{2t}$ and $\phi_{2n}=\varphi_{2n}^-$ for $\tilde{z}_{2t}^-$, $\psi_{nt-1}=\sum_{j=1}^{t}\phi_{2n}^{t-j}x_{j-1}$, and the last term $g_{n,t}=[(\hat{\rho}_n-\rho_n)\bar{x}_{n-1}-\bar{u}_n]\left(1-\phi_{2n}^{-t}\right)/(\phi_{2n}-1)$ satisfies $\max_{1\leq t\leq n}|g_{n,t}|=O_p\left(n^{-1/2}(\phi_{2n}-1)^{-1}\right)$ since $(\hat{\rho}_n-\rho_n)\bar{x}_{n-1}=O_p\left(\kappa_n^{1/2}/n\right)$ by Lemma B1(i) under C(iii) and standard local to unity asymptotics under C(ii). Note that $\max_{1\leq t\leq n}|g_{n,t}|=O_p\left(n^{-1/2}\right)$ when $\phi_{2n}=\varphi_{2n}^-$. When $n|\phi_{2n}-\rho_n|\to\infty$, (B.4) will follow from the following identity for ψ_{nt-1} in (B.38):

$$\psi_{nt-1}=\frac{1}{\phi_{2n}-\rho_n}\left(\phi_{2n}z_{2t-1}-\rho_n x_{t-1}+\phi_{2n}^t g_{n,t}'\right)\qquad\text{(B.39)}$$

where $g_{n,t}'=X_0(n)-\mu\left(1-\phi_{2n}^{-(t-1)}\right)\frac{\rho_n-1}{\phi_{2n}-1}$, with the order in (B.4) following from $(\phi_{2n}-\rho_n)^{-1}(\hat{\rho}_n-\rho_n)\max_{1\leq t\leq n}\left|g_{n,t}'\right|=o_p\left(\kappa_n^{-1/2}\rho_n^{-n}(\phi_{2n}-1)^{-1}\right)+O_p\left(\kappa_n^{-1}\rho_n^{-n}(\phi_{2n}-1)^{-1}\right)$ under C(ii)-C(iii) with both orders being majorised under C(ii) to become $o_p\left(n^{-1/2}(\phi_{2n}-1)^{-1}\right)$ and $O_p\left(n^{-1}(\phi_{2n}-1)^{-1}\right)$. To prove (B.39), substituting x_t in (6) into the expression for ψ_{nt-1} in (B.38) we obtain

$$\psi_{nt-1} = \phi_{2n}^{t-1} X_0(n) + \mu \frac{\phi_{2n}^{t-1} - 1}{\phi_{2n} - 1} + \phi_{2n}^t \sum_{i=1}^{t-1} \rho_n^{-i-1} u_i \sum_{j=i+1}^t \left(\frac{\rho_n}{\phi_{2n}}\right)^j$$
$$+\phi_{2n}^t \rho_n^{-1} (X_0(n) - \mu) \sum_{j=2}^t \left(\frac{\rho_n}{\phi_{2n}}\right)^j. \quad \text{(B.40)}$$

Evaluating the geometric progression

$$\sum_{j=i+1}^t (\rho_n/\phi_{2n})^j = \frac{\phi_{2n}}{\phi_{2n} - \rho_n} \left\{(\rho_n/\phi_{2n})^{i+1} - (\rho_n/\phi_{2n})^{t+1}\right\}$$

when $n|\phi_{2n} - \rho_n| \to \infty$, we obtain

$$\psi_{nt-1} = \phi_{2n}^{t-1} X_0(n) + \mu \frac{\phi_{2n}^{t-1} - 1}{\phi_{2n} - 1} + \frac{1}{\phi_{2n} - \rho_n} \left\{\sum_{i=1}^{t-1} \phi_{2n}^{t-i} u_i - \sum_{i=1}^{t-1} \rho_n^{t-i} u_i\right\} + \frac{(X_0(n) - \mu)\rho_n}{\phi_{2n} - \rho_n} \left(\phi_{2n}^{t-1} - \rho_n^{t-1}\right)$$

and using the expression for x_t in (6), $z_{2t} = \sum_{i=1}^t \varphi_{2n}^{t-i} u_i$ and $z_{2t}^- = \sum_{i=1}^t \left(\varphi_{2n}^-\right)^{t-i} u_i$ proves (B.39). This completes the proof of (B.4) when $n|\phi_{2n} - \rho_n| \to \infty$.

When $|\phi_{2n} - \rho_n| = O(n^{-1})$, $\rho_n/\phi_{2n} = 1 + (\rho_n - \phi_{2n})/\phi_{2n} = 1 + O(n^{-1})$ so $\sum_{j=i+1}^t (\rho_n/\phi_{2n})^j \leq nb$ for all $i < t \leq n$ and some $b > 0$. Substituting into (B.40) we conclude that

$$g_n = (\hat{\rho}_n - \rho_n) \max_{1\leq t\leq n} \left|\phi_{2n}^{-t}\psi_{nt-1}\right| \leq 4bn(\hat{\rho}_n - \rho_n) \max_{1\leq t\leq n} \left|\sum_{i=1}^{t-1} \rho_n^{-i-1} u_i\right| = O_p\left(n|\rho_n|^{-n}\left(\rho_n^2 - 1\right)^{1/2}\right)$$
$$= O_p\left(n^{-1/2}\left(\rho_n^2 - 1\right)^{-1}|\rho_n|^{-n} n^{3/2}\left(\rho_n^2 - 1\right)^{3/2}\right) = o_p\left(n^{-1/2}\left(\rho_n^2 - 1\right)^{-1}\right) = o_p\left(n^{-1/2}(\phi_{2n} - 1)^{-1}\right)$$

by Lemma B2(i) since $|\phi_{2n} - \rho_n| = O(n^{-1})$ implies that $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(iii). This completes the proof of (B.4) when $|\varphi_{2n} - \rho_n| = O(n^{-1})$.

For the remainder of part (iv), we employ (B.4) to each of R_{1n}-R_{4n}. Using (19),

$$(\varphi_{2n} - 1) v_{n,z}^{-1} \sum_{t=1}^n z_{2t-1} = v_{n,z}^{-1} z_{2n} - v_{n,z}^{-1} \sum_{t=1}^n u_t = Z_n + o_p(1) \quad \text{(B.41)}$$

since $n^{1/2} v_{n,z}^{-1} = n^{1/2}(\varphi_{2n} - 1)^{1/2} \varphi_{2n}^{-n} \to 0$ by Lemma B2(i); also $\sum_{t=1}^n x_t = O_p(\kappa_n \nu_n)$ by Lemma B1(i). Using (B.4) and the above orders for $\sum_{t=1}^n z_{2t-1}$ and $\sum_{t=1}^n x_t$ we obtain

$$R_{1n} = g_n \nu_{n,z}^{-1} O(|\phi_{2n}|^n) + \mathbf{1}\{n|\phi_{2n} - \rho_n| \to \infty\} O_p\left(\frac{(\phi_{2n} - 1)\kappa_n^{-1/2} \nu_{n,z}^{-1} \nu_n^{-1}}{\phi_{2n} - \rho_n}\right) \sum_{t=1}^n (z_{2t-1} + x_{t-1})$$
$$= O_p\left(n^{-1/2}(\phi_{2n} - 1)^{-1/2}\right) + \mathbf{1}\{n|\phi_{2n} - \rho_n| \to \infty\} O_p\left(n^{1/2}(\phi_{2n} - 1)^{1/2} \varphi_{2n}^{-n}\right) = o_p(1)$$

by Lemma B2(i). Since $\sum_{t=1}^n z_{2t-1}^- = O_p(v_{n,z})$, the above bounds for R_{1n} also show that $R_{1n}^- = o_p(1)$. For $R_{2n} = \left(\phi_{2n}^2 - 1\right) \phi_{2n}^{-n} \sum_{t=1}^n r_{nt-1} u_t$, the second term arising from (B.4) satisfies

$$\left(\phi_{2n}^2 - 1\right) \phi_{2n}^{-n} g_n \sum_{t=1}^n \phi_{2n}^{t-1} u_t = o_p\left(n^{-1/2}\left(\phi_{2n}^2 - 1\right)^{-1/2} \phi_{2n}^{-n}\right) = o_p(1)$$

by Lemma B2(i); for the first term, when $n|\varphi_{2n} - \rho_n| \to \infty$, (B.4) and the triangle inequality give

$$|R_{2n}| \leq \frac{|\hat{\rho}_n - \rho_n|(\phi_{2n} - 1)}{|\phi_{2n} - \rho_n||\phi_{2n}|^n} \left\{\left|\sum_{t=1}^n z_{2t-2} u_t\right| + \left|\sum_{t=1}^n x_{t-2} u_t\right|\right\} + o_p(1)$$
$$= \frac{\phi_{2n} - 1}{|\phi_{2n} - \rho_n|} O_p\left(|\phi_{2n}|^{-n} \kappa_n^{-1/2} \nu_n^{-1}\right) \left\{O_p\left(\kappa_n^{1/2} \nu_n\right) + O_p\left(\left(\phi_{2n}^2 - 1\right)^{-1/2} \nu_{n,z}\right)\right\} + o_p(1)$$

by (B.3). The first term above is $O_p\left(|\phi_{2n}|^{-n}\right)$ and the second term is $O_p\left(\frac{|\rho_n|^{-n}}{\kappa_n|\phi_{2n} - \rho_n|}\right)$ which is $O_p\left(|\rho_n|^{-n}\right)$ under C(iii) and $O_p\left(n^{-1}|\phi_{2n} - 1|^{-1}\right)$; since both terms are $o_p(1)$, $|R_{2n}| = o_p(1)$ under C(ii)-C(iii). Since $\sum_{t=1}^n z_{2t-2} u_t$ and $\sum_{t=1}^n z_{2t-2}^- u_t$ have the same order of magnitude, the same bound shows that $\left|R_{2n}^-\right| = o_p(1)$.

For $R_{3n} = R_{3n}' + R_{3n}''$, where $R_{3n}' = s_n^{-1} \sum_{t=1}^n r_{nt} x_t$ and $R_{3n}'' = s_n^{-1} \sum_{t=1}^n z_{2t}(x_t - x_{0t})$, we estimate the two terms arising from (B.4) for R'_{3n}: by (6)

$$s_n^{-1} g_n \sum_{t=1}^n \phi_{2n}^t x_t = s_n^{-1} g_n \left(\sum_{t=1}^n \phi_{2n}^t x_{0t} + \mu \sum_{t=1}^n \phi_{2n}^t + (X_0 - \mu) \sum_{t=1}^n (\phi_{2n}\rho_n)^t \right). \qquad \text{(B.42)}$$

Using the rate of g_n in (B.4), the third term on the right is $o_p\left(n^{-1/2}\left(\phi_{2n}^2 - 1\right)^{-1/2}\right)$. The second term is $o_p\left(n^{-1/2}\right)$ under C(ii) and $o_p\left(\kappa_n^{-1/2}(n/\kappa_n)|\rho_n|^{-n}\right) = o_p\left(\kappa_n^{-1/2}\right)$ under C(iii); the third term is $o_p\left(|\rho_n|^{-n}\kappa_n^{-1/2}\right)$ if $(|\rho_n| - 1)/(|\phi_{2n}| - 1) = O(1)$ and $O_p\left(n^{-2}(\phi_{2n} - 1)^{-3/2}(n/\kappa_n)^{3/2}\rho_n^{-n}\right) = o_p\left(n^{-1/2}\right)$ if $(|\rho_n| - 1)/(|\phi_{2n}| - 1) \to \infty$. We conclude that the second and the third terms of (B.42) are $o_p(1)$. For the first term of (B.42),

$$\begin{aligned} s_n^{-1} g_n \sum_{t=1}^n \phi_{2n}^t x_{0t} &= s_n^{-1} g_n \sum_{t=1}^n \phi_{2n}^t \sum_{j=1}^t \rho_n^{t-j} u_j = s_n^{-1} g_n \sum_{j=1}^n \rho_n^{-j} u_j \sum_{t=j}^n (\phi_{2n}\rho_n)^t \\ &= \frac{s_n^{-1} g_n}{\phi_{2n}\rho_n - 1}\left(\phi_{2n}^{n+1}\sum_{j=1}^n \rho_n^{n-j+1} u_j - \sum_{j=1}^n \phi_{2n}^j u_j\right) \\ &= s_n^{-1} g_n O_p\left(\frac{\phi_{2n}^{n+1}\nu_n}{\varphi_{2n}\rho_n - 1}\right) = O_p\left(\left(\phi_{2n}^2 - 1\right)^{1/2} g_n\right) = O_p\left(n^{-1/2}(\phi_{2n} - 1)^{-1/2}\right) \end{aligned}$$

which shows that the left side of (B.42) is $o_p(1)$. When $n|\phi_{2n} - \rho_n| \to \infty$, (B.4) gives

$$\begin{aligned} |R'_{3n}| &\le s_n^{-1}\frac{|\hat{\rho}_n - \rho_n|}{|\phi_{2n} - \rho_n|}\left(\left|\sum_{t=1}^n z_{2t-1}x_t\right| + \left|\sum_{t=1}^n x_{t-1}x_t\right|\right) \\ &\le b s_n^{-1}\frac{|\hat{\rho}_n - \rho_n|}{|\phi_{2n} - \rho_n|}\left(\left|\sum_{t=1}^n z_{2t-1}x_{t-1}\right| + \sum_{t=1}^n x_{t-1}^2\right) \qquad \text{(B.43)} \end{aligned}$$

for all but finitely many n for some $b > 0$, because (13) gives $\left|\sum_{t=1}^n z_{2t-1}x_t\right| \le |\rho_n|\left|\sum_{t=1}^n z_{2t-1}x_{t-1}\right| + |\mu(\rho_n - 1)|\left|\sum_{t=1}^n z_{2t-1}\right| + \left|\sum_{t=1}^n z_{2t-1}u_t\right|$, the first term on the right side dominates the other two terms as $n \to \infty$ and a similar inequality holds for $\left|\sum_{t=1}^n x_{t-1}x_t\right|$ with $\sum_{t=1}^n x_{t-1}^2$ dominating. By Lemma 4(i), $\left|\sum_{t=1}^n z_{2t-1}x_{t-1}\right| = O_p(s_n)$ so using the orders in (B.3), the first term on the right of (B.43) is $O_p\left(\frac{\kappa_n^{-1}\rho_n^{-n}}{|\phi_{2n} - \rho_n|}\right) = O_p\left(\frac{1}{n|\phi_{2n} - \rho_n|}\right)$ under C(ii)-C(iii) (under C(iii) by Lemma B2(i)). The second term on the right of (B.43) is $O_p\left(\kappa_n^{1/2}\nu_{n,z}^{-1}\right) = o_p\left((n(\phi_{2n} - 1))^{1/2}|\phi_{2n}|^{-n}\right) = o(1)$ by the orders in (B.3) and Lemma B2(i). This proves that $R'_{3n} = o_p(1)$. For R''_{3n}, using (6), and the computation $\sum_{t=1}^n \rho_n^t z_{2t} = \sum_{j=1}^n \rho_n^j u_j \sum_{t=0}^{n-j}(\rho_n\phi_{2n})^t = (\rho_n\phi_{2n} - 1)^{-1}\left(\phi_{2n}\rho_n^n z_{2n} - \sum_{j=1}^n \rho_n^j u_j\right)$, we obtain

$$\begin{aligned} R''_{3n} &= \mu s_n^{-1} n\bar{z}_{2n} + (X_0(n) - \mu) s_n^{-1}\sum_{t=1}^n \rho_n^t z_{2t} \\ &= \mu s_n^{-1}(\phi_{2n} - 1)^{-1} z_{2n} + (X_0(n) - \mu)\nu_{n,z}^{-1}\nu_n^{-1}\left(\phi_{2n}\rho_n^n z_{2n} - \sum_{j=1}^n \rho_n^j u_j\right) + o_p(1) \\ &= o_p\left(\kappa_n^{-1/2}(n/\kappa_n)|\rho_n|^{-n}\right) + X_0(n) O_p\left(\kappa_n^{-1/2}\right) = o_p(1) \end{aligned}$$

by Assumption 3. This shows $R_{3n} = o_p(1)$. Since $\sum_{t=1}^n z_{2t-1}x_{t-1} = \sum_{t=1}^n z_{2t-1}^+ x_{t-1}^+ = O_p(s_n)$, $\sum_{t=1}^n \rho_n^t z_{2t}^- = \sum_{t=1}^n |\rho_n|^t z_{2t}^+$ and $\bar{z}_{2n}^- = o_p(\bar{z}_{2n})$, the same bounds for R'_{3n} and R''_{3n} show that $\left|R_{3n}^-\right| = o_p(1)$.

For R_{4n}, recalling that $r_{nt} = z_{2t} - \tilde{z}_{2t}$, the identity $\tilde{z}_{2t}^2 - z_{2t}^2 = r_{nt}^2 + 2z_{2t}r_{nt}$ gives

$$\begin{aligned} R_{4n} &\le \left(\phi_{2n}^2 - 1\right)^2|\phi_{2n}|^{-2n}\left(\sum_{t=1}^n r_{nt}^2 + 2\left|\sum_{t=1}^n z_{2t}r_{nt}\right|\right) \\ &\le \left(\phi_{2n}^2 - 1\right)^2|\phi_{2n}|^{-2n}\sum_{t=1}^n r_{nt}^2 + O_p(1)\left\{\left(\phi_{2n}^2 - 1\right)^2|\phi_{2n}|^{-2n}\sum_{t=1}^n r_{nt}^2\right\}^{1/2} \end{aligned}$$

because the Cauchy-Schwarz inequality gives

$$\left(\phi_{2n}^2 - 1\right)^2|\phi_{2n}|^{-2n}\left|\sum_{t=1}^n z_{2t}r_{nt}\right| \le \left(\left(\phi_{2n}^2 - 1\right)^2|\phi_{2n}|^{-2n}\sum_{s=1}^n z_{2s}^2\right)^{1/2}\left\{\left(\phi_{2n}^2 - 1\right)^2|\phi_{2n}|^{-2n}\sum_{t=1}^n r_{nt}^2\right\}^{1/2}$$

and $\left(\phi_{2n}^2-1\right)^2|\phi_{2n}|^{-2n}\sum_{s=1}^n z_{2s}^2=O_p(1)$. We conclude that

$$R'_{4n}=\left(\phi_{2n}^2-1\right)^2|\phi_{2n}|^{-2n}\sum_{t=1}^n r_{nt}^2=o_p(1)$$

is sufficient to show that $R_{4n}=o_p(1)$. Using the identity (B.4) and the inequality $(a+b)^2\leq 2\left(a^2+b^2\right)$ we obtain

$$\begin{aligned} R'_{4n} &\leq 4\left(\frac{\hat{\rho}_n-\rho_n}{\phi_{2n}-\rho_n}\right)^2\left(\phi_{2n}^2-1\right)^2|\phi_{2n}|^{-2n}\left(\phi_{2n}^2\textstyle\sum_{t=1}^n z_{2t-1}^2+\rho_n^2\sum_{t=1}^n x_{t-1}^2\right)\mathbf{1}\left\{n\,|\phi_{2n}-\rho_n|\to\infty\right\} \\ &\quad +2g_n^2\left(\phi_{2n}^2-1\right)^2\varphi_{2n}^{-2n}\textstyle\sum_{t=1}^n\phi_{2n}^{2t}. \end{aligned}$$

The last term is $O(n^{-1}\left(\phi_{2n}^2-1\right)^{-1})=o(1)$; using (B.3), the second term is $O_p(\left(\frac{\phi_{2n}^2-1}{\phi_{2n}-\rho_n}\right)^2|\phi_{2n}|^{-2n})=O_p((n(\phi_{2n}-\rho_n))^{-2}(n(\phi_{2n}-1))^2|\phi_{2n}|^{-2n})=o_p((n(\phi_{2n}-\rho_n))^{-2})=o_p(1)$ by Lemma B2(i). By Lemma 4(i) and (B.3), the first term is $O_p(\kappa_n^{-2}\rho_n^{-2n}(\phi_{2n}-\rho_n)^{-2})$ which is $O_p((n(\phi_{2n}-\rho_n))^{-2})$ under C(ii) and $O_p((n/\kappa_n)^2\rho_n^{-2n}n^{-2}(\phi_{2n}-\rho_n)^{-2})=o_p(n^{-2}(\phi_{2n}-\rho_n)^{-2})$ under C(iii) by Lemma B2(i). The above shows that $R'_{4n}=o_p(1)$ and $R_{4n}=o_p(1)$. Since $\sum_{t=1}^n\left(z_{2t}^-\right)^2=\sum_{t=1}^n\left(z_{2t}^+\right)^2$, the above bounds for R_{4n} and R'_{4n} show that $R_{4n}^-=o_p(1)$. This completes the proof of part (iv).

For part (v), we begin by showing that

$$\epsilon_{1n}:=(\varphi_{2n}-1)\sum_{j=1}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-t}c_{t+j}\right)^2,\ \epsilon_{2n}:=(\varphi_{2n}-1)\sum_{j=1}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-(n-t+1)}c_{t+j}\right)^2 \tag{B.44}$$

satisfy $\epsilon_{1n}\to 0$ and $\epsilon_{2n}\to 0$. Choosing $m_n\to\infty$ with $m_n(\varphi_{2n}-1)\to 0$

$$\begin{aligned} \epsilon_{1n} &\leq (\varphi_{2n}-1)\left[\textstyle\sum_{j=m_n}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-t}|c_{t+j}|\right)^2+\sum_{j=1}^{m_n}\left(\sum_{t=1}^n\varphi_{2n}^{-t}|c_{t+j}|\right)^2\right] \\ &\leq (\varphi_{2n}-1)\left[\textstyle\sum_{t=1}^n\varphi_{2n}^{-t}\sum_{j=m_n}^\infty|c_{t+j}|\sum_{s=1}^n\varphi_{2n}^{-s}|c_{s+j}|+m_n\left(\sum_{t=1}^\infty|c_t|\right)^2\right] \\ &\leq \left(\textstyle\sum_{j>m_n}|c_j|\right)^2(\varphi_{2n}-1)\sum_{t=1}^n\varphi_{2n}^{-t}+(\varphi_{2n}-1)\,m_n\left(\sum_{t=1}^\infty|c_t|\right)^2\to 0 \end{aligned}$$

and, since $\sum_{t=1}^n\varphi_{2n}^{-(n-t+1)}=\sum_{t=1}^n\varphi_{2n}^{-t}$, the above bound applies to ϵ_{2n}. To show part (i) for Z_n, writing $u_t=\sum_{j=1}^t c_{t-j}e_j+\sum_{j=0}^\infty c_{t+j}e_{-j}$ and changing the order of summation of the first sum:

$$\begin{aligned} Z_n &= (\varphi_{2n}-1)^{1/2}\textstyle\sum_{j=1}^n\varphi_{2n}^{-j}\left(\sum_{t=0}^{n-j}\varphi_{2n}^{-t}c_t\right)e_j+(\varphi_{2n}-1)^{1/2}\sum_{j=0}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-t}c_{t+j}\right)e_{-j} \\ &= \left(\textstyle\sum_{t=0}^n\varphi_{2n}^{-t}c_t\right)(\varphi_{2n}-1)^{1/2}\sum_{j=1}^n\varphi_{2n}^{-j}e_j-Z_{1n}+Z_{2n} \end{aligned} \tag{B.45}$$

where $Z_{1n}=(\varphi_{2n}-1)^{1/2}\sum_{j=1}^n\varphi_{2n}^{-j}\left(\sum_{t=n-j+1}^n\varphi_{2n}^{-t}c_t\right)e_j$ and $Z_{2n}=(\varphi_{2n}-1)^{1/2}\sum_{j=0}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-t}c_{t+j}\right)$ satisfy $\mathbb{E}\left(Z_{2n}^2\right)\leq\mathbb{E}\left(e_1^2\right)\epsilon_{1n}\to 0$ by (B.44) and

$$\mathbb{E}\left(Z_{1n}^2\right)\leq\mathbb{E}\left(e_1^2\right)\left(\sum_{t=1}^\infty|c_t|\right)^2n(\varphi_{2n}-1)\varphi_{2n}^{-2(n+1)}\to 0$$

by Lemma B2(i). Since $\varphi_{2n}\to 1$ and $\sum_{t=0}^\infty|c_t|<\infty$, $\sum_{t=0}^n\varphi_{2n}^{-t}c_t\to C(1)$ by the dominated convergence theorem and $Z_n-\check{Z}_n\to_p 0$ follows from (B.45). A similar computation to that used for Z_n yields

$$Y_n=\left(\varphi_{2n}^2-1\right)^{1/2}\left(\sum_{j=1}^n\varphi_{2n}^{-j}e_{n-j+1}\sum_{t=0}^{j-1}\varphi_{2n}^t c_t+\sum_{j=0}^\infty e_{-j}\sum_{t=1}^n\varphi_{2n}^{-(n-t)-1}c_{t+j}\right)=Y_{1n}+Y_{2n}$$

in order of appearance, with $\mathbb{E}\left(Y_{2n}^2\right)\leq\mathbb{E}\left(e_1^2\right)\epsilon_{2n}\to 0$ by (B.44). Since

$$Y_{1n}-\check{Y}_n=\left(\varphi_{2n}^2-1\right)^{1/2}\sum_{j=1}^n\varphi_{2n}^{-j}e_{n-j+1}\left(\sum_{t=0}^{j-1}\left(\varphi_{2n}^t-1\right)c_t-\sum_{t=j}^\infty c_t\right),$$

we will show that $\left\|Y_{1n}-\check{Y}_n\right\|_{L_2}\to 0$ by showing that

$$p_n=\left(\varphi_{2n}^2-1\right)\sum_{j=1}^n\varphi_{2n}^{-2j}\left(\sum_{t=0}^{j-1}\left(\varphi_{2n}^t-1\right)c_t\right)^2\to 0. \tag{B.46}$$

Applying the mean value theorem to the increasing function $x\mapsto\varphi_{2n}^x$ around $(0,t)$ we obtain

$$\varphi_{2n}^t-1\leq t\varphi_{2n}^t\log\varphi_{2n} \tag{B.47}$$

and note that $\log\varphi_{2n}\to 0$ since $\varphi_{2n}\to 1$. Choosing a sequence $m_n\to\infty$ and $m_n\log\varphi_{2n}\to 0$,

$$\begin{aligned}
p_n &\leq \left(\varphi_{2n}^2-1\right)(\log\varphi_{2n})^2\sum_{j=1}^n\varphi_{2n}^{-2j}\left(\sum_{t=1}^{j-1}t\varphi_{2n}^tc_t\right)^2\\
&= \left(\varphi_{2n}^2-1\right)(\log\varphi_{2n})^2\sum_{t=1}^{n-1}t\varphi_{2n}^tc_t\sum_{s=1}^{n-t-1}s\varphi_{2n}^sc_s\sum_{j=1}^{n-t-s}\varphi_{2n}^{-2j-2t-2s}\\
&\leq \left(\log\varphi_{2n}\sum_{t=1}^{n-1}t\varphi_{2n}^{-t}|c_t|\right)^2\left(\varphi_{2n}^2-1\right)\sum_{j=1}^n\varphi_{2n}^{-2j}\\
&\leq \left(\log\varphi_{2n}\sum_{t=m_n}^{n-1}t\varphi_{2n}^{-t}|c_t|\right)^2O(1)+O\left((m_n\log\varphi_{2n})^2\right)\\
&\leq \left(\sum_{t=m_n}^{n-1}|c_t|\right)^2O(1)+O\left((m_n\log\varphi_{2n})^2\right)
\end{aligned}$$

since $\varphi_{2n}^{-t}t\log\varphi_{2n}=\left(\log\varphi_{2n}^t\right)/\varphi_{2n}^t\leq 1$ from the inequality $\log x\leq x$ for $x\geq 1$. This proves (B.46) and completes the proof of part (iv).

Proof of Corollary 1. For the last two t-statistics in (31) $T_n(\tilde\gamma_n),T_n\left(\tilde\delta_n\right)\to_d\mathcal{N}(0,1)$ follow directly from Theorem 1 by putting $v_t=u_{2t}/\sigma_\gamma$ and $v_t=u_{3t}/\sigma_\delta$ in (A.16). For $T_n(\tilde r_n)$, write

$$\begin{aligned}
\tilde r_n-r_0 &= \frac{\tilde\rho_n-\rho_n}{\tilde\gamma_n+\tilde\delta_n}-\frac{\rho_n-1}{\left(\tilde\gamma_n+\tilde\delta_n\right)(\gamma+\delta)}\left(\tilde\gamma_n-\gamma+\tilde\delta_n-\delta\right)\\
&= \left[\frac{1}{\tilde\gamma_n+\tilde\delta_n},-\frac{\rho_n-1}{\left(\tilde\gamma_n+\tilde\delta_n\right)^2},-\frac{\rho_n-1}{\left(\tilde\gamma_n+\tilde\delta_n\right)^2}\right]\begin{bmatrix}\tilde\rho_n-\rho_n\\ \tilde\gamma_n-\gamma\\ \tilde\delta_n-\delta\end{bmatrix}\\
&= \hat\upsilon_n'\frac{\sum_{t=1}^n\tilde z_{t-1}u_t}{\sum_{t=1}^nI_{t-1}\tilde z_{t-1}}
\end{aligned}$$

where $\hat\upsilon_n=\left[1/\left(\hat\gamma_n+\hat\delta_n\right),-(\hat\rho_n-1)/\left(\hat\gamma_n+\hat\delta_n\right)^2,-(\hat\rho_n-1)/\left(\hat\gamma_n+\hat\delta_n\right)^2\right]'$ and $u_t=(u_{1t},u_{2t},u_{3t})'$ in the notation of Assumption 5. Under Assumption 5, $\hat\upsilon_n\to_p\upsilon=\left[\frac{1}{\gamma+\delta},-\frac{\rho-1}{(\gamma+\delta)^2},-\frac{\rho-1}{(\gamma+\delta)^2}\right]'$, $\hat\Sigma_n\to_p\Sigma>0$ and $\hat\sigma_{r_0}^2\to_p\upsilon'\Sigma\upsilon$; hence $T_n(\tilde r_n)=[1+o_p(1)]T_n$ with T_n given by (A.16) with $v_t=\upsilon'u_t/(\upsilon'\Sigma\upsilon)^{1/2}$. By Assumption 5, (v_t) satisfies Assumption 2 with $\mathbb{E}_{\mathcal{F}_{t-1}}\left(v_t^2\right)=1$ *a.s.*, so $T_n\to_d\mathcal{N}(0,1)$ by Theorem 1.

For $T_n\left(\tilde\theta_n\right)$, denoting $\iota=(1,1,1)'$, and employing the identity $\tilde\theta_n=\tilde\rho_n+\tilde\gamma_n+\tilde\delta_n$ and a similar argument to $T_n(\tilde r_n)$ we obtain $T_n\left(\tilde\theta_n\right)=[1+o_p(1)]T_n$ with T_n given by (A.16) with $v_t=\iota'u_t/(\iota'\Sigma\iota)^{1/2}$. By Assumption 5, (v_t) satisfies Assumption 2 with $\mathbb{E}_{\mathcal{F}_{t-1}}\left(v_t^2\right)=1$ *a.s.*, so $T_n\to_d\mathcal{N}(0,1)$ by Theorem 1.

1.3 Additional Simulation Results

In this section, we present some additional simulation results. Tables B1 and B2 below contain the empirical size and Figure B1 displays the power of the two-sided test of our procedure for the predictive regression slope parameter β for $n = 1,000$ based on $10,000$ replications for a grid of points for b_1 and b_2 for $\rho_{\varepsilon u} = 0.99$ and $\rho_{\varepsilon u} = -0.99$ respectively for the case $\rho = 1$, which we use for the instrument selection of Section 4.1 of the main paper[16].

Table B1: Empirical size, $\rho_{\varepsilon u}$=0.99,n=1,000

b_1/b_2	0.650	0.675	0.700	0.725	0.750	0.775	0.800	0.825	0.850	0.875	0.900	0.925	0.950
0.650	5.01%	5.25%	5.76%	5.62%	5.63%	6.35%	6.52%	6.63%	6.16%	5.44%	5.80%	5.80%	5.98%
0.675	5.17%	5.46%	5.71%	5.52%	6.03%	5.87%	6.73%	6.63%	6.02%	5.74%	5.84%	5.94%	6.02%
0.700	5.53%	5.39%	5.61%	5.73%	6.18%	6.70%	6.69%	6.66%	6.13%	5.73%	5.79%	5.93%	6.35%
0.725	5.42%	5.51%	5.95%	5.59%	6.00%	6.72%	6.75%	6.45%	6.30%	6.20%	5.80%	5.78%	6.25%
0.750	5.33%	5.48%	5.71%	6.08%	6.03%	6.46%	6.97%	6.91%	5.70%	5.95%	5.79%	6.28%	6.34%
0.775	5.65%	5.67%	5.44%	5.66%	6.13%	6.48%	6.98%	6.62%	6.01%	5.92%	5.85%	6.04%	6.41%
0.800	5.25%	5.85%	5.56%	6.16%	5.90%	6.90%	6.64%	6.89%	6.61%	5.99%	6.21%	6.29%	5.92%
0.825	5.68%	5.44%	5.80%	6.09%	6.39%	6.83%	7.01%	6.61%	6.37%	5.89%	6.11%	6.48%	6.24%
0.850	5.57%	6.21%	5.45%	6.07%	6.39%	6.78%	7.23%	7.15%	6.35%	5.94%	5.95%	6.19%	6.59%
0.875	5.87%	6.17%	6.00%	6.04%	6.13%	6.41%	6.82%	6.71%	6.60%	6.31%	6.02%	6.56%	6.10%
0.900	5.87%	6.04%	5.77%	6.37%	6.22%	6.84%	6.72%	7.17%	6.69%	5.98%	6.01%	6.06%	7.03%
0.925	6.01%	5.83%	5.78%	6.05%	6.33%	6.83%	7.08%	6.48%	6.60%	6.21%	6.08%	5.98%	6.87%
0.950	6.46%	6.19%	5.92%	6.19%	6.40%	6.24%	7.02%	7.04%	6.70%	6.17%	6.30%	6.86%	7.16%

Figures B2 and B3 contain the empirical size of our two-sided IV-based test for correlation $\rho_{\varepsilon u}$ of -0.45 and 0.45 respectively. Figure B4 displays the proportion of times the different instruments are chosen. Figure B5 is a comparison of the length of CIs of IV and OLS under misspecification of the last observation (note, in this case, OLS has no valid coverage for the purely explosive specifications). Figures B6 and B7 present the coverage and length of CIs of our IV-based CIs and the equal-tailed two-sided intervals (ETCI) of Andrews and Guggenberger (2014) respectively. Figure B8 displays the empirical size of the OLS- and IV-based one-sided test under misspecification of the last observation. Finally, Figures B9/B11 and B10/B12 contain the empirical size and power of our one-sided IV-based test in comparison with the size and power of the Elliott et al. (2015)'s procedure for correlation $\rho_{\varepsilon u}$ of -0.45 and 0.45 respectively.

[16] We place more weight on large values for b_1 rather than large values for b_2 for three reasons: (i) power is always non-decreasing in b_1 for all autoregressive specifications, while in the explosive region power is decreasing in the value of b_2 (though this is not a serious issue since our procedure preserves the exponential rate of convergence in the explosive region $\rho_n^n n^{-b_2/2}$ regardless of the value of b_2), (ii) for power maximisation in the case $\rho = 1$, the value of b_1 is relatively more important (as can be seen from the power plots in Appendix B), since the near-stationary instrument is chosen 2/3 of the time, and (iii) values for b_2 close to unity would make our mildly explosive instrument near the boundary with local-to-unity region, which would cause the instrument to inherit local-to-unity properties and potentially some of the associated small sample distortions when working with purely explosive regressor.

Table B2: Empirical size, $\rho_{\varepsilon u}$=−0.99,n=1,000

b_1/b_2	0.650	0.675	0.700	0.725	0.750	0.775	0.800	0.825	0.850	0.875	0.900	0.925	0.950
0.650	5.88%	5.20%	5.43%	5.72%	5.64%	5.92%	6.38%	6.83%	5.93%	5.48%	6.02%	5.38%	6.38%
0.675	5.26%	5.35%	5.50%	5.57%	6.09%	6.15%	6.53%	6.63%	6.49%	5.47%	5.51%	6.20%	6.05%
0.700	5.81%	5.59%	5.17%	6.00%	5.60%	6.41%	6.65%	6.82%	6.15%	5.95%	5.88%	6.02%	6.09%
0.725	5.31%	5.72%	5.41%	5.80%	6.44%	6.28%	6.67%	6.64%	6.60%	6.17%	5.63%	5.95%	6.68%
0.750	5.75%	5.53%	5.25%	5.84%	6.14%	6.37%	7.18%	6.66%	6.21%	5.74%	6.15%	6.03%	6.28%
0.775	5.51%	5.59%	5.90%	5.93%	6.17%	6.32%	7.00%	6.91%	6.07%	6.09%	5.79%	6.18%	6.22%
0.800	4.97%	5.70%	5.45%	5.67%	5.98%	6.71%	7.06%	6.85%	5.97%	5.93%	6.09%	6.28%	5.97%
0.825	5.89%	5.82%	5.81%	5.59%	6.06%	6.04%	6.56%	6.96%	6.22%	6.00%	5.97%	5.94%	6.67%
0.850	5.79%	5.49%	5.50%	5.69%	6.11%	6.62%	7.07%	6.88%	6.63%	6.26%	6.74%	5.79%	6.63%
0.875	6.10%	5.89%	5.85%	5.98%	6.23%	6.38%	7.03%	6.84%	6.51%	6.16%	6.81%	5.83%	6.57%
0.900	5.45%	6.48%	6.36%	6.25%	6.13%	6.88%	6.85%	6.98%	6.36%	6.66%	6.13%	6.27%	6.29%
0.925	6.11%	5.80%	6.29%	6.19%	6.35%	6.68%	6.65%	7.03%	6.23%	6.54%	6.07%	6.17%	6.65%
0.950	6.21%	6.27%	6.13%	6.42%	6.40%	6.71%	6.94%	6.64%	6.69%	6.23%	6.95%	6.46%	7.19%

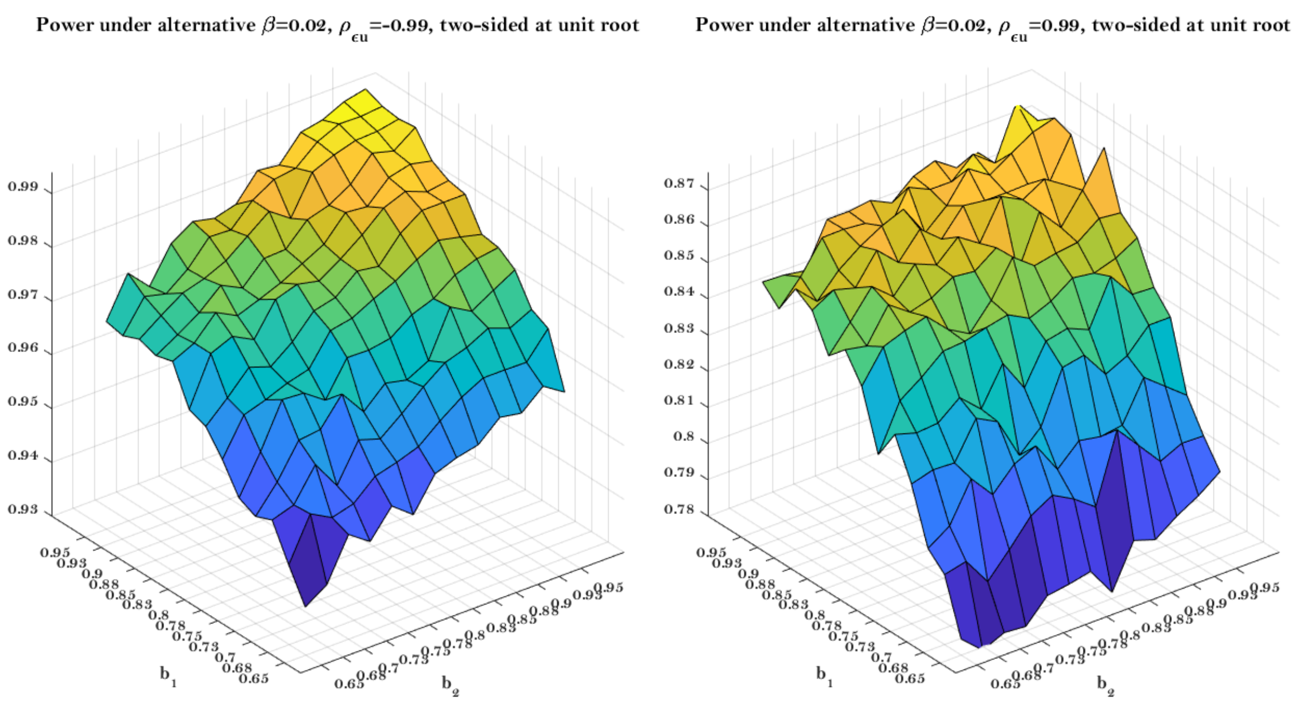


Figure B1: Power at $\rho = 1$ over a grid for b_1 and b_2

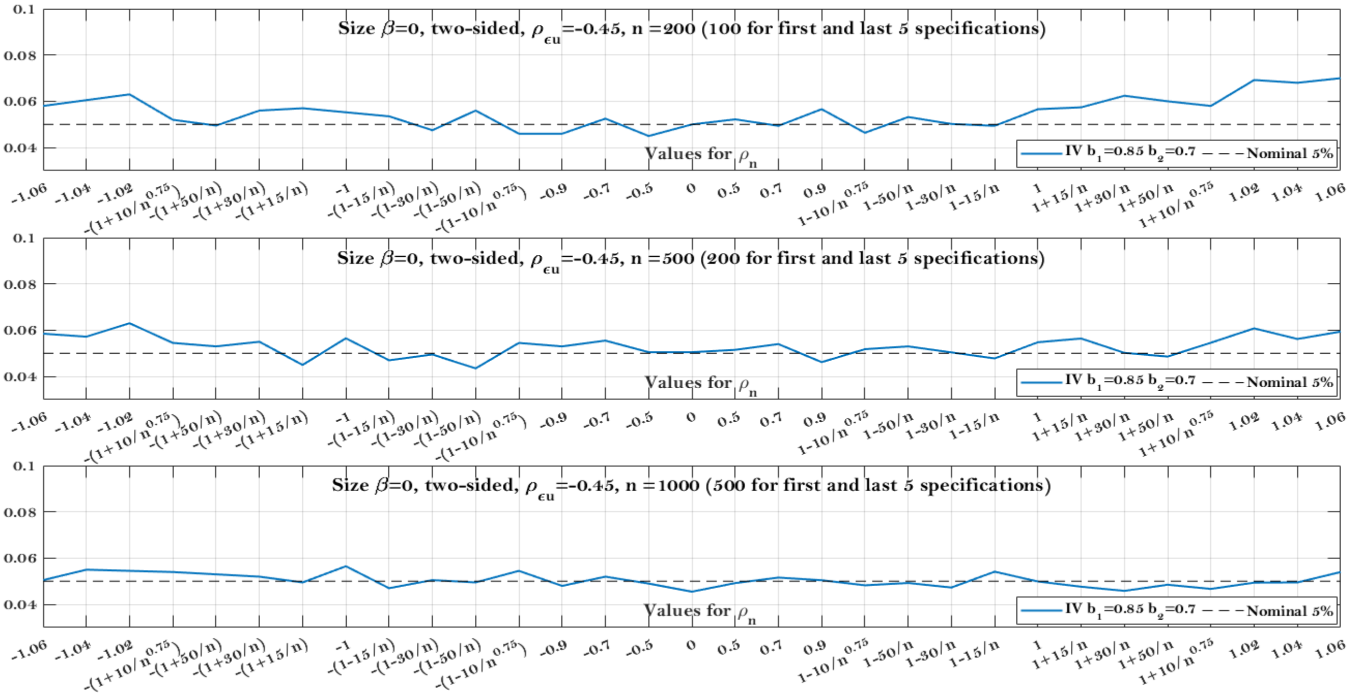


Figure B2: Empirical size of the two-sided test on β, $\rho_{\varepsilon u} = -0.45$

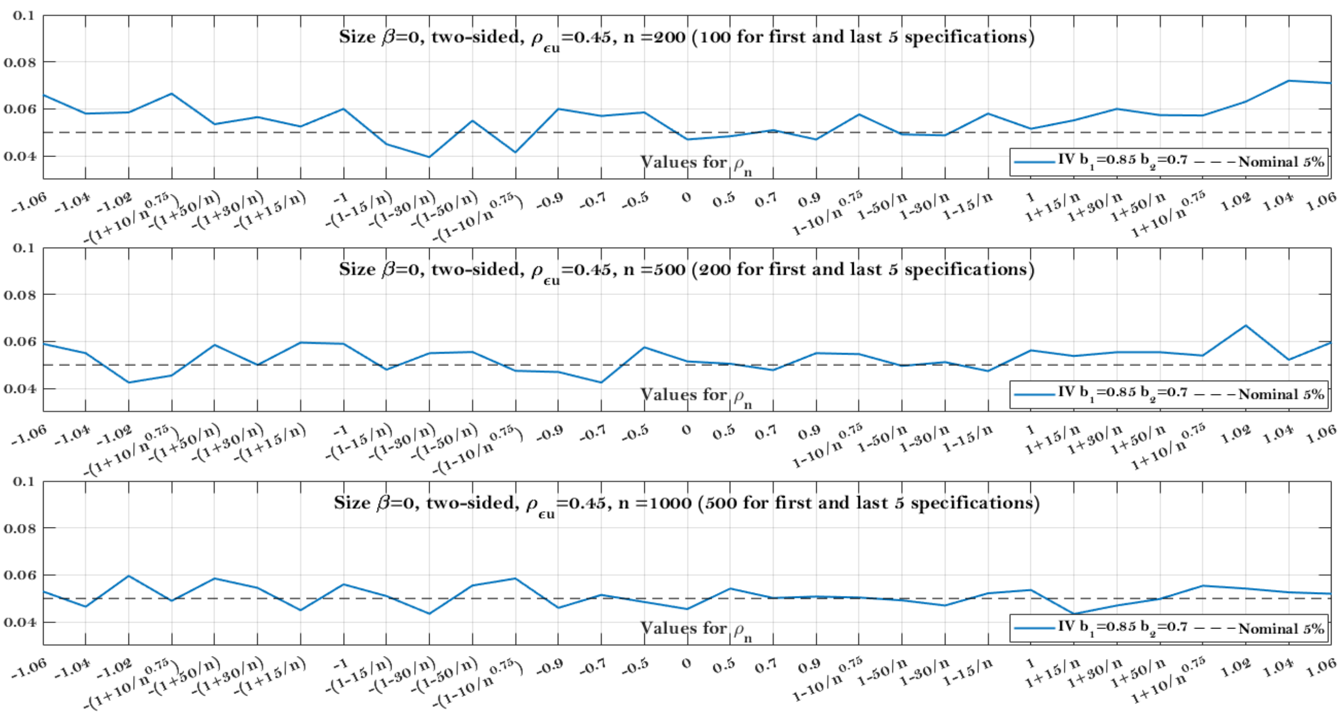


Figure B3: Empirical size of the two-sided test on β, $\rho_{\varepsilon u} = 0.45$

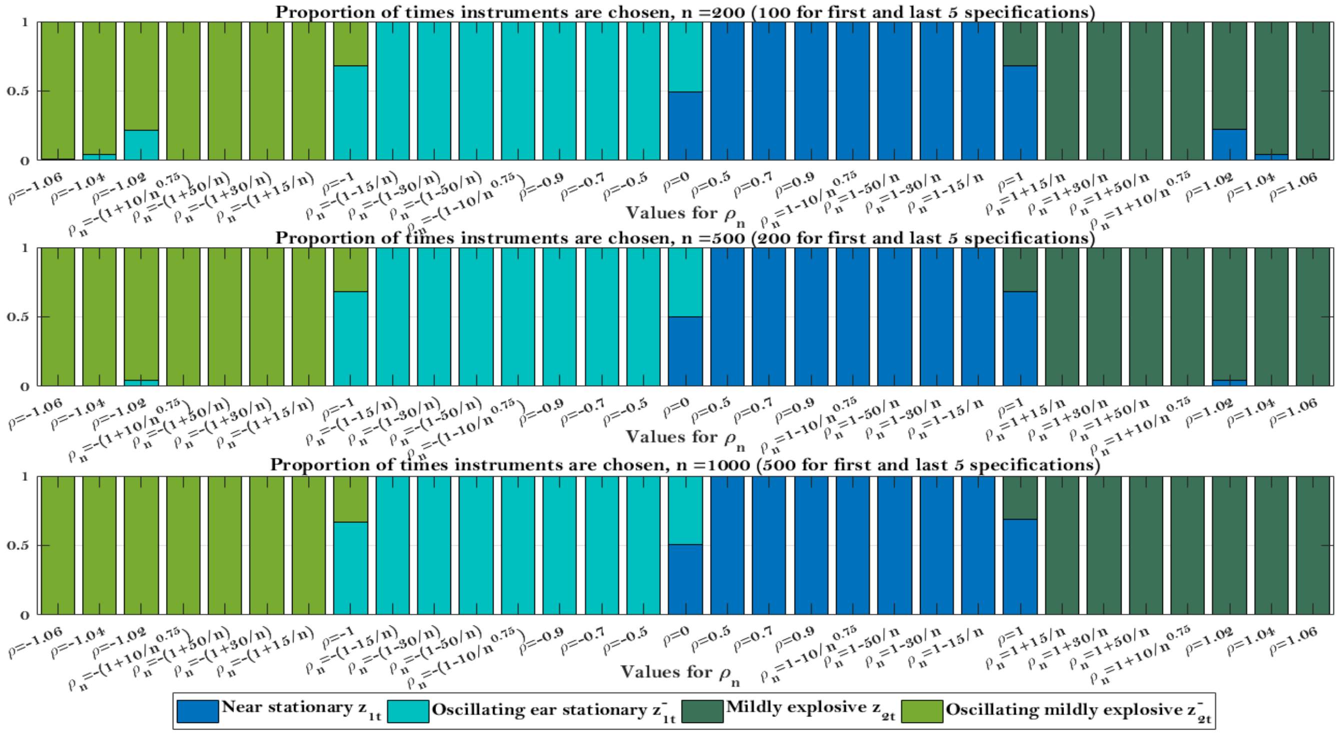


Figure B4: Proportion of times $\tilde{z}_{1t}$, $\tilde{z}^-_{1t}$, $\tilde{z}_{2t}$ and $\tilde{z}^-_{2t}$ are chosen

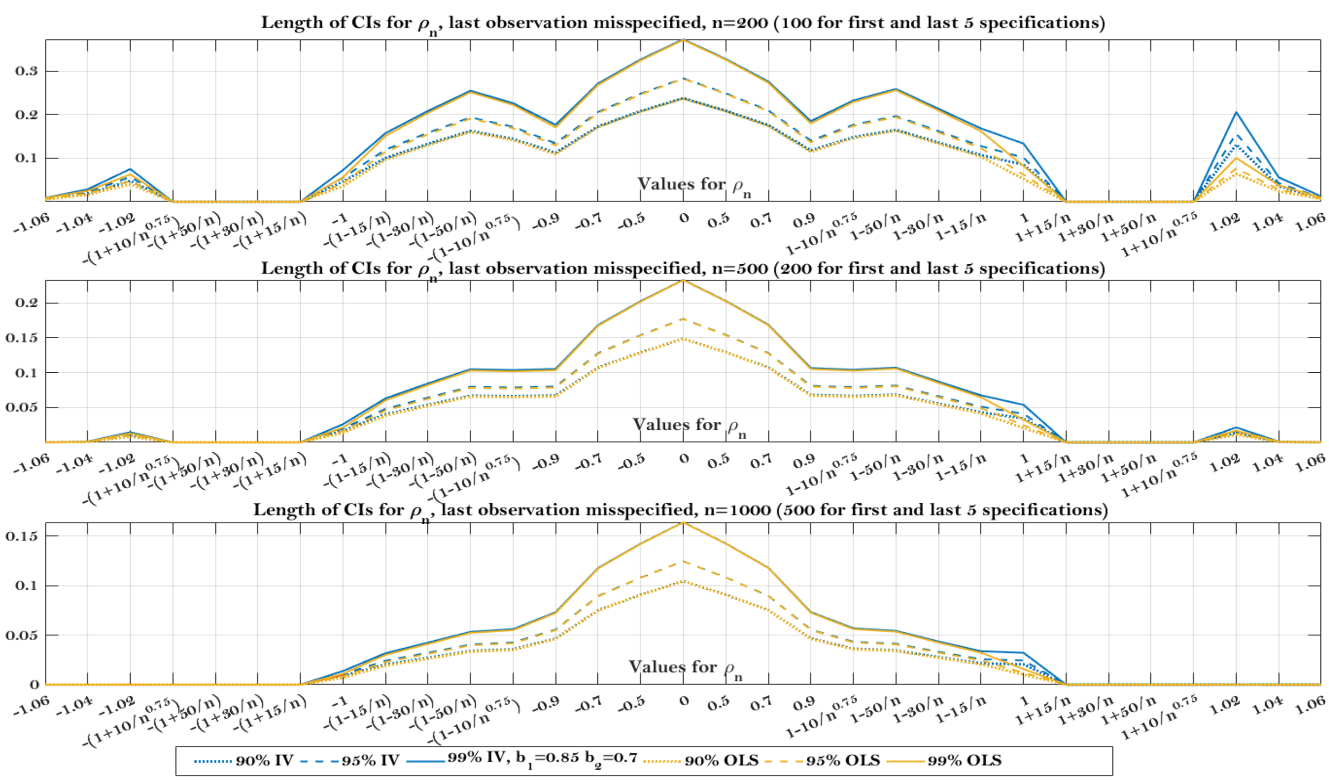


Figure B5: Length of intervals of IV and OLS under misspecification of the last observation

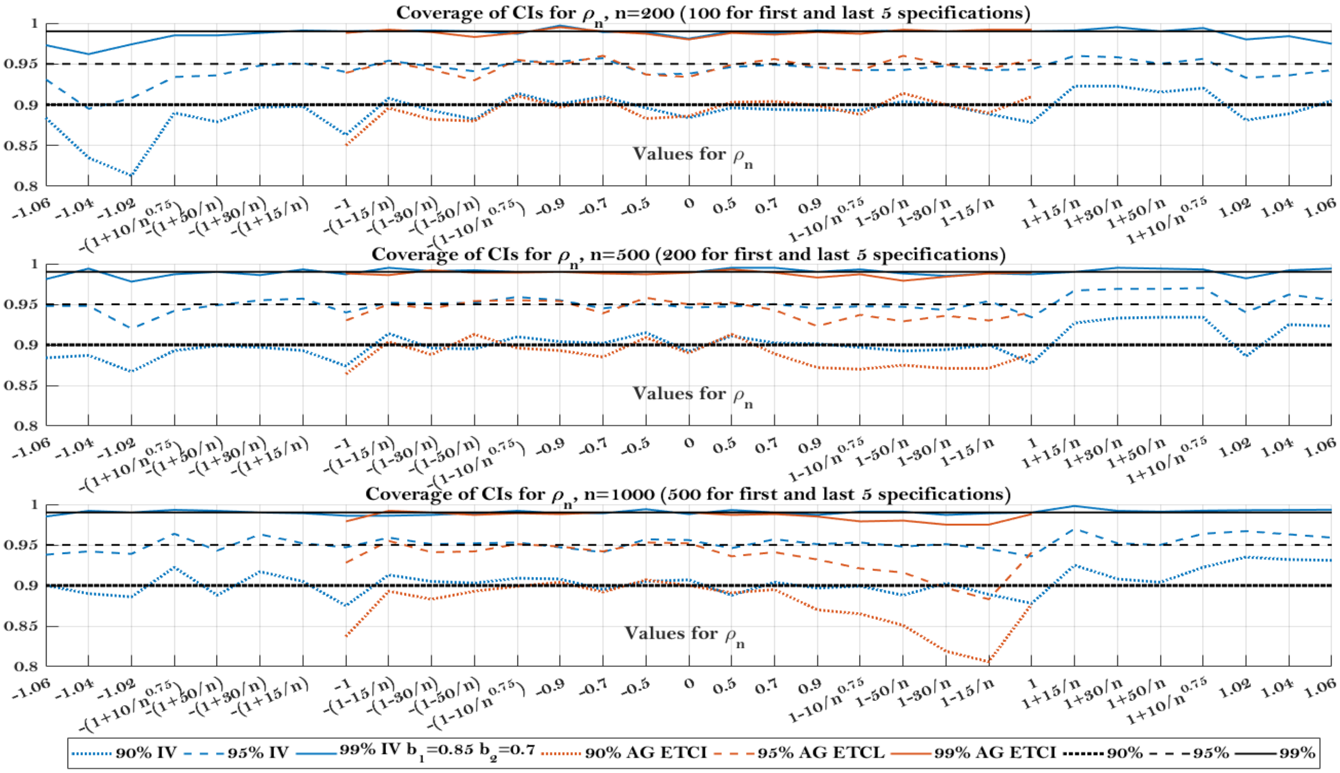


Figure B6: Coverage of CIs of IV and ETCI of Andrews and Guggenberger (2014)

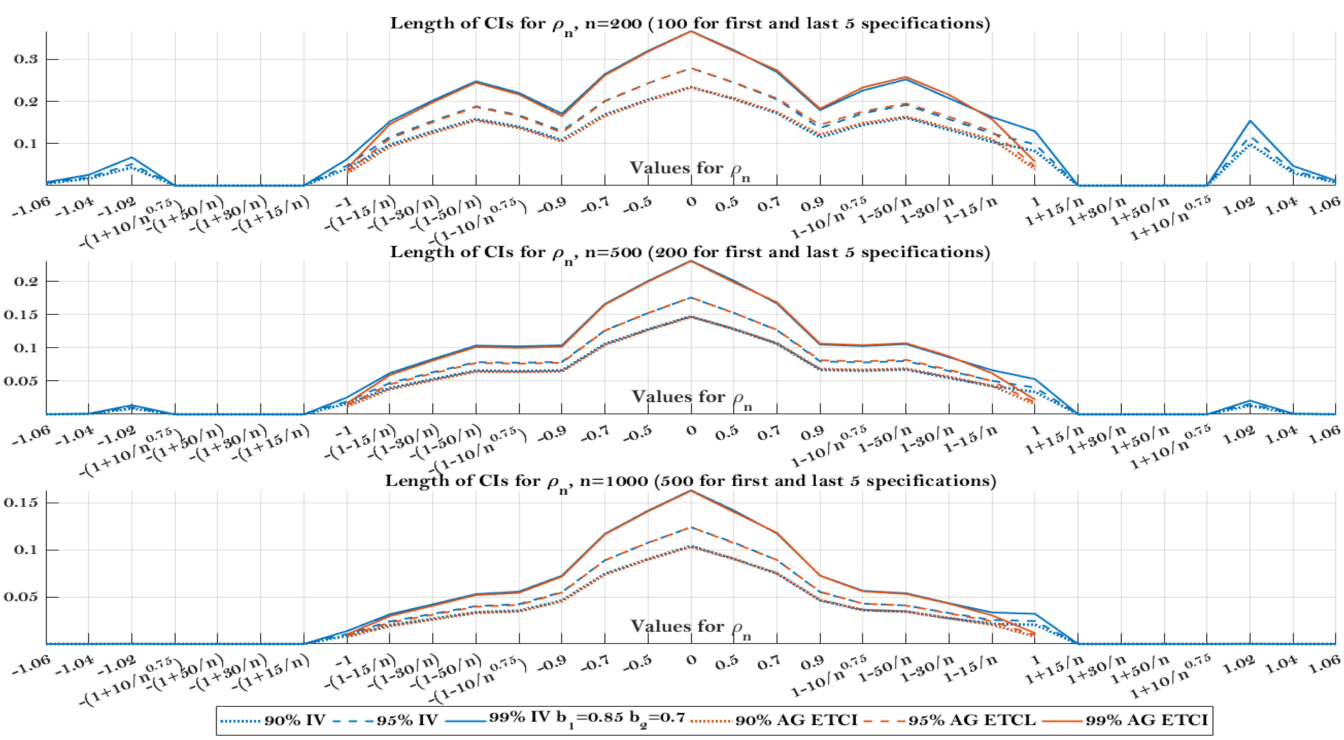


Figure B7: Length of CIs of IV and ETCI of Andrews and Guggenberger (2014)

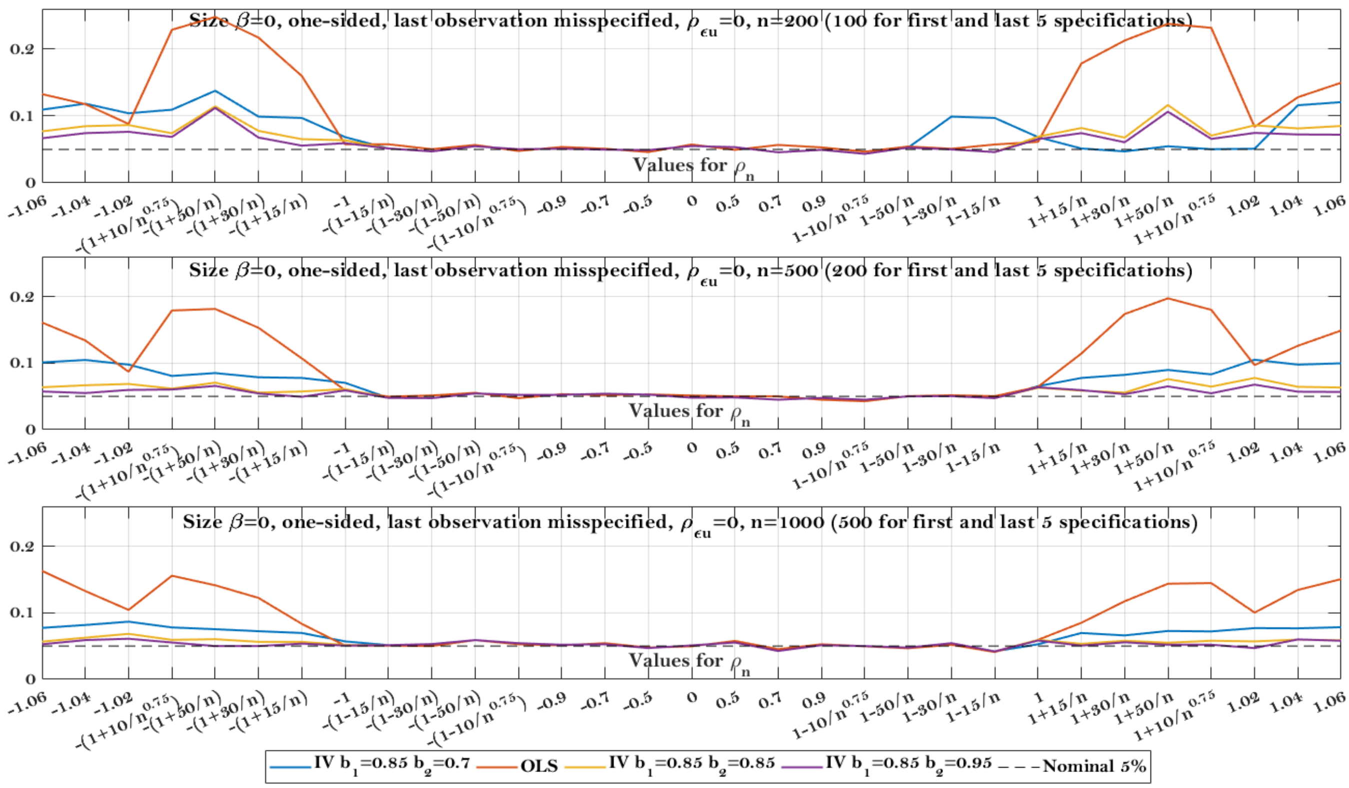


Figure B8: Size of OLS- and IV-based one-sided test under misspecification of the last observation

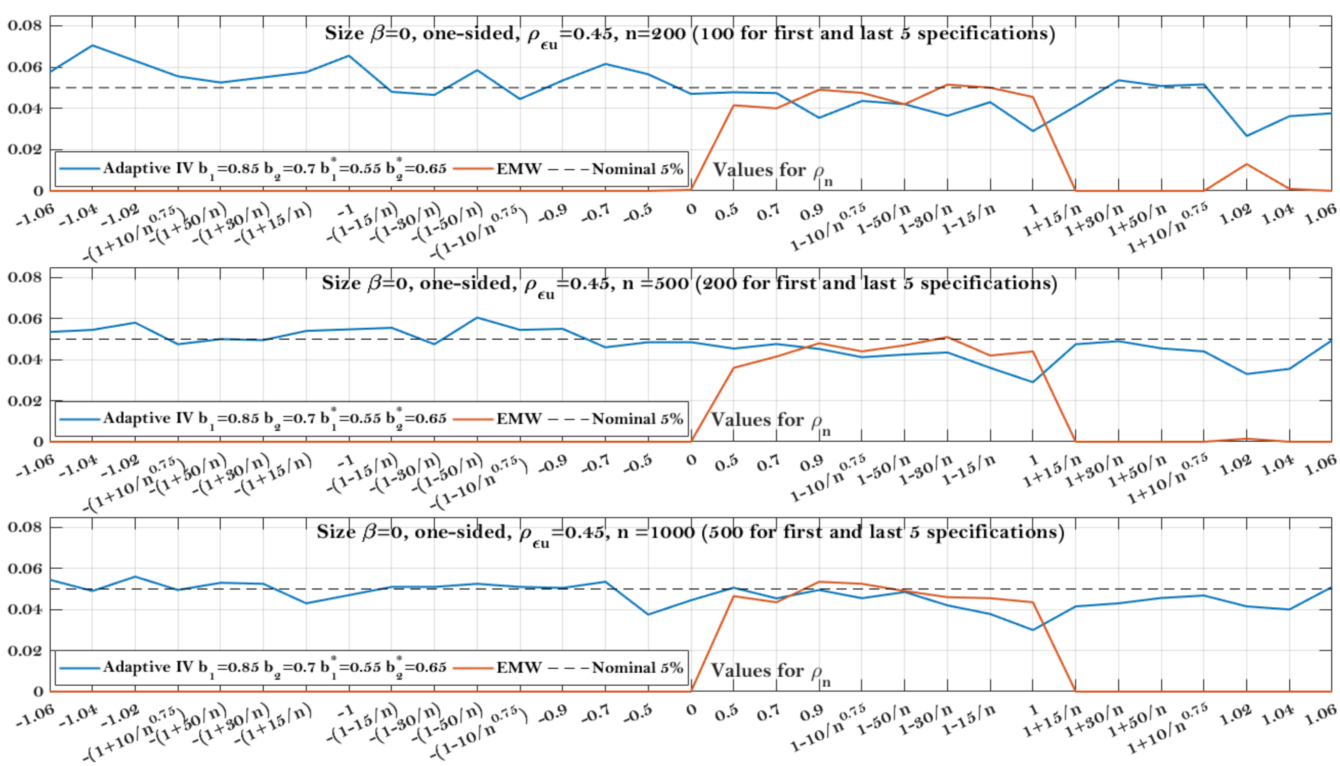


Figure B9: Empirical size of the one-sided test on β, $\rho_{\varepsilon u} = 0.45$

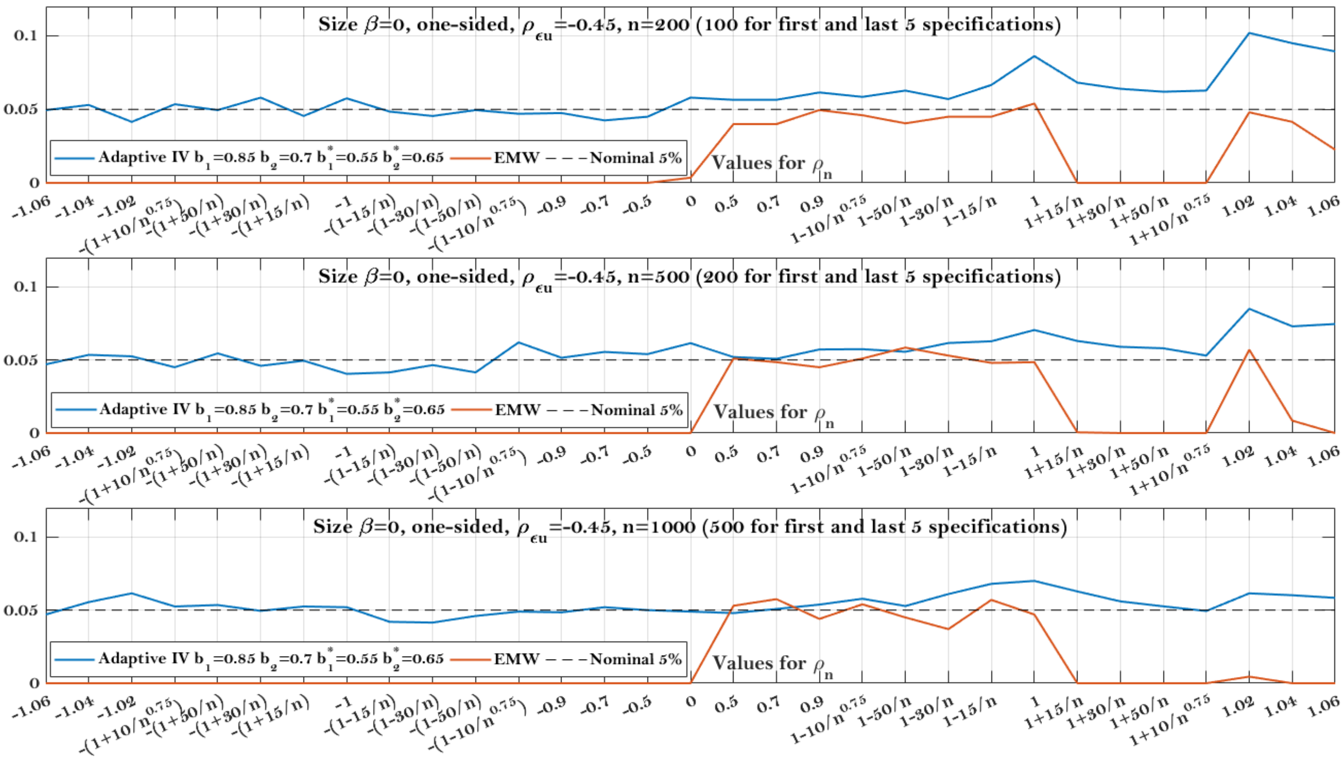


Figure B10: Empirical size of the one-sided test on β, $\rho_{\varepsilon u} = -0.45$

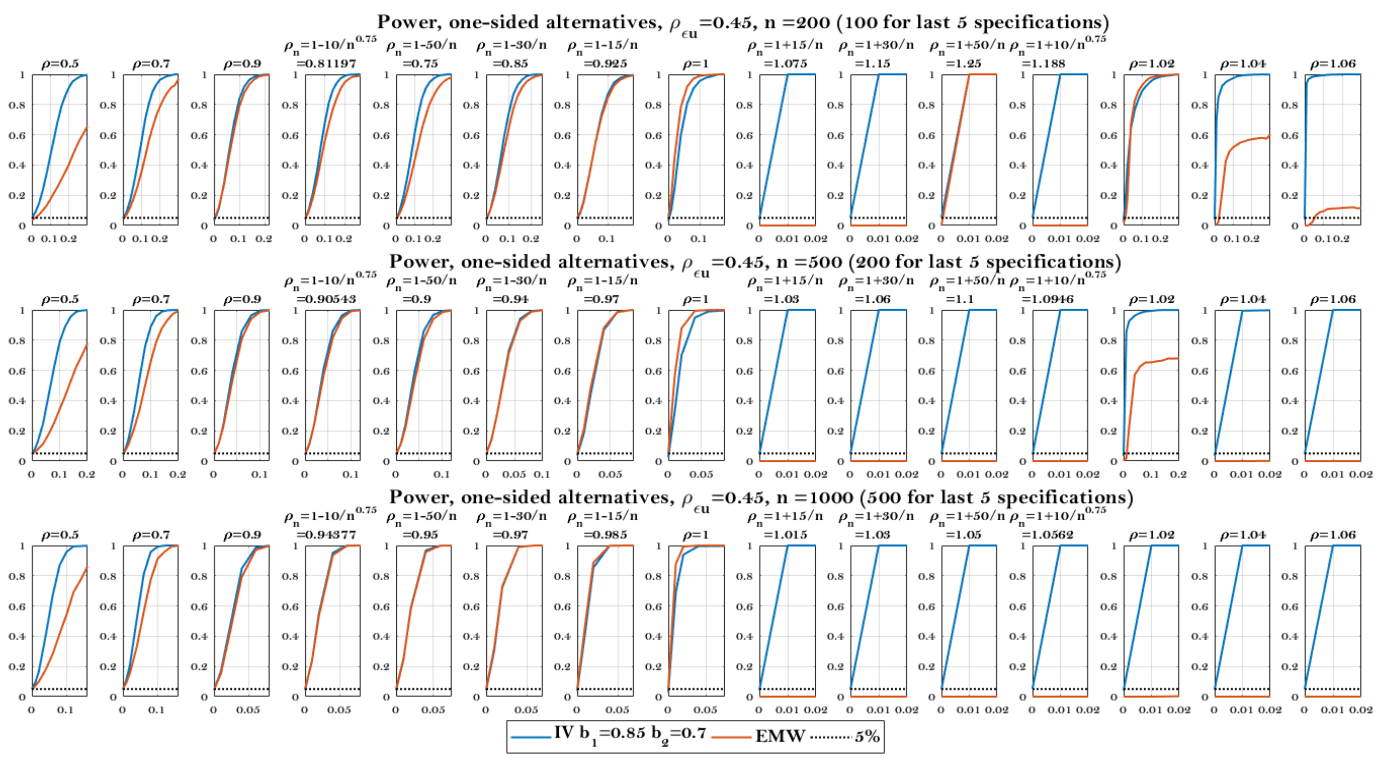


Figure B11: Power of the one-sided test on β, $\rho_{\varepsilon u} = 0.45$

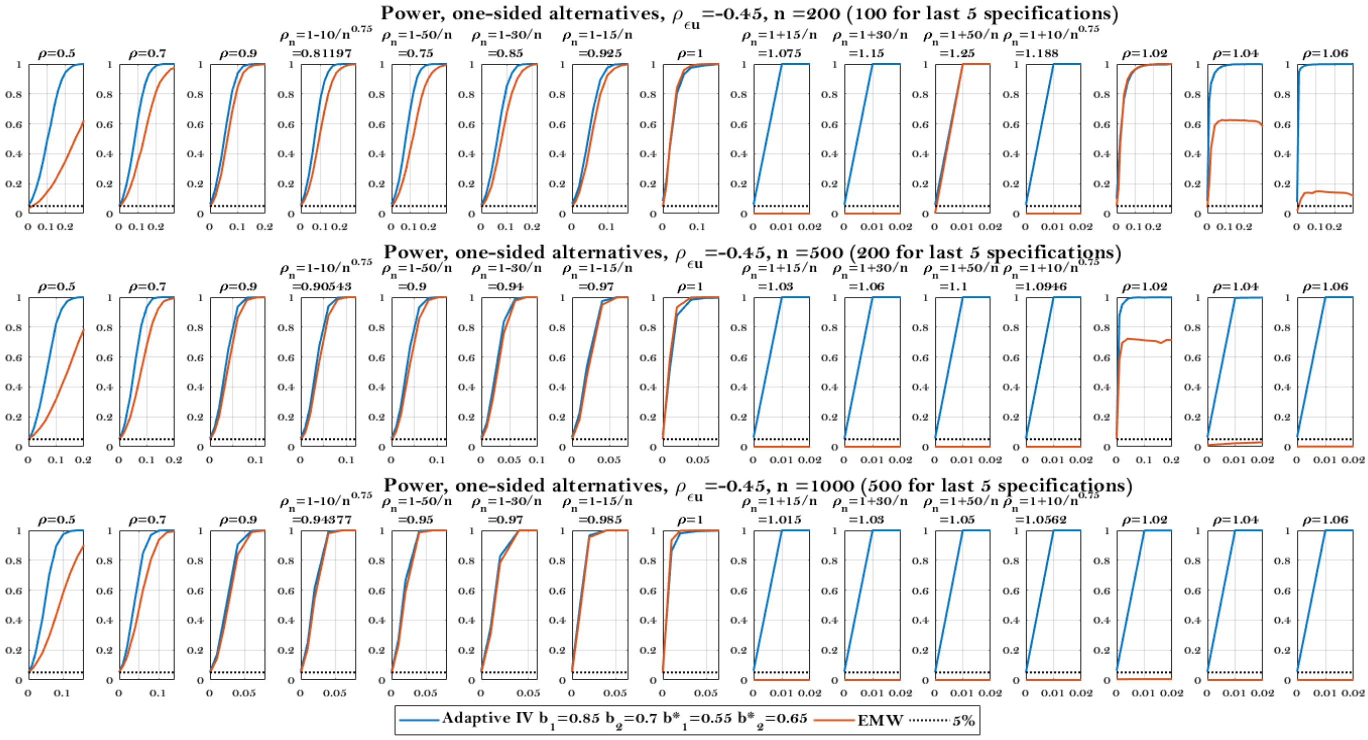


Figure B12: Power of the one-sided test on β, $\rho_{\varepsilon u} = -0.45$

References

Andrews, D.W.K. and Guggenberger, P. (2014) A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter. Review of Economics and Statistics 96(2): 376-381

Elliott, G., Müller, U. K. and Watson, M. W. (2015) Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis, Econometrica, 83(2): 771-811

Hall, P. and Heyde, C.C. (1980) Martingale Limit Theory and its Application. Academic Press.

Giraitis, L. and Phillips, P.C.B. (2006) Uniform limit theory for stationary autoregression. Journal of Time Series Analysis, 27(1): 51-60

Jacod, J. and Shiryaev, A.N. (2003) Limit Theorems for Stochastic Processes (2nd Edition). Springer.

Lai, T.L. and Wei, C.Z. (1983) A note on martingale difference sequences satisfying the local Marcinkiewicz-Zygmund condition. Bulletin of the Institute of Mathematics, Academia Sinica, 11: 1-13

Magdalinos, T. (2012) Mildly explosive autoregression under weak and strong dependence. Journal of Econometrics, 169(2): 179-187

Magdalinos, T. and Phillips, P.C.B. (2020) Econometric inference in matrix vicinities of unity and stationarity, Working paper

Phillips, P.C.B. (1987b) Towards a unified asymptotic theory for autoregression. Biometrika 74: 535–547

Phillips, P.C.B. and Magdalinos, T. (2007) Limit theory for Moderate deviations from a unit root. Journal of Econometrics, 136: 115-130