Universitat Pompeu Fabra Barcelona — Department of Economics and Business

Economics Working Paper Series

Working Paper No. 1837

Uniform and distribution-free inference with general autoregressive processes

Tassos Magdalinos and Katerina Petrova

April 2022

Uniform and distribution-free inference with general autoregressive processes*

Tassos Magdalinos
University of Southampton

Katerina Petrova
Universitat Pompeu Fabra and Barcelona School of Economics

April 22, 2022

Abstract

A unified theory of estimation and inference is developed for an autoregressive process with root in $(-1, \infty)$ 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 region $(1, \infty)$ 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 $(-1, \infty)$ 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 Kirill Evdokimov and Bo Zhou for valuable comments and discussions. Petrova acknowledges support by the Alan Turing Institute under the EPSRC grant EP/N510129/1, 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 infection in early stages of epidemics 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 and hypothesis tests. Early work on obtaining confidence intervals 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 confidence intervals 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 confidence interval 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 Demetersci (2015), Yang, Long, Peng and Cai (2020), Magdalinos and Phillips (2020), Demetersci, 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 confidence intervals with uniform asymptotic validity over the entire autoregressive parameter space $(-1, \infty)$ 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 confidence interval 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 $(-1, \infty)$ independently of the distribution of the innovations of the autoregressive process. The asymptotic mixed-Gaussianity property implies that, upon self-normalisation, the IV-based t-statistic is asymptotically standard normal and can be employed for confidence intervals construction based on standard normal quantiles. Moreover, we show that the proposed confidence intervals have 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 confidence intervals 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 on the right side of unity $(1, \infty)$, where no existing alternative approach has general asymptotic validity.

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 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 confidence intervals 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 modelling framework for 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 theoretical 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

[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 and that, in general, the limit theory of least squares estimation and inference is 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.

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 confidence intervals and hypothesis tests in comparison to the leading existing inference procedures in autoregression and predictive regression. Section 5 applies the confidence intervals of Corollary 1 to Covid-19 infections across a panel of countries and Section 6 concludes. All mathematical proofs are collected in Appendix A. Some auxiliary mathematical results, the proof of Corollary 1 and additional simulation results can be found in the supplementary online Appendix B.

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, with an innovation sequence $(u_t)_{t\in\mathbb{N}}$ and an initialisation X_0. It is easy to see that (1) yields an autoregressive process

$$x_t = \mu(1-\rho_n) + \rho_n x_{t-1} + u_t \tag{2}$$

$$= \mu + (X_0(n) - \mu)\rho_n^t + x_{0t}, \quad x_{0t} = \sum_{j=1}^n \rho_n^{t-j} u_j \tag{3}$$

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).

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* $\rho_n \to \rho \in (-1, \infty)$.

In order to establish an asymptotic theory of estimation (Theorem 3 below), 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* $c := \lim_{n\to\infty} n(\rho_n - 1)$ *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 $c = \infty$.

Stationary autoregressions with fixed root in $(-1, 1)$ as well as near-stationary autoregressions are included in class C(i), pure unit root processes with $c = 0$ as well as local departures from unity are included in C(ii) and explosive (fixed root in $(1, \infty)$) and mildly explosive autoregressions are included in class C(iii). 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 $\rho_n \to \rho > 1$.

[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 conficence 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.

When $\rho = 1$ in Assumption 1a, Assumption 1b is more restrictive than Assumption 1a[4]: for example, the oscillating sequence

$$(\rho_n)_{n\in\mathbb{N}} = 1 + (-1)^n / k_n \quad k_n \to \infty \tag{4}$$

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 $\left(\rho_{m_n}\right)_{n\in\mathbb{N}}$ of $(\rho_n)_{n\in\mathbb{N}}$ there exists a further subsequence $\left(\rho_{s_n}\right)_{n\in\mathbb{N}}$ of $\left(\rho_{m_n}\right)_{n\in\mathbb{N}}$ such that $\left(\rho_{s_n}\right)_{n\in\mathbb{N}}$ 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-normality 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 confidence intervals of Theorems 1 and 2.

Assumption 2 (innovation sequence). *Given a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, the innovation sequence 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. and $(u_t^2)_{t\in\mathbb{Z}}$ is a uniformly integrable sequence. In the explosive case C_0(iii) we assume in addition that*

$$\liminf_{t\to\infty} \mathbb{E}_{\mathcal{F}_{t-1}} |u_t| > 0 \ \ a.s.. \tag{5}$$

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*

$$X_0(n) = \max\left\{O_p(1), o_p\left(\kappa_n^{1/2}\right)\right\}, \ \ \text{where} \ \ \kappa_n := |\rho_n - 1|^{-1} \wedge n. \tag{6}$$

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) as it is of sufficient generality to give rise to uniform asymptotic size for hypothesis tests and confidence intervals over the parameter space defined by Assumptions 1a, 2 and 3. For each $n \in \mathbb{N}$, letting $\mathcal{A}_n = \{(F_n, X_0(n)) : \{u_1, ..., u_n\} \ \text{have c.d.f.} \ F_n, \ F_n \ \text{and} \ X_0(n) \ \text{satisfy Assumptions 2 and 3}\}$ we define the following parameter space for the problem of conducting inference in (1):

$$\Theta = \{(\rho, F, X_0) : \rho \in [-1 + \delta, M] \, (\forall \delta, M > 0) \ \text{and} \ (F, X_0) \in \mathcal{A}\} \tag{7}$$

where $\mathcal{A} := \liminf_{n\to\infty} \mathcal{A}_n = \cup_{n\geq 1} \cap_{j\geq n} \mathcal{A}_j$. The inference procedure developed in the paper gives rise to confidence intervals for the autoregressive parameter in (1) with correct asymptotic coverage probability uniformly over the parameter space Θ in (7).

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 $n^{-1}(1 - \rho_n) \sum_{t=1}^n x_t^2$ satisfies a law of large numbers and $n^{-1/2}(1 - \rho_n)^{1/2} \sum_{t=1}^n x_{t-1} u_t$ satisfies a central limit theorem. 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). The validity of confidence

[4] When $\rho \neq 1$, Assumptions 1a and 1b are equivalent.

interval 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) for the case of conditionally homoskedastic and conditionally heteroskedastic innovations u_t respectively. The current paper proposes a confidence interval for the autoregressive parameter with uniform coverage probability over the entire autoregressive parameter space $(-1, \infty)$. Extensions of the procedure of the paper are possible over the parameter space $(-\infty, \infty)$; we abstract from considering the region $(-\infty, -1]$ for brevity, since it requires a more involved construction of the instrument and several additional cases to be added in Assumption 1b and Theorem 3. Moreover, nonstationary and explosive oscillations with autoregressive roots in $(-\infty, -1]$ seem to be of limited empirical relevance in economics.

Assumption 2 requires u_t to be a conditionally homoskedastic[5] 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 $n^{-1} \sum_{t=1}^{n} u_t^2 \rightarrow_{L_1} \sigma^2$ and a functional central limit theorem $n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} u_t \Rightarrow B(r)$ on $D[0,1]$ where $B(r)$ is a Brownian motion with variance σ^2. The additional condition of Assumption 2 in the explosive case C_0(iii) ensures that the random variable

$$X_\infty = \left(\rho^2 - 1\right)^{1/2} \left(\sum_{j=1}^{\infty} \rho^{-j} u_j + X_0 - \mu \right) \tag{8}$$

is non-zero $a.s.$: condition (5) and $\mathbb{E}_{\mathcal{F}_{j-1}} \left(u_j^2\right) = \sigma^2$ imply the local Marcinkiewicz-Zygmund condition, see equation (1.1) of Lai and Wei (1983). Corollary 2 of Lai and Wei (1983) shows that the random variable in (8) satisfies $X_\infty \neq 0$ $a.s.$ when $(u_j, \mathcal{F}_j)$ is a martingale difference sequence.

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 processes (processes in C(iii) satisfying $\rho_n \rightarrow 1$). 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 (8) 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 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 on the initial condition X_0 of (1) is standard for processes in C(i) and C(ii) and in the mildly explosive case but significantly generalises 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

[5] 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).

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 of a predictive regression

$$y_t = \gamma + \beta x_{t-1} + \varepsilon_t \tag{9}$$

driven by an autoregressive process x_t in (1). While the parameter of interest in such models is β and the autoregressive root of (1) is a nuisance parameter, it is well-documented that 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 autoregressive 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 near-explosive region C(iii) is not considered by the above papers and 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 (9) 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 and the distribution of the innovations ε_t and u_t.

Conducting inference on β instead of the autoregressive parameter is possible under a more general assumption on the autoregressive innovations u_t than Assumption 2 above.

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 $C(1) := \sum_{j=0}^{\infty} c_j \neq 0$. 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 such that $\mathbb{E}_{\mathcal{F}_{t-1}}(v_t v_t') = \Sigma_v > 0$ a.s. for all t and* $\left(\|v_t\|^2\right)_{t\in\mathbb{Z}}$ is a uniformly integrable sequence. *When* $\rho_n \to \rho > 1$ under C_0(iii), *we assume in addition that (5) holds with u_t replaced by e_t and that $C_\rho(1) := \sum_{j=0}^{\infty} \rho^{-j} c_j \neq 0$.*

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 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:

$$\begin{aligned} I_{t+1} &= I_t\left(1 + \theta S_t/N - \gamma - \delta\right) \qquad (10) \\ S_{t+1} &= S_t\left(1 - \theta I_t/N\right), \ R_{t+1} = R_t + \gamma I_t, \ D_{t+1} = D_t + \delta I_t \end{aligned}$$

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 t, 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 $\theta, \gamma, \delta \in (0, 1]$[6]: 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 and there is no possibility of re-infection. The model's dynamics is driven by the *basic reproduction number* which in the model (10) is given by

$$r_0 = \theta / (\gamma + \delta), \tag{11}$$

measuring the number of infections per infected individual. When $r_0 \geq 1$ the underlying disease will escalate into an epidemic and will continue 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 and to conduct statistical inference on the model's parameters, we use next generation matrix (NGM) approach and linearise the system in (10) around the DFE ($I = R = D = 0, S = N$)[7]. 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 (11)) 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 will always display exponential growth[8].

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

$$\begin{bmatrix} I_t \\ \Delta R_t \\ \Delta D_t \end{bmatrix} = \begin{bmatrix} 1+\theta-\gamma-\delta & 0 & 0 \\ \gamma & 0 & 0 \\ \delta & 0 & 0 \end{bmatrix} \begin{bmatrix} I_{t-1} \\ R_{t-1} \\ D_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \tag{12}$$

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 (12).

Assumption 5. *The autoregressive parameter $\rho_n := 1+\theta-\gamma-\delta$ of I_t in (12) satisfies Assumption 1a; I_0 satisfies Assumption 3. The innovation sequence $u_t = [u_{1t}, u_{2t}, u_{3t}]'$ in (12) satisfies: $(u_t, \mathcal{F}_t)_{t\in\mathbb{N}}$ is a martingale difference sequence such that $\mathbb{E}_{\mathcal{F}_{t-1}}(u_t u_t') = \Sigma > 0$ for all t a.s. and*

[6]The requirements that $\theta, \gamma, \delta \leq 1$ imposes that the discrete period Δt is less than: (i) the average time required for a successful contact, (ii) less than the average infection period, and (iii) less than the average period the disease results into death; this requirement guarantees that the discretised SIR model approximates well the underlying continuous-time system.

[7]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.

[8]This result is not specific to our choice of SIR model and holds for more elaborate models; in larger systems it can be shown that the spectral radius of the resulting autoregressive parameter matrix for the vector of I_t exceeds 1 whenever $r_0 > 1$; see, for example, Theorem 2.1 from Allen and Van den Driessche (2008).

$(\|u_t\|^2)_{t\in\mathbb{Z}}$ *is a uniformly integrable sequence. In addition, under* C_0*(iii), (5) is satisfied with* $|u_t|$ *replaced by* $|u_{1t}|$.

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 through, for example, pre-testing 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 (9) that deliver a unified asymptotic theory of hypothesis testing and confidence interval construction 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). 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 purely explosive region C_0(iii), where least squares asymptotic inference is valid only under i.i.d. Gaussian innovations.

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 least squares estimator for ρ: for each $n \in \mathbb{N}$, define the event

$$F_n = \{n(\hat{\rho}_n - 1) \leq 0\} \tag{13}$$

where $\hat{\rho}_n$ is the OLS estimator and let $\bar{F}_n$ denote the complement of F_n. Asymptotic separation of the C(i) and C(iii) autoregressive classes can be achieved by employing the information contained in (13): under C(i) $n(\hat{\rho}_n - 1) \to_p -\infty$ which implies that $\mathbf{1}_{\bar{F}_n} = 0$ for all but finitely many n with probability tending to 1, whereas under C(iii) $n(\hat{\rho}_n - 1) \to_p \infty$ which implies that $\mathbf{1}_{F_n} = 0$ for all but finitely many n with probability tending to 1. This insight is formalised by the following result, the proof of which can be found in the Appendix.

Lemma 2. *Let* $(m_n)_{n\in\mathbb{N}}$ *be an arbitrary sequence of positive numbers such that* $m_n \to \infty$. *Under Assumption 4:* **(i)** *if* $(\rho_n)_{n\in\mathbb{N}}$ *belongs to C(i) then* $m_n\mathbf{1}_{\bar{F}_n} \to_p 0$ **(ii)** *if* $(\rho_n)_{n\in\mathbb{N}}$ *belongs to C(iii) then* $m_n\mathbf{1}_{F_n} \to_p 0$.

With the desired asymptotic separation guaranteed at an arbitrary rate by Lemma 2, we proceed to describing our instrumentation procedure. Given a sequence (v_t), we denote $\bar{v}_{n-j} := n^{-1}\sum_{t=1}^{n} v_{t-j}$ and $\underline{v}_{t-j} := v_{t-j} - \bar{v}_{n-j}$ for each $j \in \{0, ..., t-1\}$. In this notation, subtracting $\bar{x}_t$ from (2) yields

$$\underline{x}_t = \rho_n \underline{x}_{t-1} + \underline{u}_t; \tag{14}$$

the OLS estimator and residuals from (1)-(2) are given by

$$\hat{\rho}_n = \left(\sum_{t=1}^n \underline{x}_{t-1}^2\right)^{-1} \sum_{t=1}^n \underline{x}_t \underline{x}_{t-1} \text{ and } \hat{u}_t = \underline{x}_t - \hat{\rho}_n \underline{x}_{t-1}. \tag{15}$$

Recalling the definition of the event in (13), we define

$$\tilde{u}_t = \Delta x_t \mathbf{1}_{F_n} + \hat{u}_t \mathbf{1}_{\bar{F}_n} \text{ and } \rho_{nz} = \varphi_{1n} \mathbf{1}_{F_n} + \varphi_{2n} \mathbf{1}_{\bar{F}_n} \tag{16}$$

where $(\varphi_{1n})_{n\in\mathbb{N}}$ and $(\varphi_{2n})_{n\in\mathbb{N}}$ are *chosen* sequences in C(i) and C(iii) respectively: $\varphi_{1n} \to 1$ with $n(\varphi_{1n} - 1) \to -\infty$ and $\varphi_{2n} \to 1$ with $n(\varphi_{2n} - 1) \to \infty$. We construct an instrument process by accumulating the stochastic sequence $\tilde{u}_t$ in (16) 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 the orthogonal decomposition

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

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

$$\tilde{z}_{1t} = \varphi_{1n} \tilde{z}_{1t-1} + \Delta x_t \text{ and } \tilde{z}_{2t} = \varphi_{2n} \tilde{z}_{2t-1} + \hat{u}_t. \tag{19}$$

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

$$\tilde{\rho}_n = \frac{\sum_{t=1}^n \underline{x}_t \tilde{z}_{t-1}}{\sum_{t=1}^n \underline{x}_{t-1} \tilde{z}_{t-1}} = \mathbf{1}_{F_n} \tilde{\rho}_{1n} + \mathbf{1}_{\bar{F}_n} \tilde{\rho}_{2n} \tag{20}$$

where $\tilde{\rho}_{1n} = \sum_{t=1}^n \underline{x}_t \tilde{z}_{1t-1} / \sum_{t=1}^n \underline{x}_{t-1} \tilde{z}_{1t-1}$ and $\tilde{\rho}_{2n} = \sum_{t=1}^n \underline{x}_t \tilde{z}_{2t-1} / \sum_{t=1}^n \underline{x}_{t-1} \tilde{z}_{2t-1}$ employ the near-stationary and near-explosive instruments in (19) and (19) respectively. Filtering in (17) and (20) 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 near-stationary and near-nonstationary processes. The new 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) and differs from the IVX estimator based on $\tilde{z}_{1t}$ in two important ways: firstly, the instrument construction is based on the OLS residuals $\hat{u}_t$ which (unlike Δx_t) approximate well the true innovation process u_t in (1) in explosive autoregression; secondly, a mildly explosive (instead of a near-stationary) root is employed in the instrument generation.

The main contribution of the current approach, is to combine the novel near-explosive instrument $\tilde{z}_{2t}$ for regions C(ii) and C(iii) with the near-stationary instrument $\tilde{z}_{1t}$ in a data-driven way to provide inference for autoregressive roots in $(-1, \infty)$. Combining $\tilde{z}_{1t}$ with $\tilde{z}_{2t}$ to unify inference on both sides of unity is intuitively appealing but the asymptotic validity of such an approach is not obvious: the asymptotic mixed-Gaussianity (AMG) property of the estimator in (20) is established in Section 3.3. However, an essential step in the right direction is provided by the asymptotic separation property established by Lemma 2: since the sequence $(m_n)_{n\in\mathbb{N}}$ is allowed to diverge to ∞ arbitrarily fast, Lemma 2 implies that the asymptotic behaviour of $\tilde{\rho}_n$ in (20) will be driven exclusively by the component $\tilde{\rho}_{1n}$ involving the near-stationary instrument $\tilde{z}_{1t}$ when x_t is in C(i) and exclusively by the component $\tilde{\rho}_{2n}$ involving the mildly explosive instrument $\tilde{z}_{2t}$ when x_t is in C(iii). The fact that the contribution of both components in the near-nonstationary case C(ii) preserves the AMG property of the IV estimator in (20), as intended, is a central result of Section 3.3.

For the predictive regression model in (1) and (9), the same instrument (17) is employed, giving rise to the estimator

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

where $\tilde{\beta}_{1n} = \sum_{t=1}^{n} \underline{y}_t \tilde{z}_{1t-1} / \sum_{t=1}^{n} \underline{x}_{t-1} \tilde{z}_{1t-1}$ and $\tilde{\beta}_{2n} = \sum_{t=1}^{n} \underline{y}_t \tilde{z}_{2t-1} / \sum_{t=1}^{n} \underline{x}_{t-1} \tilde{z}_{2t-1}$ employ the near-stationary and near-explosive instruments in (19) and (18) respectively.

3.2 Asymptotic inference for autoregression and predictive regression

Estimation of the autoregressive root in (1) by the IV estimator in (20) has the advantage that the limit distribution of the normalised and centred estimator $\tilde{\rho}_n$ belongs to the mixed-Gaussian family of distributions in all C(i)-C(iii) cases, independently of the distribution of the innovations u_t in (1). This is in contrast to the OLS estimator which does not have a mixed-Gaussian limit distribution in the near-nonstationary case 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 estimator $\tilde{\rho}_n$ to Section 3.3. In this section, we focus on the implication of the above mixed-Gaussian property to inference, namely that self-normalised statistics based on $\tilde{\rho}_n$ in (20) have a standard normal limit distribution along the entire autoregressive parameter space of Assumption 1a, independently of the distribution of the innovations in (1) or the initial condition.

Denoting the lagged data vectors $X = (X_0, x_1, ..., x_{n-1})'$, $\tilde{Z} = (0, \tilde{z}_1, ..., \tilde{z}_{n-1})'$ and $\underline{X} = (X_0 - \bar{x}_{n-1}, ..., x_{n-1} - \bar{x}_{n-1})'$, we define a t-statistic based on $\tilde{\rho}_n$ as follows:

$$T_n(\tilde{\rho}_n) = \frac{\left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{1/2}}{\hat{\sigma}_n} (\tilde{\rho}_n - \rho_n) \tag{22}$$

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 (22) can be used to test hypotheses or to construct a $(1-\alpha)\%$ confidence interval for the autoregressive root ρ_n,

$$I_n(\tilde{\rho}_n, \alpha) = [\tilde{\rho}_n - c_n(\alpha), \tilde{\rho}_n + c_n(\alpha)], \quad c_n(\alpha) = \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\hat{\sigma}_n}{\left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{1/2}} \tag{23}$$

where Φ denotes the $\mathcal{N}(0,1)$ distribution function. As a consequence of the AMG property of $\tilde{\rho}_n$, established by Theorem 3 in Section 3.3., the t-statistic in (22) and the confidence interval in (23) enjoy asymptotic properties that are very convenient for inference, presented in Theorem 1 below.

Theorem 1. *Consider the process (1) satisfying Assumptions 1a, 2 and 3, the process $\tilde{z}_t$ defined in (16)-(17) and the IV estimator $\tilde{\rho}_n$ in (20). The t-statistic in (22) satisfies $T_n(\tilde{\rho}_n) \to_d \mathcal{N}(0,1)$ as $n \to \infty$ and the confidence interval in (23) satisfies $\lim_{n\to\infty} \inf_{\theta\in\Theta} \mathbb{P}[\rho \in I_n(\tilde{\rho}_n, \alpha)] = 1 - \alpha$, where Θ denotes the parameter space in (7).*

Remarks.

1. Theorem 1 shows that the methodology of the paper delivers distribution-free inference for autoregressions satisfying Assumption 1a. To our knowledge, this is the first procedure that provides a unified framework of inference and confidence interval construction when data originate from autoregressive time series encompassing the stationary, nonstationary, explosive and all intermediate regions described in C(i)-C(iii), without *a priori* knowledge or the need for pre-testing on the type of autoregressive process that generates the data. The generality of our methodology makes it suitable for empirical application such as the stochastic evolution of Covid-19 infections, 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 details.

2. In view of the general validity of the standard normal limit distribution for the t-statistic in (22), inference for the autoregressive parameter based on the confidence interval (23) has uniform asymptotic validity over the parameter space Θ in (7): in the terminology of Andrews, Cheng and

Guggenberger (2020), the CI in (23) has correct asymptotic size and is uniformly asymptotically similar.

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 (20), established by Theorem 3 below. The instrumentation of the autoregressive process by a combination of a near-stationary and mildly explosive process 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) of the parameter space) to construct the estimator (20) that enjoys the AMG property across the spectrum of autoregressive classes C(i)-C(iii).

4. A particular advantage of the inferential framework of (20), (22) and (23) is that it 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. On the other hand, the IV estimator $\tilde{\rho}_n$ in (20) has the AMG property irrespective of the distribution of (u_t), as Theorem 3 shows, giving rise to the distribution-free and correctly-sized asymptotic confidence interval in (23). To our knowledge, the t-statistic in (22) and the associated confidence interval in (23) provide the first solution to the problem of distribution-free asymptotic inference in the explosive autoregression.

5. The asymptotic normality result of Theorem 1 includes oscillating sequences 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 (4) with $k_n = n$. The standard t-statistic $T_n(\hat{\rho}_n)$ based on the OLS estimator $\hat{\rho}_n$ satisfies $T_{2n}(\hat{\rho}_{2n}) \to_d R_1$ and $T_{2n-1}(\hat{\rho}_{2n-1}) \to_d R_{-1}$ where $R_c = \left(\sigma^2 \int_0^1 J_c(r)^2 dr\right)^{-1/2} \int_0^1 J_c(r)\, dB(r)$. Since the random variables R_1 and R_{-1} do not have the same distribution, the sequence $\{T_n(\hat{\rho}_n) : n \in \mathbb{N}\}$ does not converge in distribution. For the above sequence $(\rho_n)_{n\in\mathbb{N}}$, the IV estimator $\tilde{\rho}_n - \rho_n$ in (20) also has, after appropriate normalisation, two accumulation points in distribution (along the odd and even integers). However, as Theorem 3 shows, both accumulation points will have the AMG property; as a result the t-statistic $T_n(\tilde{\rho}_n)$ in (22) satisfies $T_{2n}(\tilde{\rho}_{2n}) \to_d \mathcal{N}(0,1)$ and $T_{2n-1}(\tilde{\rho}_{2n-1}) \to_d \mathcal{N}(0,1)$, implying that the entire sequence $\{T_n(\tilde{\rho}_n) : n \in \mathbb{N}\}$ converges in distribution to $\mathcal{N}(0,1)$. The proof of Theorem 1 employs Lemma 1 to show that the above asymptotic behaviour of the t-statistic in (22) is typical and $T_n(\tilde{\rho}_n) \to_d \mathcal{N}(0,1)$ only requires the weaker Assumption 1a.

For the predictive regression model in (1) and (9), we employ a similar studentisation to (22) based on the IV estimator $\tilde{\beta}_n$ in (21):

$$T_n\left(\tilde{\beta}_n\right) = \frac{\left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{1/2}}{\hat{\sigma}_\varepsilon}\left(\tilde{\beta}_n - \beta_n\right) \tag{24}$$

where $\hat{\sigma}_\varepsilon^2$ is the OLS estimator of the variance of ε_t in (9). While the t-statistic in (24) is shown to be asymptotically standard normal in Theorem 2 below, the estimation of the intercept in (9) induces a finite sample size distortion when x_t has a unit root and a near-stationary instrument is employed, as documented by Kostakis et al. (2015), Hosseinkouchack and Demetrescu (2020) and Harvey, Leybourne and Taylor (2021). The problem occurs because the sample moment that drives mixed normality is given by $\sum_{t=1}^n \tilde{z}_{1t-1}\varepsilon_t - \bar{z}_{1n-1}\bar{\varepsilon}_n$ and, while the first term on the right-hand side dominates and is asymptotically normally distributed, $\bar{z}_{n-1}\bar{\varepsilon}_n$ is not asymptotically mixed-Gaussian and has more pronounced finite sample effects when x_t is a unit root process (see Remark 2 below). Given that the finite sample distortion only occurs very close to unity, one solution is to employ the fully-modified (FM) transformation of Phillips and Hansen (1990) that orthogonalises the innovations ε_t of (9) 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 process. The FM-corrected IV estimator $\tilde{\beta}_{1n}$ in (21) (the component of $\tilde{\beta}_n$ generated by a near-stationary 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 $\sigma^2_\varepsilon = \mathbb{E}\left(\varepsilon^2_t\right)$, $\omega^2_u = \sum_{k=-\infty}^{\infty} \mathbb{E}\left(u_t u_{t-k}\right)$ and $\rho_{\varepsilon u} = corr\left(\varepsilon_t, u_t\right)$. With the above correction, the IV estimator of β becomes

$$\beta^*_n = \mathbf{1}_{F_n} \beta^*_{1n} + \mathbf{1}_{\bar{F}_n} \tilde{\beta}_{2n} \tag{25}$$

with $\tilde{\beta}_{2n}$ defined as in (21). A simple computation of the standard error of the estimator β^*_n above gives rise to the following t-statistic:

$$T_n\left(\beta^*_n\right) = \frac{1}{\hat{\sigma}_\varepsilon} \frac{\sum_{t=1}^n \underline{x}_{t-1} \tilde{z}_{t-1}}{\left(\sum_{t=1}^n \tilde{z}^2_{t-1} - n\bar{z}^2_{1,n-1} \left(1 - \hat{\rho}^2_{\varepsilon u}\right) \mathbf{1}_{F_n} \right)^{1/2}} \left(\beta^*_n - \beta_n\right). \tag{26}$$

Theorem 2. *Consider the predictive regression model (1) and (9) satisfying Assumptions 1a, 3 and 4, the filtered process $\tilde{z}_t$ defined by (16)-(17) and the IV estimators $\tilde{\beta}_n$ and β^*_n in (21) and (25). The statistics in (24) and (26) satisfy $T_n\left(\tilde{\beta}_n\right) \to_d \mathcal{N}(0,1)$ and $T_n\left(\beta^*_n\right) \to_d \mathcal{N}(0,1)$.*

Remarks.

1. The standard normal limit distribution of the t-statistics in (24) and (26) is invariant to the nuissance parameter c of Assumption 1b that defines the autoregressive categories C(i)-C(iii). Consequently, hypothesis tests on β with critical regions based on Theorem 2 will have uniform asymptotic size and the corresponding confidence interval for β, $I_n\left(\beta^*_n, \alpha\right) = \left[\beta^*_n - c_n(\alpha), \beta^*_n + c_n(\alpha)\right]$ with $c_n(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \hat{\sigma}_\varepsilon \left(\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{t-1}\right)^{-1} \left(\sum_{t=1}^n \tilde{z}^2_{t-1} - n\bar{z}^2_{1,n-1}\left(1 - \hat{\rho}^2_{\varepsilon u}\right)\mathbf{1}_{F_n}\right)^{1/2}$ has uniform asymptotically correct coverage probability (in the sense of Theorem 1) over the parameter space Θ in (7) with Assumption 2 replaced by Assumption 4 for the innovations (u_t).

2. While $T_n\left(\beta^*_n\right)$ and $T_n\left(\tilde{\beta}_n\right)$ have the same limit distribution, the test based on $T_n\left(\tilde{\beta}_n\right)$ may suffer from finite sample distortion due to the fact that the estimation of the intercept γ in (9) 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 $o_p(1)$, $n\pi_n^{-1}\bar{z}_{1n-1}\bar{\varepsilon}_n = O_p(n^{-1/2}\left(1 - \varphi_{1n}\right)^{-1/2})$ under C(ii) in the notation of Theorem 3, $n\bar{z}_{1n-1}\bar{\varepsilon}_n$ is asymptotically equivalent to $\left(1 - \varphi_{1n}\right)^{-1} x_n \sum_{t=1}^n \varepsilon_t$ 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 $T_n\left(\tilde{\beta}_n\right)$ exhibits finite sample distortions when the following occur *jointly*: (i) the autoregressive root of x_t is very close to unity; (ii) $\rho_{\varepsilon u} = corr\left(\varepsilon_t, u_t\right)$ is close to 1 in absolute value; (iii) φ_{1n} is chosen close to 1. The FM transformation of Phillips and Hansen (1990), $\varepsilon_{0t} = \varepsilon_t - \omega^{-1}\mathbb{E}\left(\varepsilon_t u_t\right) u_t$, 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 trasformation and the corresponding t-statistic $T_n\left(\beta^*_n\right)$ have significantly improved finite sample properties whenever ρ_n is close to 1 with large

$|\rho_{\varepsilon u}|$, while both $T_n\left(\tilde{\beta}_n\right)$ and $T_n\left(\beta_n^*\right)$ perform equally well in all other cases.

3. Practical implementation of the test procedures of Theorems 1 and 2 requires a choice for φ_{1n} and φ_{2n} in (16) for the construction of the instrument $\tilde{z}_t$. Since our procedure is designed to work across the autoregressive parameter space $(-1,\infty)$, we require one 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 minimising the worst finite sample distortion scenario: from Remark 2 above, we know that our test procedure on β suffers its worst small-sample distortions in the case of a unit root regressor with large correlation $|\rho_{\varepsilon u}|$. 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; 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 φ_{1n} and φ_{2n} in Section 4.1 works very well (both in terms of size and power) for all autoregressive specifications in $(-1,\infty)$ both in the autoregression and in the predictive regression setups.

4. The above methodology may be extended to multivariate predictive regression models where both x_t and y_t in (1) and (9) 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 roots in $(1,\infty)$ 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 (16)-(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 confidence intervals for ρ and β with uniform asymptotic validity). The general multivariate setup, where x_t is an unrestricted VAR with possible cointegrating relationships and feedback effects between near-nonstationary and near-explosive components, is more challenging, as it requires the development of new VAR representation theory, of the Granger-Johansen type.

We now turn to the problem of conducting inference for the parameters of the epidemiological model in (12) and, in particular, of constructing robust confidence intervals for the basic reproduction number r_0 in (11) regardless of whether r_0 is above, equal or below unity. Denoting the autoregressive parameter of I_t in the first equation of (12) by $\rho_n := 1+\theta-\gamma-\delta$, (12) 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)-(20):

$$\tilde{\rho}_n = \frac{\sum_{t=1}^n I_t \tilde{z}_{t-1}}{\sum_{t=1}^n I_{t-1}\tilde{z}_{t-1}},\quad \tilde{\gamma}_n = \frac{\sum_{t=1}^n \Delta R_t \tilde{z}_{t-1}}{\sum_{t=1}^n I_{t-1}\tilde{z}_{t-1}}\ \text{ and }\ \tilde{\delta}_n = \frac{\sum_{t=1}^n \Delta D_t \tilde{z}_{t-1}}{\sum_{t=1}^n I_{t-1}\tilde{z}_{t-1}} \tag{27}$$

where the instrument $\tilde{z}_t$ is constructed from the first equation of (12) by

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

where $\hat{u}_{1t}$ are the OLS residuals obtained from the first equation of (12), the events F_n and $\bar{F}_n$ are defined in (13) and ρ_{nz} is chosen as in (16). 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 $\gamma+\delta$) and the expression for ρ_n as:

$$\tilde{r}_n = 1 + (\tilde{\rho}_n - 1)/\left(\tilde{\gamma}_n + \tilde{\delta}_n\right)\ \text{ and }\ \tilde{\theta}_n = \tilde{\rho}_n - 1 + \tilde{\gamma}_n + \tilde{\delta}_n \tag{28}$$

where $\tilde{\rho}_n$, $\tilde{\gamma}_n$ and $\tilde{\delta}_n$ are the IV estimators in (27). Adjusting for the asymptotic variance of $\tilde{r}_n$

and $\tilde{\theta}_n$, we may construct studentised version of these estimators as follows:

$$\left[T_n\left(\tilde{r}_n\right), T_n\left(\tilde{\theta}_n\right), T_n\left(\tilde{\gamma}_n\right), T_n\left(\tilde{\delta}_n\right)\right] = \left(X'P_{\tilde{Z}}X\right)^{1/2}\left[\frac{\tilde{r}_n - r_0}{\hat{\sigma}_{r_0}}, \frac{\tilde{\theta}_n - \theta}{\hat{\sigma}_\theta}, \frac{\tilde{\gamma}_n - \gamma}{\hat{\sigma}_\gamma}, \frac{\tilde{\delta}_n - \delta}{\hat{\sigma}_\delta}\right] \tag{29}$$

where $X = [I_0, I_1, ..., I_{n-1}]'$, $\tilde{Z} = [0, \tilde{z}_1, ..., \tilde{z}_{n-1}]'$, $\hat{\sigma}^2_{r_0} = \hat{\upsilon}'_n\hat{\Sigma}_n\hat{\upsilon}_n$, $\hat{\sigma}^2_\theta = \iota'\hat{\Sigma}_n\iota$, $\hat{\sigma}^2_\gamma = e'_2\hat{\Sigma}_n e_2$, $\hat{\sigma}^2_\delta = e'_3\hat{\Sigma}_n e_3$, $\hat{\Sigma}_n = n^{-1}\sum_{t=1}^n \hat{u}_t\hat{u}'_t$ with $\hat{u}_t$ denoting the OLS residuals of (12), $\iota = [1, 1, 1]'$, $e_2 = [0, 1, 0]'$, $e_3 = [0, 0, 1]'$ and $\hat{\upsilon}_n = \left[1/\left(\hat{\gamma}_n + \hat{\delta}_n\right), (1 - \hat{\rho}_n)/\left(\hat{\gamma}_n + \hat{\delta}_n\right)^2, (1 - \hat{\rho}_n)/\left(\hat{\gamma}_n + \hat{\delta}_n\right)^2\right]'$ based on the OLS estimators $\hat{\rho}_n$, $\hat{\gamma}_n$ and $\hat{\delta}_n$ in (12). Letting $c^g_n(\alpha) = \left(X'P_{\tilde{Z}}X\right)^{-1/2}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\hat{\sigma}_g$ for $g \in \{r_0, \theta, \gamma, \delta\}$ and denoting $[a \pm b] = [a - b, a + b]$ for brevity, we may construct confidence intervals based on the studentised estimators in (29): $I_n(\tilde{r}_n, \alpha) = [\tilde{r}_n \pm c^{r_0}_n(\alpha)]$, $I_n(\tilde{\theta}_n, \alpha) = [\tilde{\theta}_n \pm c^\theta_n(\alpha)]$, $I_n(\tilde{\gamma}_n, \alpha) = [\tilde{\gamma}_n \pm c^\gamma_n(\alpha)]$ and $I_n\left(\tilde{\delta}_n, \alpha\right) = [\tilde{\delta}_n \pm c^\delta_n(\alpha)]$. The asymptotic distribution of the t-statistics in (29) and the asymptotic coverage probabilities of the corresponding confidence intervals can be easily deduced from the analysis leading to Theorem 1 above.

Corollary 1. *Consider the model (I_t, R_t, D_t) in (12) satisfying Assumption 5 with parameters r_0, θ, γ and δ estimated in (27) and (28). The t-statistics in (29) all converge in distribution to $\mathcal{N}(0, 1)$ and the associated $1 - \alpha$ confidence intervals $I_n(\tilde{r}_n, \alpha)$, $I_n(\tilde{\theta}_n, \alpha)$, $I_n(\tilde{\gamma}_n, \alpha)$ 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 normalised and centred IV estimator in (20) under Assumptions 1b, 2 and 3. The main result, Theorem 3 below, is preceded by a discussion of the stochastic properties of the instrument process $\tilde{z}_t$ in (17) and three results, Lemmata 3-5, that provide an insight into the mechanics that yield the AMG property and facilitate the proof of Theorem 3.

We first provide a brief informal discussion of the behaviour of the instrument under the different regimes C(i)-C(iii). 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 belonging to 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

$$z_{1t} = \varphi_{1n}z_{1t-1} + u_t = \sum_{j=1}^t \varphi_{1n}^{t-j}u_j \tag{30}$$

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 its construction, the instrument $\tilde{z}_{2t}$ in (19) is always approximated by mildly explosive process

$$z_{2t} = \varphi_{2n}z_{2t-1} + u_t = \sum_{j=1}^t \varphi_{2n}^{t-j}u_j \tag{31}$$

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 A1 in the Appendix.

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

$$\Gamma_n = \sum_{k=1}^{\infty} \rho_n^{k-1} \gamma_u(k) \quad \text{and} \quad \Gamma = \sum_{k=1}^{\infty} \rho^{k-1} \gamma_u(k). \tag{32}$$

By Assumption 1a, $\rho_n \to \rho$ and $\Gamma = \lim_{n\to\infty} \Gamma_n$ exists by the dominated convergence theorem since $\sum_{k=1}^{\infty} |\gamma_u(k)| < \infty$ by Assumption 4. Note that, 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 the limit c in Assumption 1b is a real number, define the Ornstein-Uhlenbeck processes

$$W_c(t) = \int_0^t e^{c(t-s)} dW(s) \quad \text{and} \quad J_c(t) = \int_0^t e^{c(t-s)} dB(s) \tag{33}$$

and the Dickey-Fuller-type ratio

$$K_c = \int_0^1 J_c(r)\, dB(r) / \int_0^1 J_c(r)^2\, dr. \tag{34}$$

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*

$$\begin{aligned} \tilde{\Psi}_n &= (1+\rho_n)\left[\sigma^2 + 2\rho_n \Gamma_n + (2\rho_n - 1)\left(\frac{1}{n}\sum_{t=1}^n x_{0t-1} u_t - \Gamma_n\right)\right] \\ &\quad -\rho_n\left(1-\rho_n^2\right)\frac{1}{n}\sum_{t=1}^n x_{0t-1}^2 - 2\frac{x_{0n}}{n^{1/2}}\frac{1}{n^{3/2}}\sum_{j=1}^n x_{0j-1} \end{aligned} \tag{35}$$

$\tilde{\Psi}(c) = \sigma^2 + 2\rho\Gamma + \left(J_c(1)^2 - 2J_c(1)\int_0^1 J_c(r)\, dr\right) \mathbf{1}\{c \in \mathbb{R}\}$ *and* x_{0t} is defined in *(3).*

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

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

where Γ_n *and* Γ *are defined in (32),* $J_c(\cdot)$ *in (33) 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

$$\nu_n = \left(\rho_n^2 - 1\right)^{-1/2} \rho_n^n \mathbf{1}\{c = \infty\} + n^{1/2}\mathbf{1}\{c \in \mathbb{R}\} \quad \text{and} \quad \nu_{n,z} = \left(\varphi_{2n}^2 - 1\right)^{-1/2} \varphi_{2n}^n \tag{36}$$

where c denotes the limit in Assumption 1b, and

$$s_n = (\rho_n \varphi_{2n} - 1)^{-1} \nu_{n,z} \nu_n. \tag{37}$$

Following Phillips and Magdalinos (2007), the limit theory for the mildly explosive instrument's sample moments will be driven by the stochastic sequences

$$[Y_n, Y_n^{\varepsilon}, Z_n] := \left(\varphi_{2n}^2 - 1\right)^{1/2}\left[\sum_{t=1}^n \varphi_{2n}^{-(n-t)-1} u_t, \sum_{t=1}^n \varphi_{2n}^{-(n-t)-1} \varepsilon_t, \sum_{j=1}^n \varphi_{2n}^{-j} u_j\right]. \tag{38}$$

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

$$X_n := \frac{x_n}{\nu_n} = \left(\rho_n^2 - 1\right)^{1/2}\left(\sum_{j=1}^n \rho_n^{-j} u_j + X_0(n) - \mu\right)\mathbf{1}\{c = \infty\} + \frac{x_n}{\sqrt{n}}\mathbf{1}\{c \in \mathbb{R}\}. \tag{39}$$

The following result characterises the joint asymptotic behaviour of the sequences Y_n, Z_n and X_n and the sample moment asymptotic behaviour of the instrument $\tilde{z}_{2t}$ for the autoregressive classes C(ii)-C(iii).

Lemma 4. *Let* X_∞ *be the random variable defined in (8) and* Y_n, Y_n^{ε}, Z_n *and* X_n *be the stochastic sequences in (38) and (39) and let* Y, Y^{ε}, Z, X *denote* $\mathcal{N}(0, \omega^2)$ *random variables. Under Assumptions 3 and 4 and C(ii)-C(iii) of Assumption 1b, the following hold as* $n \to \infty$*:*

(i) $[Y_n, Z_n] \to_d [Y, Z]$ *and* $[Y_n^{\varepsilon}, Z_n] \to_d [Y^{\varepsilon}, Z]$ *where* Z *is independent of* Y *and* Y^{ε}*. In* *addition,*

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

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

(iii) *Under Assumption 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* $\mathcal{MN}\left(0,\omega^2 g^2(X_\infty)\right)$ *variables.*

Part (iii) of Lemma 4 deserves special attention because it establishes a central limit theorem to a mixed-Gaussian distribution in the purely explosive case C_0(iii) and 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 (20) 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 will have a $\mathcal{MN}\left(0,\sigma^2/X_\infty^2\right)$ limit distribution in the explosive case, establishing the asymptotic mixed normality 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 near-nonstationary 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 limit in distribution of the sequence of events F_n in (13). Additional complication is introduced by the randomness of the limits of the signals ($2\Psi_n\to_d\sigma^2+J_c(1)^2$ from $\tilde{z}_{1t}$ and $X_nZ_n\to_d J_c(1)Z$ 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) and Chan and Wei (1987), the sequences F_n, Ψ_n and X_n in (13), (35) and (39) 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 is sufficient to prove the independence of $[U(1), Y]$ and the Brownian motion B; it is established in the following result.

Lemma 5. *Define the following random elements in* $D[0,1]$: $B_n(s)=n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor}u_t$, $U_n(s)=\left(n\left(1-\varphi_{1n}^2\right)^{-1}\right)^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor}z_{1t-1}e_t$ *and* $Y_n(s)=\left(\varphi_{2n}^2-1\right)^{1/2}\sum_{t=1}^{\lfloor ns\rfloor}\varphi_{2n}^{-(\lfloor ns\rfloor-t)-1}u_t$. *Under Assumptions 1b and 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 may now employ the limit theory of Lemmata 2-5 to establish the AMG property of the IV estimator $\tilde{\rho}_n$ in (20) within each of the autoregressive classes C(i)-C(iii). For $c\in\mathbb{R}$, define $\underline{W}_c(t):=W_c(t)-\int_0^1 W_c(r)\,dr$ and the random variables

$$\Psi_-(c)=1+W_c(1)^2-2W_c(1)\int_0^1 W_c(r)\,dr \quad\text{and}\quad \Psi_+(c)=2\underline{W}_c(1) \tag{40}$$

where the event F_c and its complement $\bar{F}_c$ are defined by $F_c=\{K_c+c\leq 0\}$ with K_c being the ratio in (34).

Theorem 3. *Consider the autoregression (1) and the predictive regression model (9) and (1) under Assumptions 1b and 3, and the IV estimators* $\tilde{\rho}_n$ *in (20),* $\tilde{\beta}_n$ *in (21) and* β_n^* *in (25). The following asymptotic approximations apply as* $n\to\infty$ *to* $\pi_n(\tilde{\rho}_n-\rho_n)$ *under Assumption 2 and to* $\pi_n\left(\tilde{\beta}_n-\beta\right)$ *under Assumption 4 respectively:*

(i) *Under part C(i) of Assumption 1b,* $\pi_n=\left(n\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1}\right)^{1/2}$,

$$\pi_n(\tilde{\rho}_n-\rho_n)\to_d\mathcal{N}(0,1)\quad\text{and}\quad\pi_n\left(\tilde{\beta}_n-\beta_n\right)\to_d\mathcal{N}\left(0,\sigma_\varepsilon^2/\left(\sigma^2+2\Gamma\right)\right).$$

(ii) *Under part C(ii) of Assumption 1b,* $\pi_n = n^{1/2}\left(\left(1-\varphi_{1n}^2\right)^{-1}\mathbf{1}_{F_n} + \left(\varphi_{2n}^2 - 1\right)^{-1}\mathbf{1}_{\bar{F}_n}\right)^{1/2}$

$$\pi_n\left(\tilde{\rho}_n - \rho_n\right) \to_d \mathcal{MN}\left(0, \Psi\left(c\right)^{-2}\right) \quad \text{and} \quad \pi_n\left(\tilde{\beta}_n - \beta_n\right) \to_d \mathcal{MN}\left(0, \frac{\sigma_\varepsilon^2}{\omega^2}\Psi\left(c\right)^{-2}\right)$$

where the events F_n *and the random variable* $\Psi\left(c\right) = \Psi_-\left(c\right)\mathbf{1}_{F_c} + \Psi_+\left(c\right)\mathbf{1}_{\bar{F}_c}$ *are defined in (13) and (40) respectively.*

(iii) *Under part C(iii) of Assumption 1b,* $\pi_n = \left(\varphi_{2n}^2 - 1\right)^{1/2}\left(\rho_n\varphi_{2n} - 1\right)^{-1}\left(\rho_n^2 - 1\right)^{-1/2}\rho_n^n$

$$\pi_n\left(\tilde{\rho}_n - \rho_n\right) \to_d Y/X =_d \mathcal{MN}\left(0, \sigma^2/X\right) \quad \text{and} \quad \pi_n\left(\tilde{\beta}_n - \beta_n\right) \to_d \tilde{Y}/X =_d \mathcal{MN}\left(0, \sigma_\varepsilon^2/X\right)$$

where $Y =_d \mathcal{N}\left(0, \sigma^2\right)$, $\tilde{Y} =_d \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, *and* X *is independent of* $\left(Y, \tilde{Y}\right)$ *with* $X =_d \mathcal{N}\left(0, \omega^2\right)$ *when* $\rho_n \to 1$ *and* $X = X_\infty$ *in (8) when* $\rho_n \to \rho > 1$.

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

Remarks.

1. The data-filtering procedure proposed in the paper guarantees that the resulting estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ in (20), and (21) respectively exhibit a AMG property along the entire spectrum of autoregressive regressor processes, including stationary, non-stationary, explosive processes and all intermediate regimes. 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 (9): 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 can be a linear process under Assumption 4. The only component that depends on the distribution of (u_t) is the mixing variate $\tilde{X}_\infty$ 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 $\tilde{\rho}_n$ is in sharp contrast with the dependence of large sample OLS inference on the distribution of (u_t) in explosive autoregressions. Hence, in addition to producing robust inference along all autoregressive classes, our proposed estimation procedure is the first that achieves distribution-free asymptotic inference in explosive autoregression and is asymptotically invariant to the initialisation X_0 of the time series in (1).

2. The key element of the procedure that delivers the AMG property and the distributional invariance to the autoregressive innovations across the autoregressive classes C(i)-C(iii) is the newly proposed combined instrument $\tilde{z}_t$ in (16)-(17). This instrument employs information from the OLS estimator of the autoregressive 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 $\tilde{z}_{1t}$ and the resulting IV estimator $\tilde{\rho}_n$ is asymptotically equivalent to the IVX estimator of Phillips and Magdalinos (2009). When $c = \infty$, $\tilde{z}_t$ takes the form of a mildly-explosive instrument $\tilde{z}_{2t}$ constructed from the least squares residuals $\hat{u}_t$ and the resulting IV estimator $\tilde{\rho}_n$ based on $\tilde{z}_{2t}$ is shown to achieve distribution-free inference on the non-stationary side of unity, including the explosive region. When $c \in \mathbb{R}$, the autoregression is of the near-nonstationary type C(ii) in which case $\tilde{z}_t$ takes the form of a random linear combination of $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$. 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 in (13) which, like the limit distribution of Ψ_n in (35), 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 $\tilde{z}_{2t}$ 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 crucial feature of our estimation procedure that delivers the robust and distribution-free inference based on the t-statistic of Theorems 1 and 2.

3. It is worth making a comparison of the current instrumentation procedure to the IVX method of Phillips and Magdalinos (2009), particularly since the latter was applied by Phillips and Lee (2016) to conduct inference in predictive regression in the presence of mildly explosive time series. The original IVX approach was designed to address local-to-unit-root type of nonstationarity; for this reason, the filtered process in (17) is constructed by using Δx_t (instead for the OLS residuals $\hat{u}_t$) and a near-stationary root. However, in an explosive setup, differencing the explosive regressor x_t will not produce an $I(0)$ process and the IVX instrumentation effect vanishes asymptotically: as Phillips and Lee (2016) show, under C(iii) the original IVX estimator reduces asymptotically to the OLS estimator. Hence, the original IVX estimator inherits the limitations of least squares limit theory in the explosive case and cannot resolve its lack of central limit theory and its dependence on the distribution of the innovations (u_t) and the initial condition. On the other hand, by employing the residuals $\hat{u}_t$ and a mildly explosive root φ_{2n} for the construction of the instrument $\tilde{z}_{2t}$ in (17), the instrumentation of this paper ensures that the limit distribution of $\tilde{\rho}_n$ is driven by the mildly explosive component z_{2t} in equation (A.1) of the Appendix under C(iii) 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 invariance and subsequent distribution-free inference is a reduction in the convergence rate of $\tilde{\rho}_n - \rho_n$ by an order of $(\varphi_{2n}^2 - 1)^{1/2}$ compared to the ρ_n^n-OLS rate. However, given that the above order satisfies $(\varphi_{2n}^2 - 1)^{-1/2} = o(n^{1/2})$ and that the exponential part ρ_n^n of the OLS rate is maintained in the convergence rate of Theorem 3(iii), the efficiency loss associated with employing $\tilde{\rho}_n$ is small compared to the benefit of an estimation procedure that gives rise to test statistics and confidence intervals of general asymptotic validity.

4. The limit distribution theory of Theorem 3 under C(i) and C(iii) shares some similarity with the corresponding results for the OLS estimator: in particular, in the case when $\rho_n \to 1$ both asymptotic distributions are Gaussian under C(i) and Cauchy under C(iii) (Y/X is Cauchy distributed when $X =_d \mathcal{N}(0, \sigma^2)$). This is not surprising since in both cases the instrument $\tilde{z}_t$ has similar time series properties (near-stationary or mildly explosive) as the original autoregressive process x_t. Building the instrument $\tilde{z}_{1t}$ based on Δx_t rather than the autoregressive residuals (as for $\tilde{z}_{2t}$) maintains the asymptotic optimality of the IVX procedure of Phillips and Magdalinos (2009) under C(i) when φ_{1n} is chosen closer to unity than ρ_n.

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 explosivity. In Section 4.2 we provide an illustration of the failure of general asymptotic inference based on the OLS estimator in the explosive region: 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 in these cases. Next, we compare the finite sample properties of our procedure to the leading existing approaches: in Section 4.3.1, we provide a comparison of our confidence intervals in (23) for the autoregressive parameter to the procedure of Andrews and Guggenburger (2014); in Section 4.3.2, we compare the size and power of our testing procedure in (26) 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 to the left of unity (including local-to-unity, near- and purely stationary regions) except for the case of exact unit root. Crucially, our procedure also provides correct inference on the right side of unity, in the local-to-unit-root, mildly and purely explosive regions, where no existing alternative approach has general asymptotic validity.

4.1 Instrument Selection

Practical implementation of our procedure requires a choice for φ_{1n} and φ_{2n} in (16) for the instrument construction in (17). While theoretically, any values of $\varphi_{1n} \to 1$ belonging to C(i) and $\varphi_{2n} \to 1$ belonging to C(iii) deliver correct asymptotic inference, finite sample performance may vary considerably with the choice for φ_{1n} and φ_{2n}. Choosing

$$\varphi_{1n} = 1 - 1/n^{b_1}, \ \varphi_{2n} = 1 + 1/n^{b_2} \tag{41}$$

reduces the problem to selecting values for b_1 and b_2 in $(0, 1)$. Since we require an instrument selection with good performance along all autoregressive regions without *a priori* knowledge, we adopt a conservative approach: from Remark 2 after Theorem 2, we know that our inference procedure 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 (9)[9]. Therefore, we base our selection procedure for the values of b_1 and b_2 on the principle of minimising the worst case scenario and select values that deliver satisfactory test size in this most unfavourable case. We consider a grid of values for b_1 and b_2 in (41) with very strong positive and negative correlation $|\rho_{\varepsilon u}| = 0.99$. Tables B1 and B2 of the online Appendix B contain the empirical size 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 various combinations of b_1 and b_2 for $\rho_{\varepsilon u} = 0.99$ and $\rho_{\varepsilon u} = -0.99$ respectively. The power in this case (plots of which for the grid points can be found in Figure B1 of the online Appendix B) is 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%. 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 the online Appendix B), since the near-stationary instrument is chosen 2/3 of the time (this is since the OLS distribution in the unit root case is left-skewed with values below unity occurring with probability 2/3), and (iii) values for b_2 close to unity would imply that our mildly explosive instrument is near the boundary with local-to-unity region, which would cause the instrument inheriting local-to-unity properties and potentially some of the associated small sample distortions when working with purely explosive regressor. From Tables B1 and B2 of the online Appendix B, we suggest using $b_1 = 0.85$ and $b_2 = 0.7$ in (41) since in both cases, the empirical size in these unfavourable cases does not exceed 5.99%. We use these values throughout the rest of the Monte Carlo section (and in the empirical application in Section 5) and demonstrate that our choice works well for all autoregressive specifications in both a predictive and autoregressive setup.

[9] In the autoregressive setup, such finite sample distortions are less pronounced.

FIGURE 1

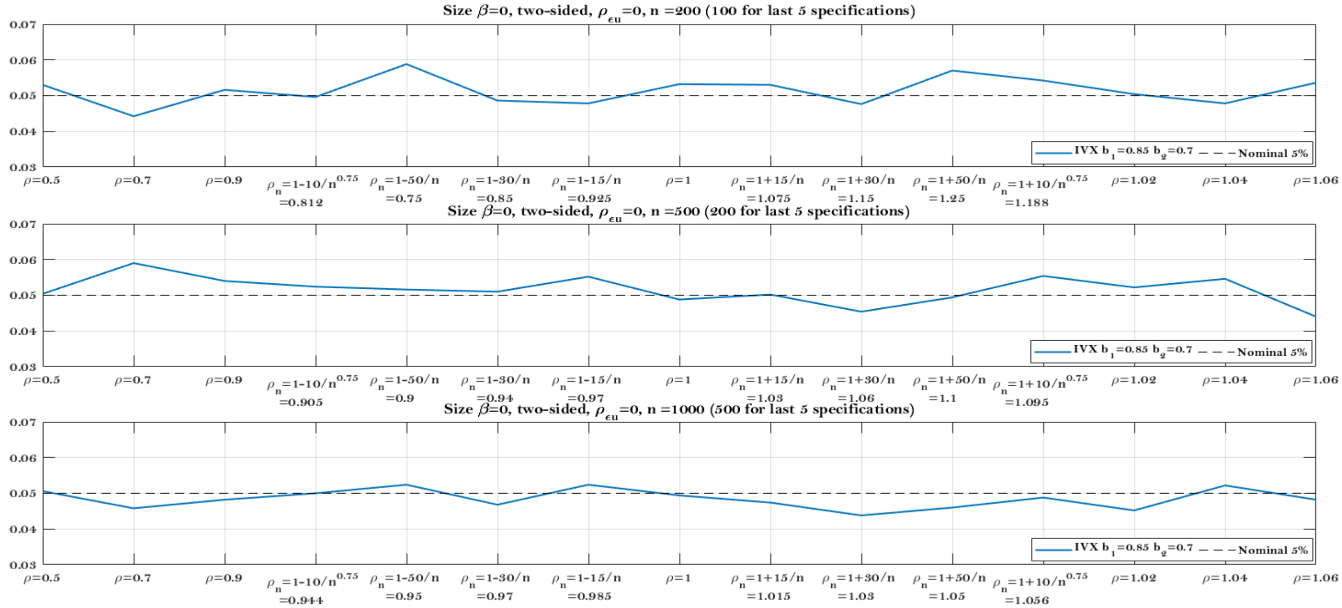
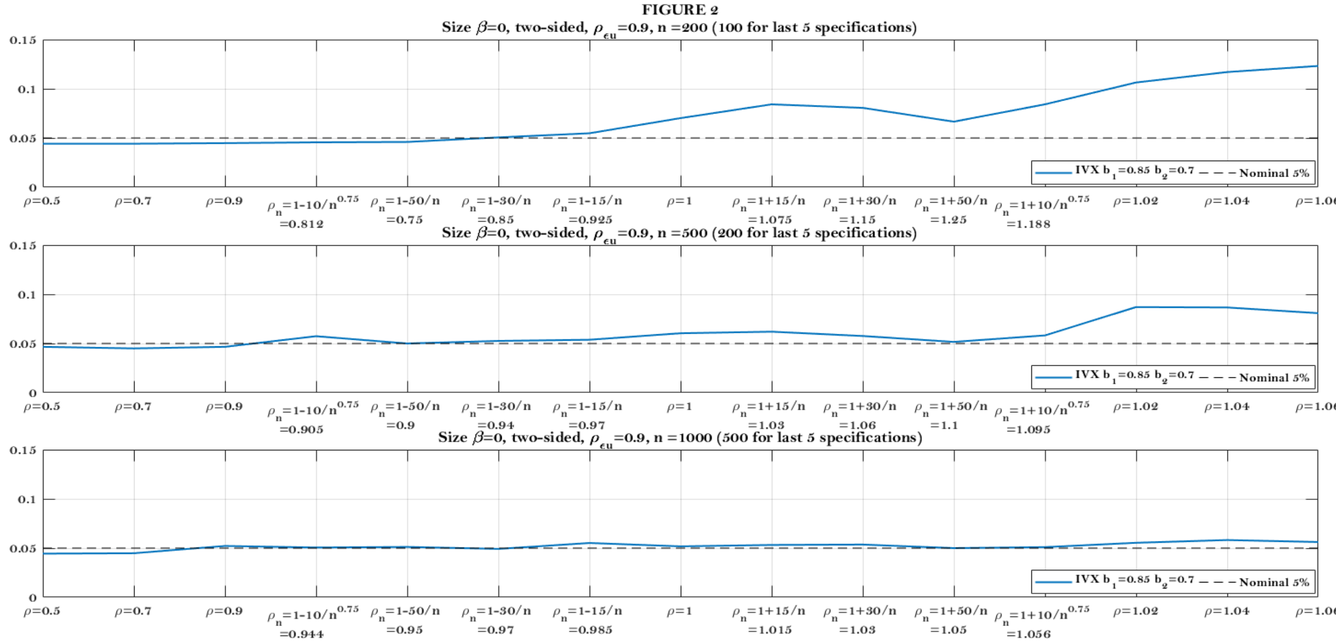


FIGURE 2

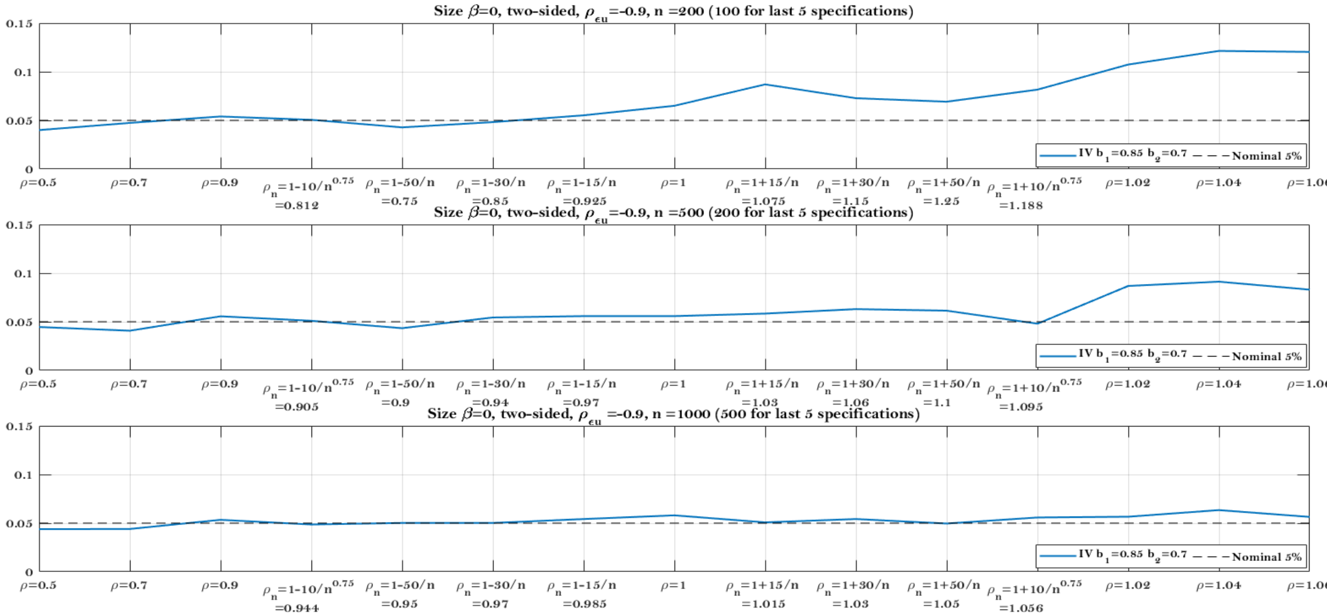


We first implement our choice of instrument in the predictive regression setup (9) along different autoregressive regimes for x_t in (1):

$$\begin{aligned}\rho_n \;&\in\; \{0.5,\ 0.7,\ 0.9,\ 1-10/n^{0.75},\ 1-50/n,\ 1-30/n,\ 1-15/n,\ 1,\ 1+15/n,\\ &\qquad 1+30/n,\ 1+50/n,\ 1+10/n^{0.75},\ 1.02,\ 1.04,\ 1.06\},\ \text{with } X_0=0,\ \mu=\mu_y=0, \quad (42)\\ \varepsilon_t \;&\sim\; \mathcal{N}\left(0,\sigma_\varepsilon^2\right),\ u_t \sim \mathcal{N}\left(0,\sigma^2\right),\ \sigma_\varepsilon=\sigma=1,\ \rho_{\varepsilon u}\in\{-0.9,-0.45,0,0.45,0.9\}. \quad (43)\end{aligned}$$

For each specification, we compute the empirical size of the two-sided test statistic in (26) based on 5,000 simulated samples for sample sizes $n \in \{200, 500, 1000\}$. Throughout the entire Monte Carlo section, we always use reduced sample sizes $n \in \{100, 200, 500\}$ for the explosive specifications $\rho_n \in \{1+50/n, 1+10/n^{0.75}, 1.02, 1.04, 1.06\}$. We do this for two reasons: (i) it facilitates comparison since the exponential rate of convergence for these specifications implies extremely precise estimates with standard errors of the range of 10^{-30} for sample sizes of 500, and (ii) it prevents Matlab rounding such standard errors to 0 (resulting to point confidence intervals) without the need of committing excessive memory.

FIGURE 3



Figures 1-3 display the rejection probability of our test procedure in (26) 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. The online 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 the mildly explosive instrument is chosen throughout the different autoregressive specifications. As expected, the mildly explosive instrument is never chosen in the stationary region even for small samples, and is chosen in the pure unit root case around 33% of the time.

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 region 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 (9), 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 with 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.

FIGURE 4

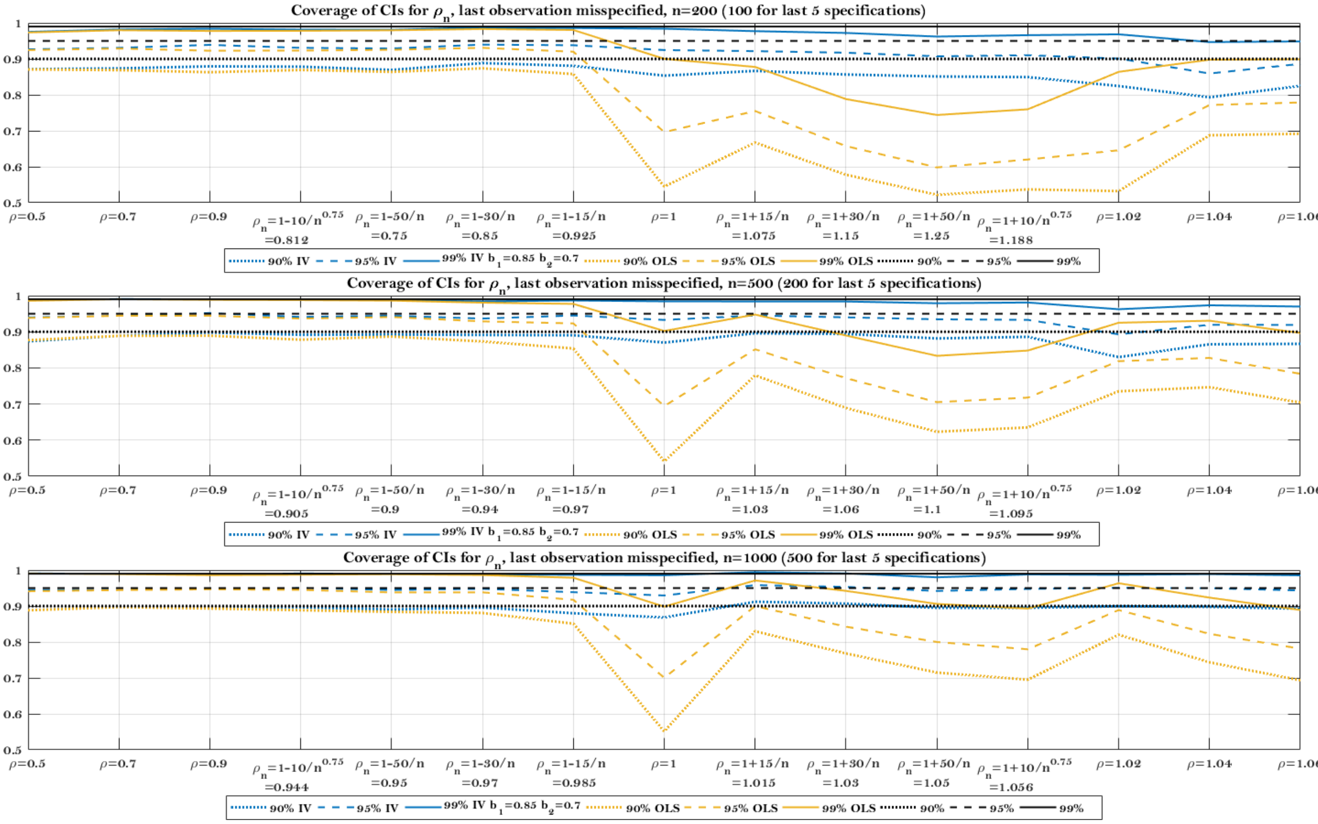
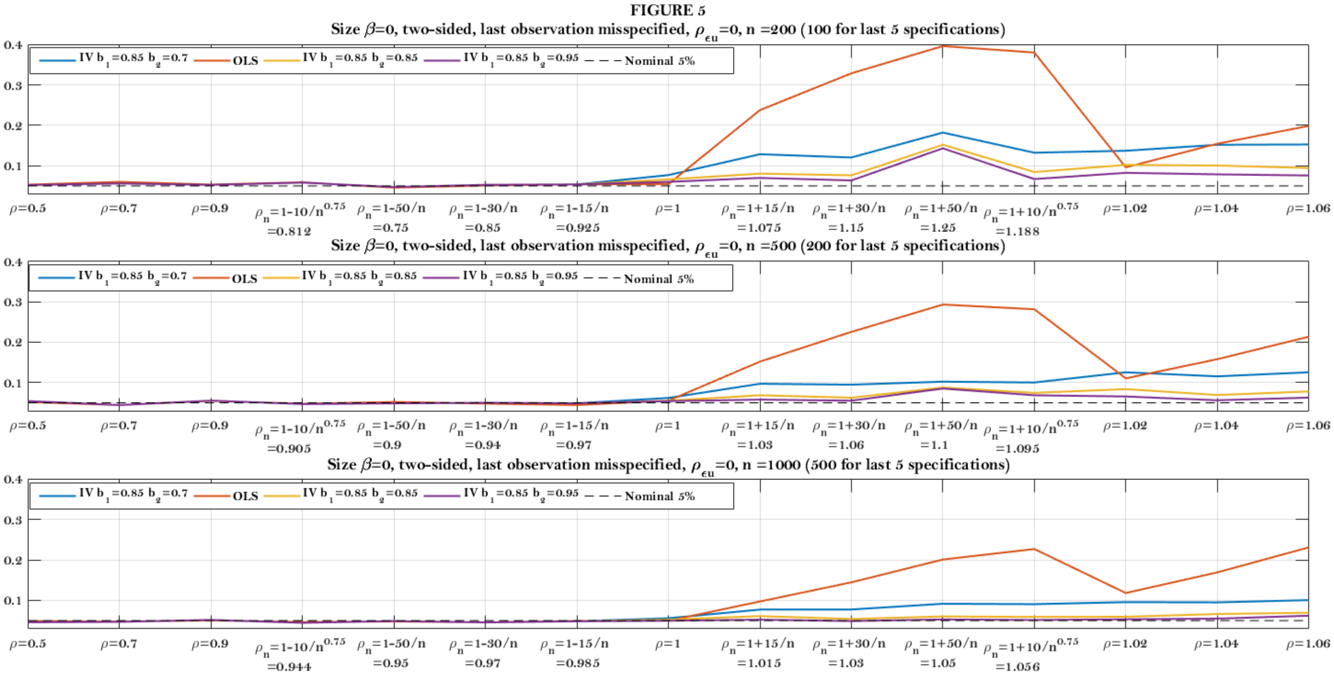


FIGURE 5



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 (42) 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% confidence intervals implied by the IV and OLS respectively, based on 5,000 replications. 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 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 $\rho_n = 1 + 10/n^{0.75}$ 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 $\rho_n = 1.06$, the 90% OLS confidence interval contains the truth 70% of the time irrespective of increases in the sample size.

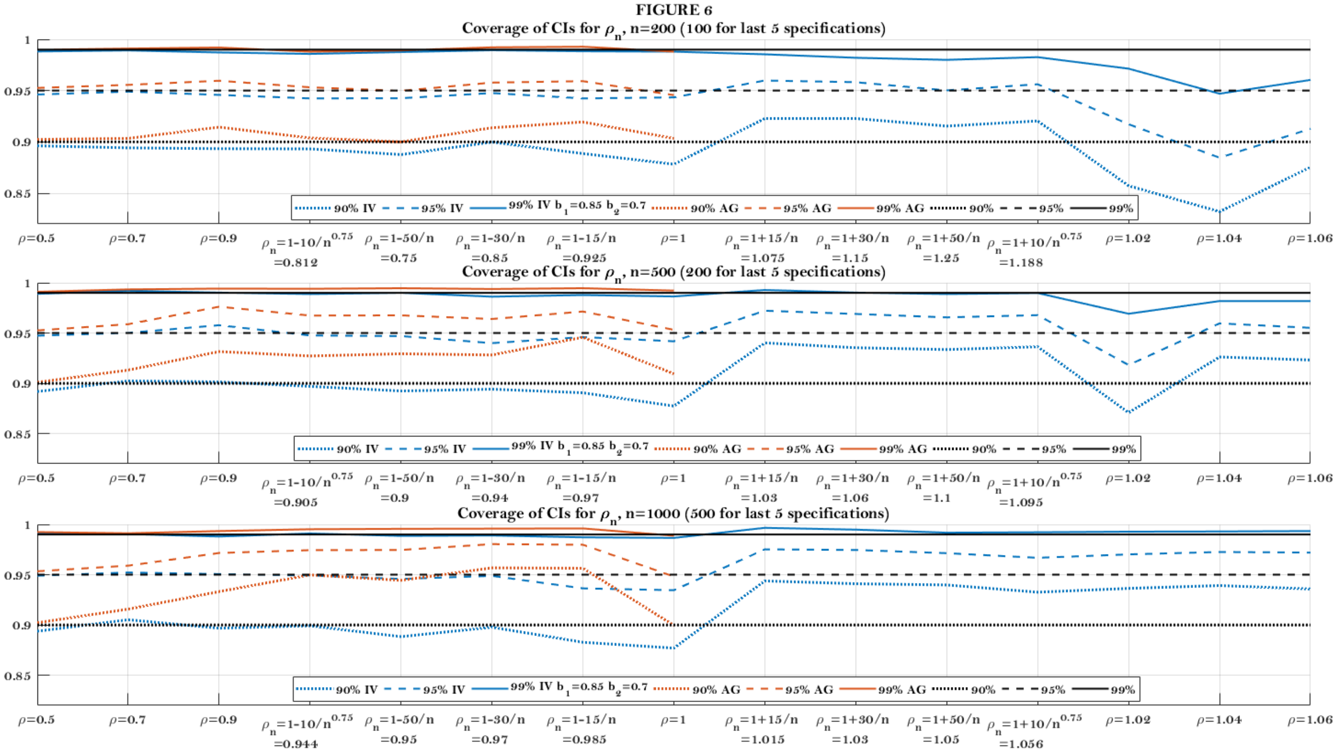
We find similar results in the predictive regression setup. In Figure 5, we report the rejection probability of the OLS under the null $\beta = 0$ against a two-sided alternative[10] for the same specifications and sample sizes. We present the rejection probability of the IV procedure for the choice 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 level as the sample size increases, as suggested by the theoretical results of Theorem 2.

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 autoregression and predictive regression for $\rho \in (-1, 1]$. We first evaluate our proposed autoregressive confidence intervals in (23) and we compare them to the procedure by Andrews and Guggenberger (2014)[11], which constructs the intervals by inverting the OLS t-statistic, which under the null is asymptotically nuisance-parameter-free.

FIGURE 6

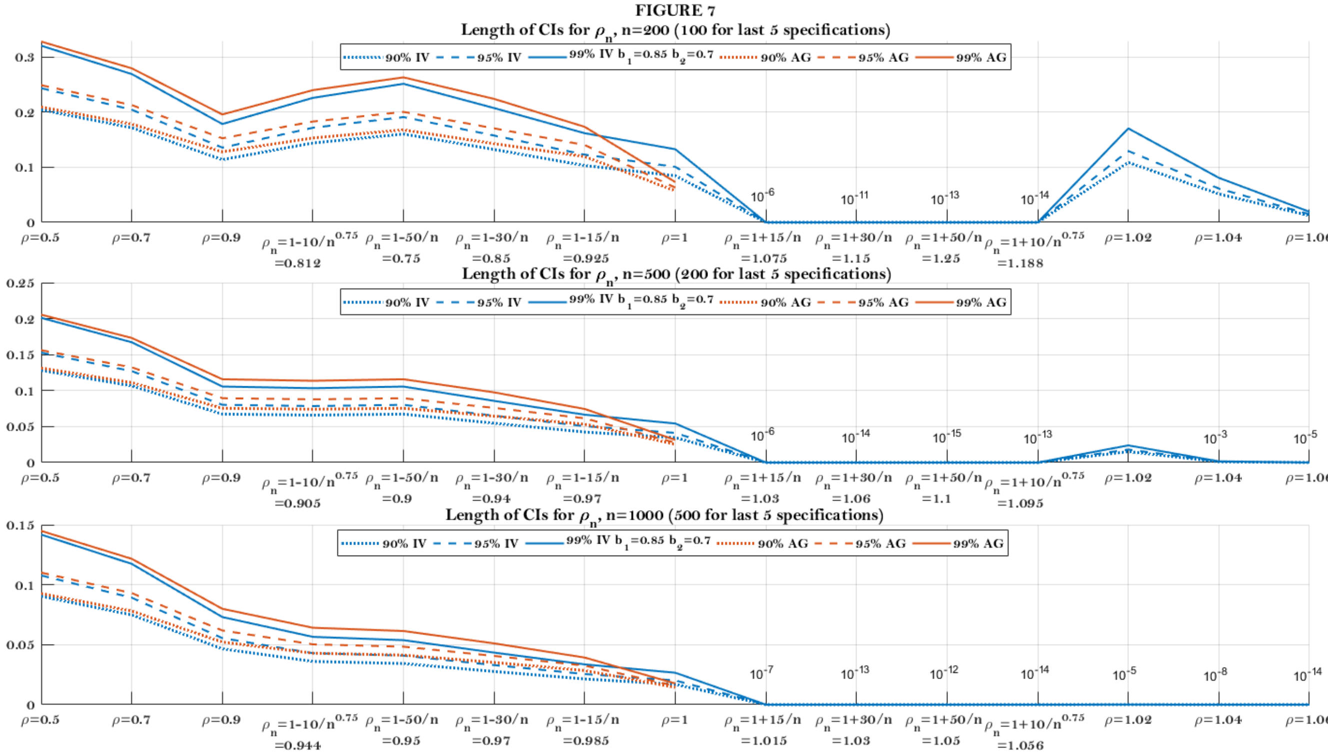


In Figure 6 and 7, we report the 90%, 95% and 99% coverage rates and lengths of the intervals

[10] The online Appendix B contains additional comparison for the corresponding one-sided rejection probabilities.
[11] The Gauss code for the procedure was kindly provided by Don Andrews and Patrik Guggenberger and translated into Matlab.

respectively for the IV estimator and Andrews and Guggenberger (2014)'s procedure (AG) 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; the online Appendix B also contains the equal-tailed two-sided intervals of Andrews and Guggenberger (2014). Figure 6 presents evidence that the IV procedure works well and is comparable to the AG procedure on the left side of unity, while also providing valid inference for ρ_n on the right side of unity in the local-to-unity, mildly and purely explosive regions. In terms of length of the intervals, from Figure 7, it can be seen that our intervals are always shorter[12] than those of AG (which translates into higher power) for all specifications except for the exact unit root case. The differences in interval length in the unit root case are not large and become negligible for large samples.

FIGURE 7



4.3.2 Size and power comparison in the predictive regression model

Next, we evaluate the inference based on the IV-based t-statistic in (26) in the predictive regression setup (9) and we compare it to the one-sided test procedure by Elliott et al. (2015)[13], 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 (9) for the specifications of (42) and (43). 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.

[12] This result on the shorter length of the IV intervals continues to hold in the comparison for the 90% and 99% equal-tailed two-sided intervals of AG, which can be found in the online Appendix.

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

Instead, we propose using the following adaptive t-statistic:

$$T_n^A\left(\tilde{\beta}_n\right) = \mathbf{1}\left\{\hat{\rho}_{\varepsilon u} \leq L\right\} T_n\left(\tilde{\beta}_{1n}\left(\breve{z}_{1,t}\right)\right) + \mathbf{1}\left\{\hat{\rho}_{\varepsilon u} > L\right\} T_n\left(\tilde{\beta}_{2n}\left(\breve{z}_{2,t}\right)\right) \tag{44}$$

where $T_n\left(\tilde{\beta}_{1n}\left(\breve{z}_{1,t}\right)\right)$ and $T_n\left(\tilde{\beta}_{2n}\left(\breve{z}_{2,t}\right)\right)$ are the t-statistics 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 fitted OLS residuals for u_t and ε_t, and L is some threshold level below which we use a more conservative instrument. In this way, we can resolve the size distortion in the unit root under strong negative correlation, without affecting the power of the our procedure in all other cases.

FIGURE 8

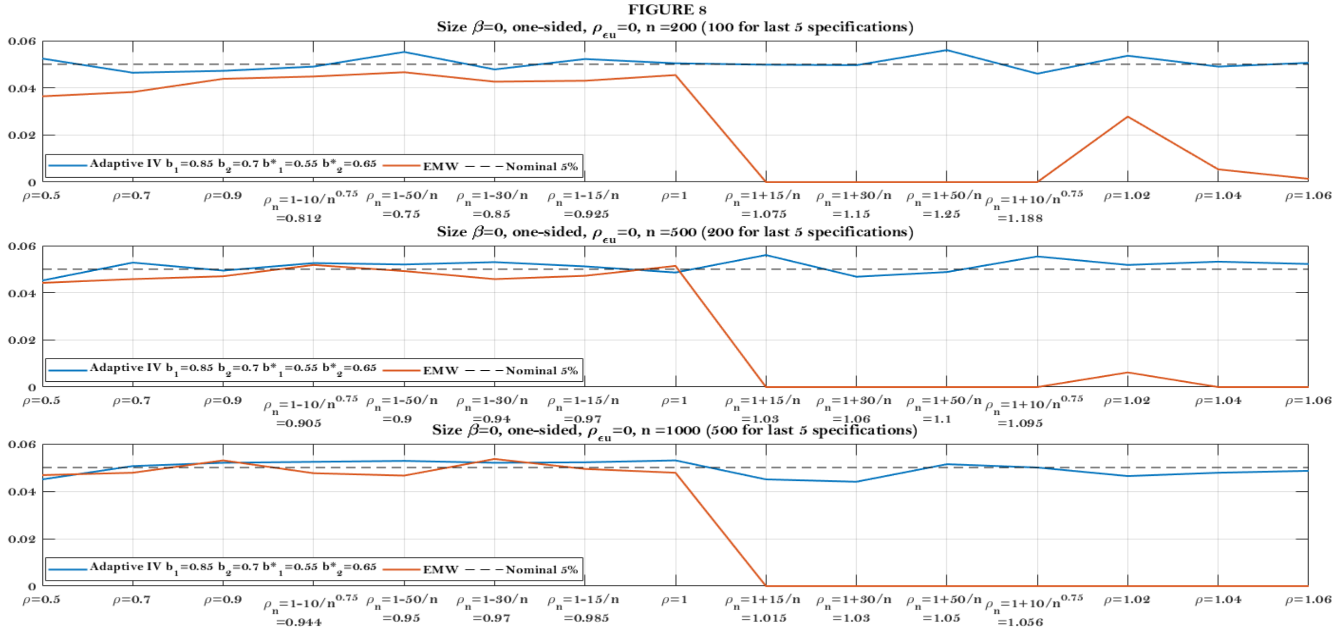


FIGURE 9

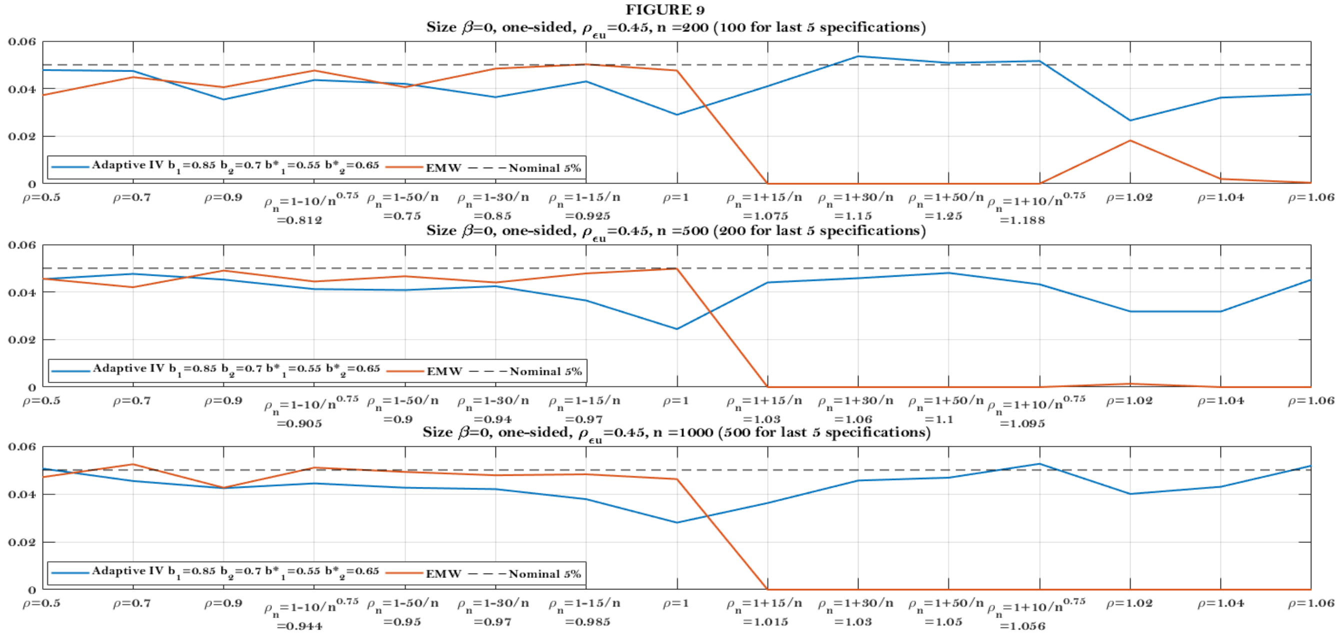


FIGURE 10

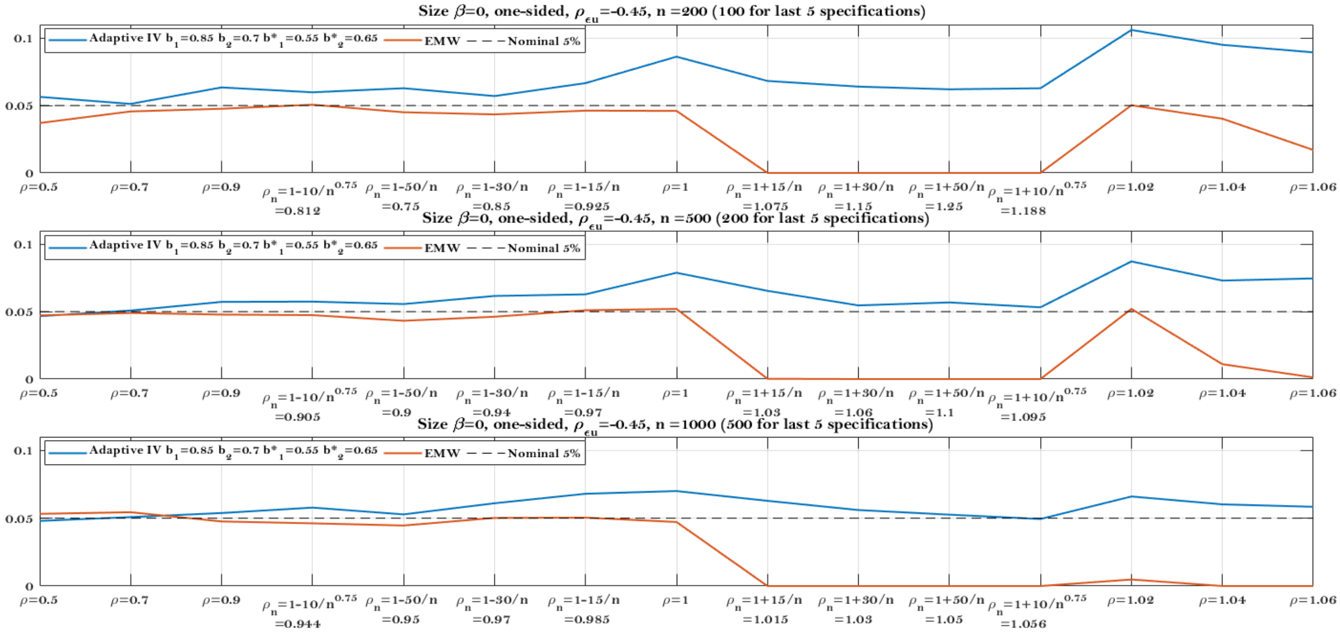
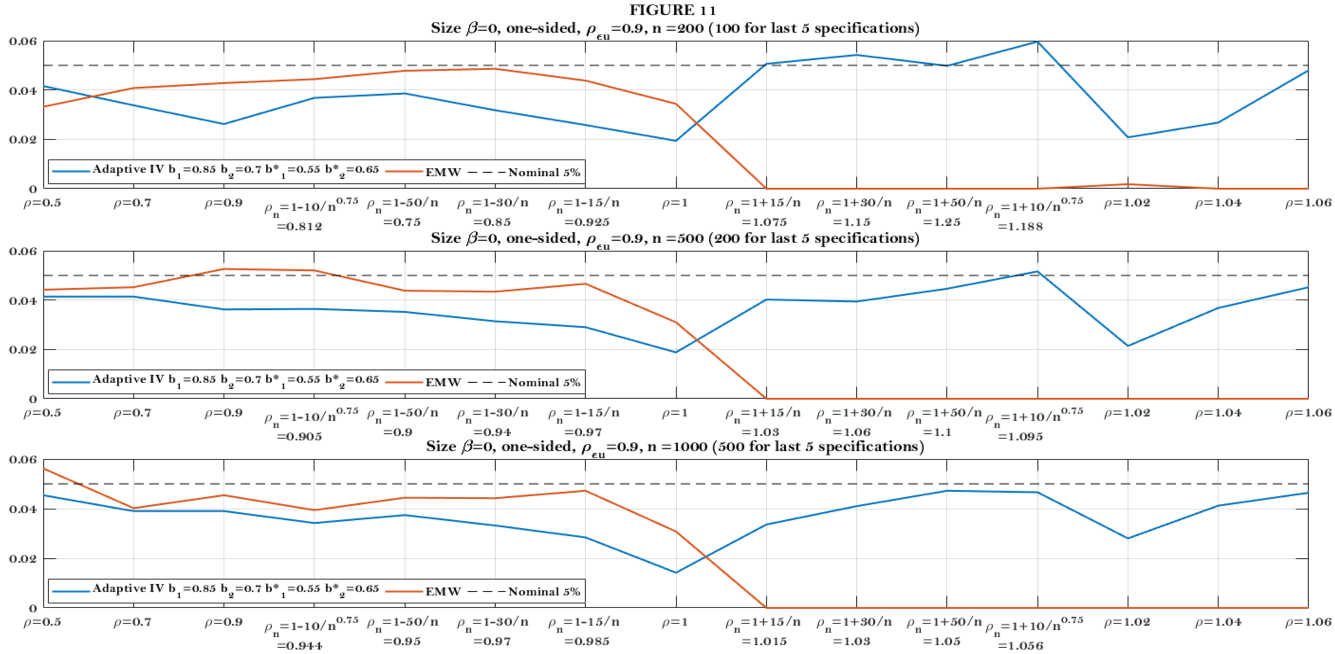


FIGURE 11



We set $L = -0.7$, and for the conservative instrument $\breve{z}_{1t}$, we use (41) with $b_1 = 0.55$ and $b_2 = 0.65$. For $\breve{z}_{2t}$, we continue to use the choice of instrument from Section 4.1 with $b_1 = 0.85$ and $b_2 = 0.7$. In the case of $\rho_{\varepsilon u} = -0.9$ in Figure 12, 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 (44) to illustrate the effect of using the adaptive procedure. For all other cases, Figures 8-11, 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 13-17 present the corresponding power curves. 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 hence it is invalid) on the right side of unity.

FIGURE 12

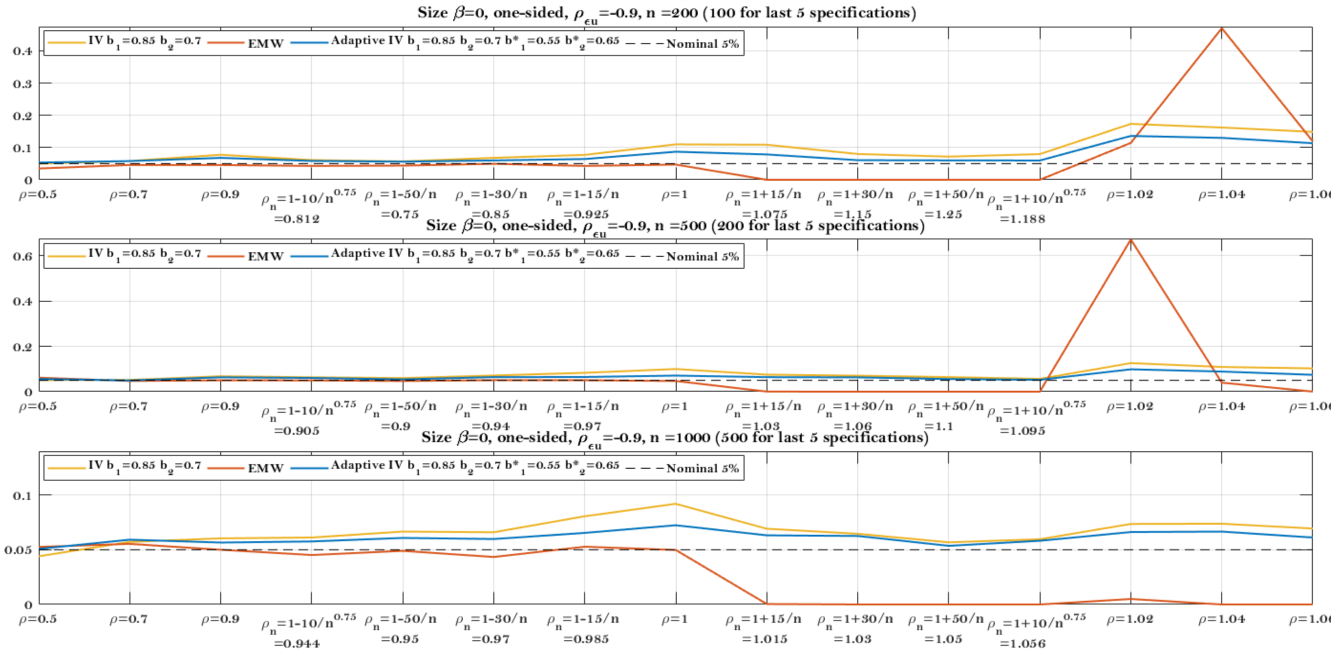
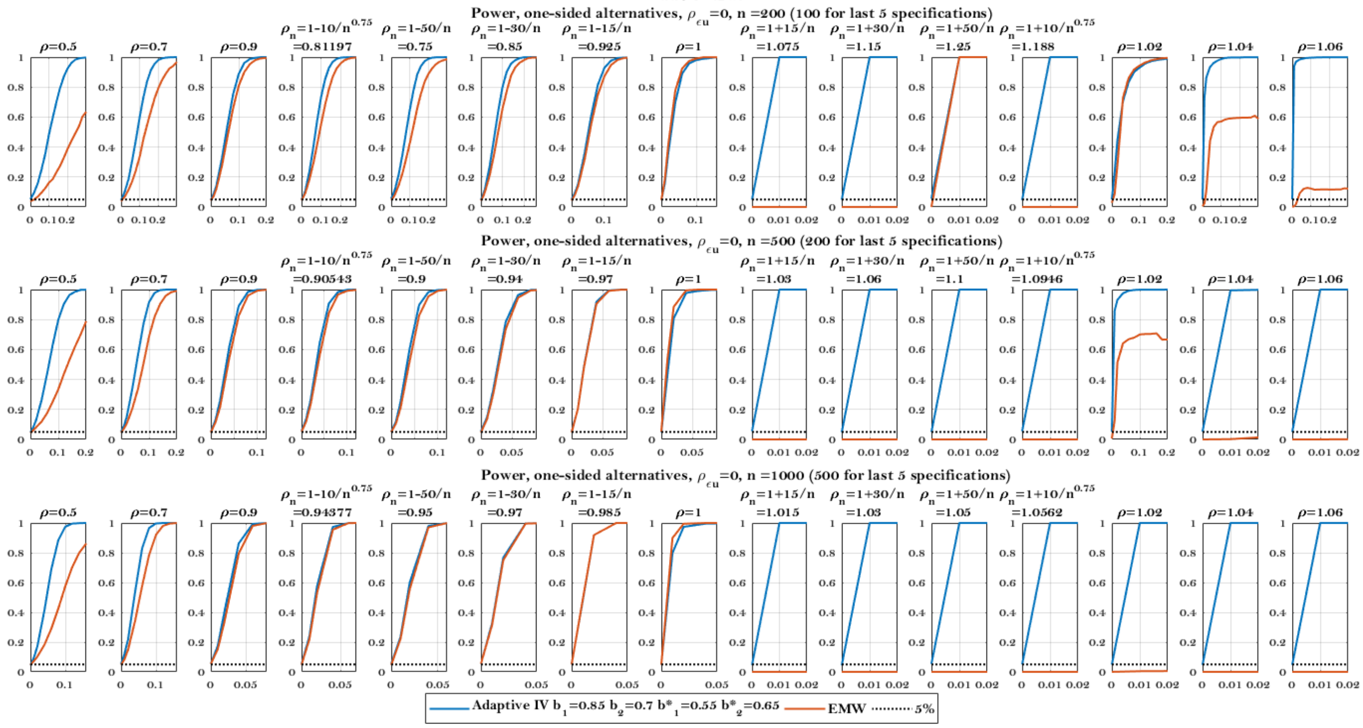


FIGURE 13



There are several important conclusions from the size and power comparisons in Figures 8-17. First, our adaptive procedure in (44) 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.

FIGURE 14

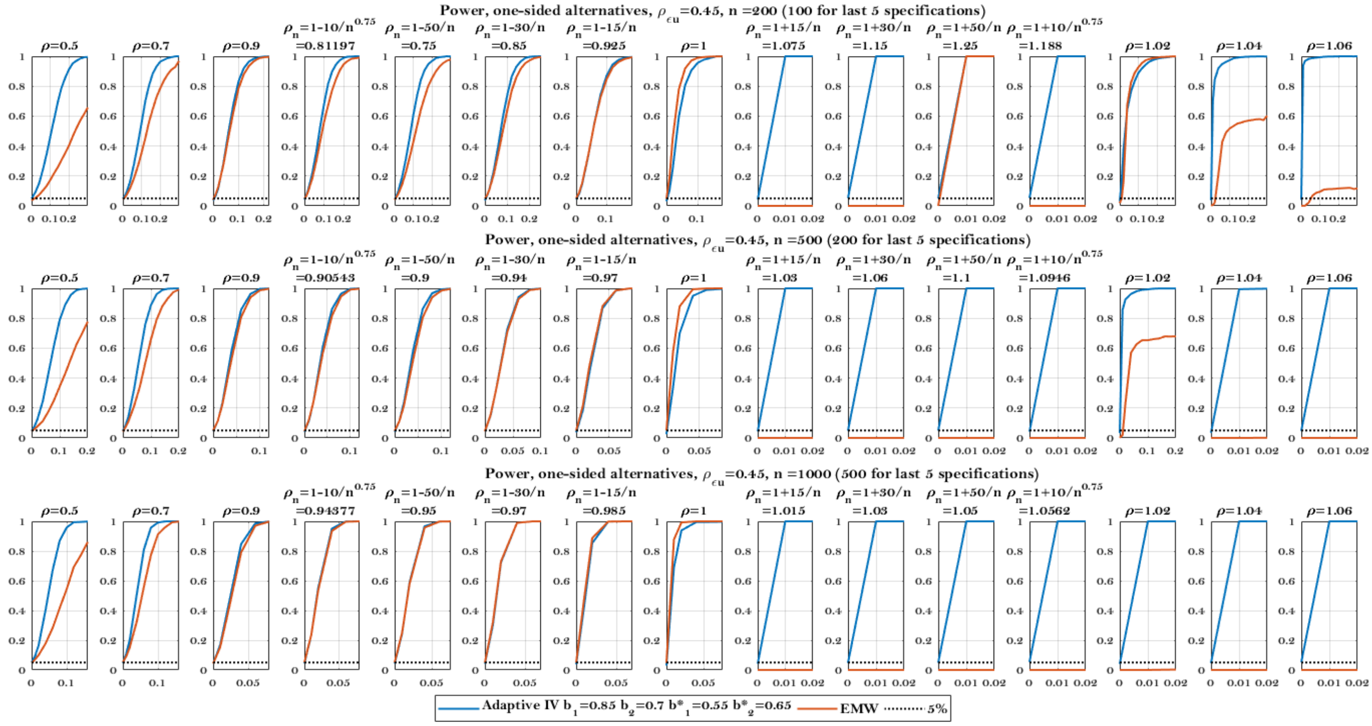
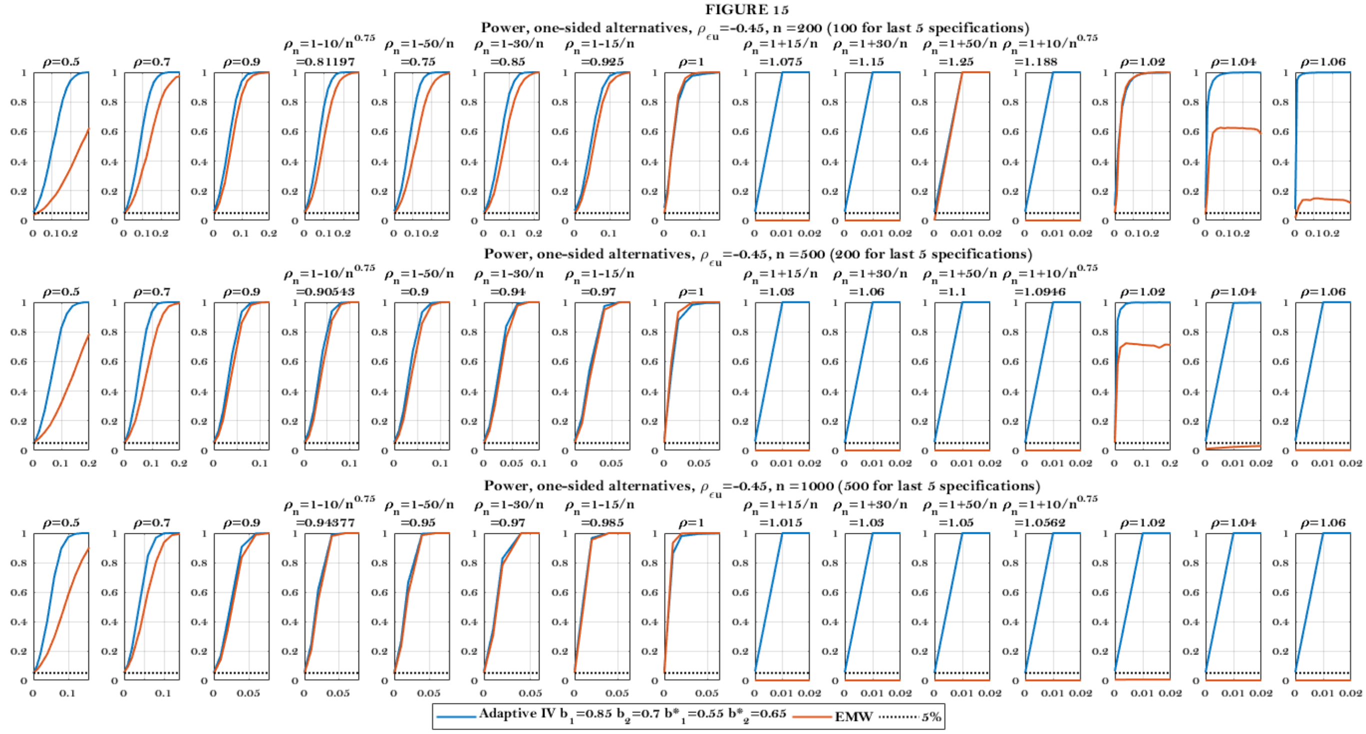


FIGURE 15



In terms of power, we reach a very similar conclusion to the one in the comparison with Andrews and Guggenberger (2014) in Section 4.3.1: namely, our procedure is always more powerful than EMW (which is asymptotically nearly-optimal under our Gaussian DGP for the innovations) 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. Crucially, our procedure extends in the right-side of unity and provides correct inference in the right side of local-to-unity, mildly explosive and pure explosive regions, for which alternative approaches are invalid.

FIGURE 16

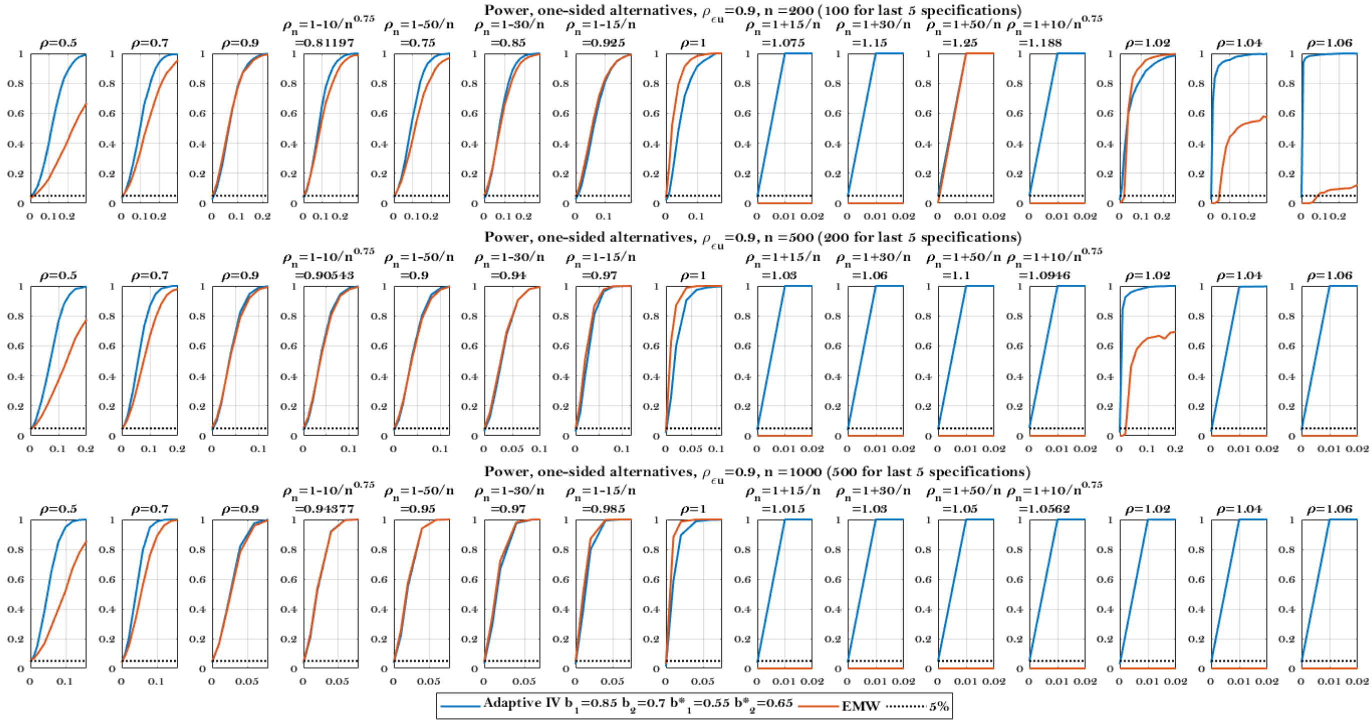
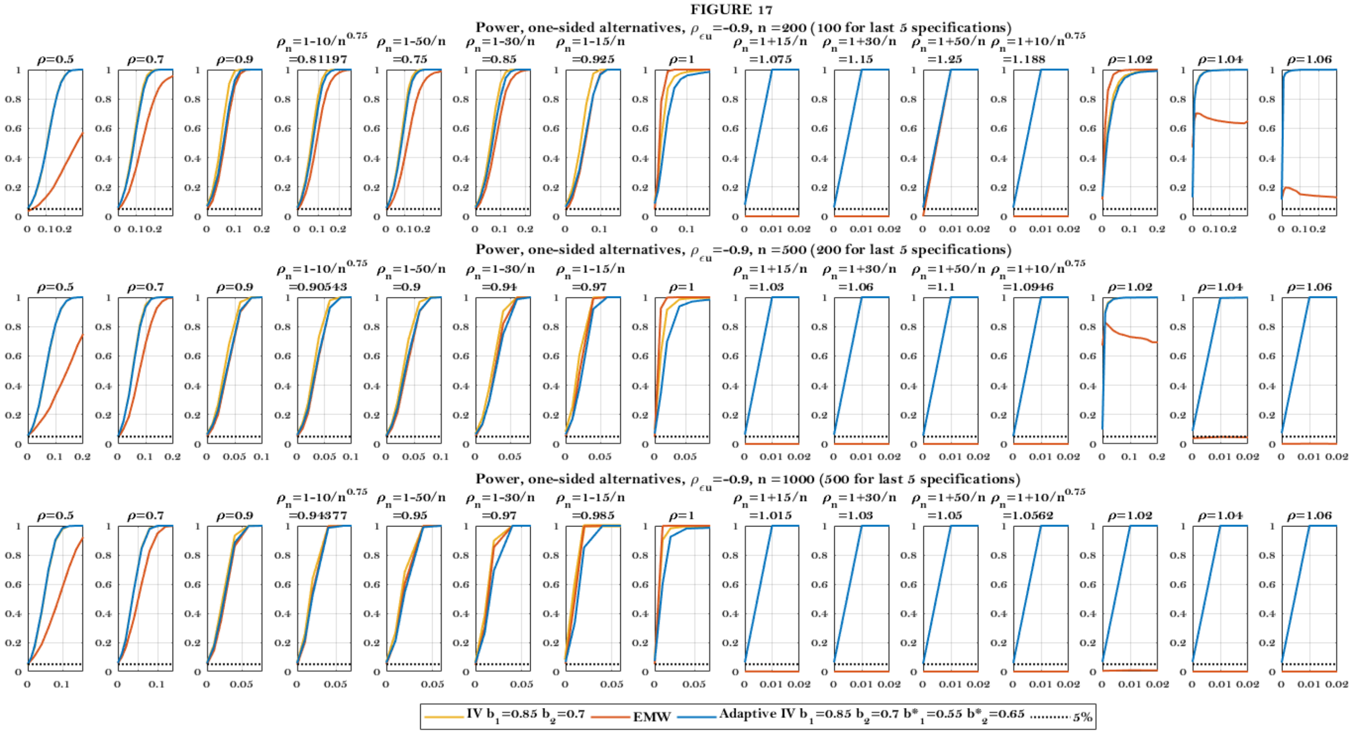


FIGURE 17



5 Inference in a linearised SIR model

In this section, we apply the procedure proposed in the paper on the linearised SIR model (12) on Covid-19 data in order to construct confidence intervals for the parameters θ, γ, δ and for the basic reproduction number r_0 across a panel of countries. As discussed in Section 2.3, the triangular system in (12) 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 has been to reduce r_0 (and hence ρ) 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 (for example, Liu, Moon and Schorfheide (2021) perform a fully parameteric 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 (12) 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 (confidence intervals), 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 assumptions or homogeneity 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 confidence intervals with correct coverage regardless of the value of $r_0 \in (0, \infty)$ is of great importance for policymakers.

We use a dataset on daily number of confirmed cases, recovered and deceased obtained from the John Hopkins University database[14] for Italy, Germany, Austria, Denmark, Israel and South Korea[15]. 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[16]. 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[17] (first reported death: 24/07/2020; 25/07/2020:26/11/2020, 27/11/2020:31/03/2021, 01/04/2020:04/08/2021). We construct the confidence intervals for θ, γ, δ and r_0 for each country and subsample, using the IV confidence intervals in Corollary 1. For the instrument construction, we use (41) with $b_1 = 0.85$ and $b_2 = 0.7$, which are the values we show work well for all autoregressive regions in the Monte Carlo exercise of Section 4. Our choice to conduct inference over subsamples is motivated by the unlikelihood that the model's parameters, for example r_0, have remained constant over time; this is since aggressive

[14] https://github.com/CSSEGISandData/COVID-19

[15] Our choice of countries is motivated by the availability and quality of series on the number of recovered individuals; for example, for many countries, data for recovered are not reported; and in many cases when series are available, they are often of poor quality or stop being updated at some point in the sample.

[16] Series on recovered after August 2021 are not available. Arguably in late 2021, the SIR model becomes inappropriate, since we start observing many re-infections due to mutations of the virus, so an SIS model (where there is probability of re-infection) may be more appropriate for analysis.

[17] To avoid any arbitrary sample split, we use the same dates for all countries, since they give us roughly the same number of observations in each subsample. We find that our results are robust to alternative sample splits.

government policies aimed at controlling the early dynamics of the epidemic were aimed at containing the outbreak either by 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), or later on, through vaccination by reducing the proportion of susceptibles S_0/N.

FIGURE 18

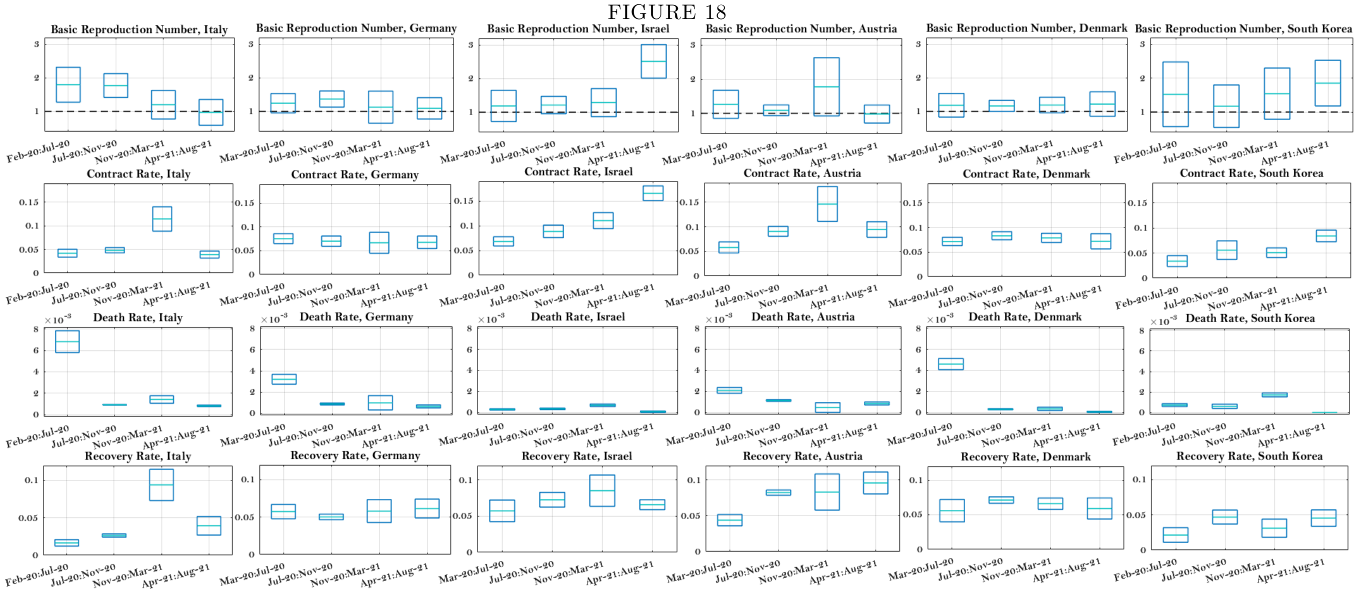


Figure 18 presents the IV estimates and 95% confidence intervals for r_0, θ, δ, and γ for each country and subsample. There are three main conclusions from our empirical analysis. First, the death rate has considerably fallen over time in all countries, and the recovery rate has increased over time for 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 value is usually above unity, one is most of time included in the 95% confidence interval. 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 we recognise that the linearised SIR model in (12) is a very simple and stylised model and that the data on Covid-19 infections 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 inference procedure proposed in this paper. Its main advantage is that it gives rise to confidence intervals 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 very close to unity depending on the various stages of the epidemic.

6 Conclusion

The paper proposes a unified, distribution-free framework of inference in both an autoregressive and predictive regression models, when the regressor's autoregressive root is in $(-1, \infty)$. This includes: (i) stable and near-stable regressor processes, (ii) unit root and local-to-unity regressors, and (iii) regressors that exhibit stochastic exponential growth (i.e. explosive and mildly explosive processes).

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 near-stationary and near-explosive root. The resulting IV estimators for the autoregressive 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 and gives rise to asymptotically correctly-sized confidence intervals. To our knowledge, this is the first method that delivers central limit theory and, consequently, general distribution-free asymptotic inference when the regressor is purely explosive. Crucially, the method also allows for inference with less persistent processes to the right of unity, i.e. mildly explosive and local-to-unity processes with $c > 0$ (which are assumed away by the local-to-unity literature), while remaining valid when the process is with root in the standard range of $(-1, 1]$.

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 the right side of unity $(1, \infty)$, where no existing alternative approach has general asymptotic validity.

Finally, we show that the basic SIR model for modelling epidemics' dynamics upon linearisation around the disease-free equilibrium, 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 the early dynamics of the Covid-19 epidemic across a panel of countries and construct confidence intervals 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.

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 autoregressive 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

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*, Forthcoming

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

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

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*

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

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 Lee, J.H. (2016) Robust Econometric Inference with Mixed Integrated and Mildly Explosive Regressors. *Journal of Econometrics*, **192**: 433-450

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

7 Appendix A

This Appendix contains an additional result (Lemma A1), and the mathematical proofs of Lemmata 1-5 and Theorems 1-3 of the paper. Some further auxiliary results, as well as the proof of Lemma A1 and Corollary 1, can be found in the Online Appendix B.

Lemma A1. *Let* $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ *denote the instruments in (19),* z_{1t}*, and* z_{2t} *the processes in (30) and (31). Let* x_{0t} *denote the zero-intercept autoregression in (3) and* $\tilde{z}_{0t} = \sum_{j=1}^{t} \varphi_{1n}^{t-j} \Delta x_{0j}$ *be an instrument generated by* x_{0t}*. 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) *Under C(i)-C(ii), the sequences* $r_{1n} = \pi_n^{-1} \sum_{t=1}^{n} (\tilde{z}_{1t-1} - \tilde{z}_{0t-1}) u_t$, $r_{2n} = \pi_n^{-2} \sum_{t=1}^{n} (\tilde{z}_{1t}^2 - \tilde{z}_{0t}^2)$, $r_{3n} = \pi_n^{-2} \sum_{t=1}^{n} (\tilde{z}_{1t} x_t - \tilde{z}_{0t} x_{0t})$ *with* $\pi_n = n^{1/2} (1-\rho_n^2 \varphi_{1n}^2)^{-1/2}$ *are all* $o_p(1)$.

(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 (6); 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),* $\tilde{z}_{2t}=z_{2t}-r_{nt}$, *where*

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

and $g_n=O_p\left(n^{-1/2}\left(\varphi_{2n}-1\right)^{-1}\right)$ *independently of* t. *Moreover, all sequences defined by* $R_{1n}=(\varphi_{2n}-1)\,\nu_{n,z}^{-1}\left(\sum_{t=1}^{n}\tilde{z}_{2t}-\sum_{t=1}^{n}z_{2t}\right)$, $R_{2n}=\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left(\sum_{t=1}^{n}\tilde{z}_{2t-1}u_t-z_{2t-1}u_t\right)$,

$R_{3n}=s_n^{-1}\sum_{t=1}^{n}\left(\tilde{z}_{2t}x_t-z_{2t}x_t\right)$, $R_{4n}=\left(\varphi_{2n}^2-1\right)^2\varphi_{2n}^{-2n}\left(\sum_{t=1}^{n}\tilde{z}_{2t}^2-\sum_{t=1}^{n}z_{2t}^2\right)$ *are* $o_p(1)$.

(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 (38). The following approximation holds:* $\left(\check{Y}_n,\check{Z}_n\right)-(Y_n,Z_n)\to_p 0$.

The proof of Lemma A1 can be found in the Online Appendix B.

Proof of Lemma 1. Convergence of $(\rho_n)_{n\in\mathbb{N}}$ to $\rho\in(-1,\infty)$ 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. First, we show that, when $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(i) or C(iii), the OLS estimator $\hat{\rho}_n$ satisfies

$$\epsilon_n:=\frac{\hat{\rho}_n-\rho_n}{1-\rho_n}\to_p\epsilon<1 \tag{A.2}$$

where ϵ is a non-random constant. When $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(i) $n(1-\rho_n)\to\infty$, and $\hat{\rho}_n-\rho_n=O_p\left(n^{-1/2}(1-\rho_n)^{1/2}\right)$ under Assumption 2 by Theorem 1 of Giraitis and Phillips (2006), giving $\epsilon_n=O_p\left([n(1-\rho_n)]^{-1/2}\right)=o_p(1)$ and $\epsilon=0$; under Assumption 4, Lemma A0(i) implies that

$$\epsilon_n=(1+\rho_n)\left(1-\rho_n^2\right)^{-1}(\hat{\rho}_n-\rho_n)\to_p\epsilon(\rho):=\frac{(1+\rho)\,\Gamma}{\sigma^2+2\rho\Gamma},$$

and we need to show that $\epsilon(\rho)<1$ for all $\rho\in(-1,1)$. Since $\Gamma\leq 0\Rightarrow\epsilon(\rho)\leq 0$, it is sufficient to consider $\Gamma>0$. Differentiating $\epsilon(\rho)$ we obtain

$$\epsilon'(\rho)=\frac{\left(\sigma^2-2\Gamma\right)\Gamma}{\left(\sigma^2+2\rho\Gamma\right)^2}\ \text{ and }\ sign\left\{\epsilon'(\rho)\right\}=sign\left(\sigma^2-2\Gamma\right)$$

since $\Gamma>0$. Hence if $\sigma^2>2\Gamma$, $\epsilon(\cdot)$ is increasing so $\epsilon(\rho)\leq\lim_{\rho\to 1}\epsilon(\rho)=2\Gamma/\left(\sigma^2+2\Gamma\right)<1$; if $\sigma^2<2\Gamma$, $\epsilon(\cdot)$ is decreasing so $\epsilon(\rho)\leq\lim_{\rho\to -1}\epsilon(\rho)=0$; if $\sigma^2=2\Gamma$, $\epsilon(\rho)=1/2$. In all cases, $\epsilon(\rho)<1$ completing the proof of (A.2) under C(i). When $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(iii), Theorem 4.3 of Phillips and Magdalinos (2007) gives $\hat{\rho}_n-\rho_n=O_p\left(\rho_n^{-n}\left(\rho_n^2-1\right)\right)$, so $\epsilon_n=O_p\left(\rho_n^{-n}\right)$ proving (A.2) with $\epsilon=0$. Recalling the expression for F_n in (13), write

$$n(\hat{\rho}_n-1)=n(\rho_n-1)+n(\hat{\rho}_n-\rho_n)=n(\rho_n-1)(1-\epsilon_n)$$

where ϵ_n is defined in (A.2). By (A.2), we may choose some $\eta\in(0,1-\epsilon)$; for arbitrary $\delta>0$ and $m_n\to\infty$ we obtain

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

When $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(i), $n(\rho_n - 1) \to -\infty$ so there exists $n_0 \in \mathbb{N}$ such that $n(\rho_n - 1) < 0$ for all $n \geq n_0$; hence $\mathbb{P}(m_n \mathbf{1}_{\bar{F}_n} > \delta) \leq \mathbb{P}(\epsilon_n > 1-\eta)$ for all $n \geq n_0$ and all $\delta > 0$ so part (i) follows since $\mathbb{P}(\epsilon_n > 1-\eta) \to 0$ by (A.2) and the choice $\eta \in (0, 1-\epsilon)$. When $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(iii), using the same argument and choice of $\eta \in (0, 1-\epsilon)$, we may write

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

Since $n(\rho_n - 1) \to \infty$ under C(iii), there exists $n_1 \in \mathbb{N}$ such that $n(\rho_n - 1) > 0$ for all $n \geq n_1$ and all $\delta > 0$; hence, $\limsup_{n\to\infty} \mathbb{P}(m_n \mathbf{1}_{F_n} > \delta) \leq \limsup_{n\to\infty} \mathbb{P}(\epsilon_n > 1-\eta) = 0$ by (A.2).

Proof of Lemma 3. Using the approximation for r_{3n} in Lemma A1(ii), we may write

$$\begin{aligned}(1-\rho_n\varphi_{1n})\frac{1}{n}\sum_{t=1}^{n} x_{t-1}\tilde{z}_{t-1} &= (1-\rho_n\varphi_{1n})\frac{1}{n}\sum_{t=1}^{n} x_{0t-1}\tilde{z}_{0t-1} + o_p(1) \\ &= \sigma^2 + \frac{1}{n}\sum_{t=1}^{n}\tilde{z}_{0t-1}u_t + \frac{1}{n}(2\rho_n - 1)\sum_{t=1}^{n} x_{0t-1}u_t \\ &\quad + \rho_n(\rho_n - 1)\frac{1}{n}\sum_{t=1}^{n} x_{0t-1}^2 + o_p(1) \qquad \text{(A.3)}\end{aligned}$$

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 A1(iii) and $\bar{x}_{0n-1} = O_p\left(n^{-1/2}\kappa_n\right)$ 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 A1(iii) yields

$$n^{-1}(1-\rho_n\varphi_{1n})n\bar{z}_{1n-1}\bar{x}_{n-1} = \frac{x_{0n}}{n^{1/2}}\frac{1}{n^{3/2}}\sum_{j=1}^{n} x_{0j-1} + o_p(1). \qquad \text{(A.4)}$$

Combining (A.3)-(A.4) and using $(1-\rho_n^2\varphi_{1n}^2)/(1-\rho_n\varphi_{1n}) \sim 1+\rho_n$, we obtain that

$$\left(1-\rho_n^2\varphi_{1n}^2\right)\frac{1}{n}\sum_{t=1}^{n}\underline{x}_{t-1}\tilde{z}_{t-1} = \tilde{\Psi}_n + o_p(1) \qquad \text{(A.5)}$$

with $\tilde{\Psi}_n$ defined in (35) under C(i)-C(ii), with the term in (A.4) 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})

$$\frac{1}{n}\left(1-\rho_n^2\right)\sum_{t=1}^{n} x_{0t-1}^2 = \frac{1}{n}\sum_{t=1}^{n} u_t^2 + 2\rho_n\frac{1}{n}\sum_{t=1}^{n} x_{0t-1}u_t - \frac{1}{n}x_{0n}^2 \to_p \sigma^2 + 2\rho\Gamma \qquad \text{(A.6)}$$

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 A1(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 (A.6) 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 $n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)\sum_{t=1}^n x_{0t-1}\tilde{z}_{0t-1} \to_p \omega^2$ (recall that we are under C(i)), the result follows from

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

since $\varphi_{1n} \to 1$ and $\rho_n \to 1$. For part (iii), in view of the approximation for r_{1n} in Lemma A1(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 A1(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 (2) and (31) give

$$\begin{aligned}(\rho_n\varphi_{2n}-1)\sum_{t=1}^n x_{t-1}z_{2t-1} &= x_n z_{2n} - \varphi_{2n}\sum_{t=1}^n z_{2t-1}u_t - \rho_n\sum_{t=1}^n x_{t-1}u_t - \sum_{t=1}^n u_t^2 \\ &\quad +\varphi_{2n}\mu\left(\rho_n-1\right)\sum_{t=1}^n z_{2t-1} + \mu\left(\rho_n-1\right)\sum_{t=1}^n u_t \\ &= x_n z_{2n} + o_p\left(\nu_n\nu_{n,z}\right) \qquad\qquad (A.7)\end{aligned}$$

where the order of magnitude follows from: $\nu_n^{-1}\left(\rho_n-1\right)\sum_{t=1}^n u_t$ is of order $O_p(\rho_n^{-n}\left(n\left(\rho_n-1\right)\right)^{1/2}) = o_p(1)$ by Lemma A1(i) under C(iii) and $O_p\left(n^{-1}\right)$ under C(ii); $\nu_n^{-1}\nu_{n,z}^{-1}\sum_{t=1}^n z_{2t-1}u_t$ is of order $O_p[\nu_n^{-1}\left(\varphi_{2n}-1\right)^{-1/2}] = o_p(\rho_n^{-n}\left[n\left(\rho_n-1\right)\right]^{1/2}) = o_p(1)$ under C(iii) (by Lemma A1(i)) and $O_p[(n\left(\varphi_{2n}-1\right))^{-1/2}]$ under C(ii); by (B.3) $\nu_n^{-1}\nu_{n,z}^{-1}\sum_{t=1}^n x_{t-1}u_t$ is of order $O_p(\varphi_{2n}^{-n}\left[n\left(\varphi_{2n}-1\right)\right]^{1/2}) = o_p(1)$ by Lemma A1(i); finally, the recursion (31) gives

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

since $z_{2n} = O_p\left(\nu_{n,z}\right)$ and $\sum_{t=1}^n u_t = O_p\left(n^{1/2}\right)$. The proof of (A.7) follows by Lemma A1(i) and the fact that $n\left(\varphi_{2n}-1\right) \to \infty$. By (A.7), 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 (38) and (39).

For part (ii), the martingale approximation of Lemma A1(v) implies that

$$[Y_n, X_n]' = C(1)\sum_{j=1}^n c_{nj}e_j + o_p(1) \quad \text{with } c_{nj} = \left[\left(\varphi_{2n}^2-1\right)^{1/2}\varphi_{2n}^{-(n-j)-1}, \left(\rho_n^2-1\right)^{1/2}\rho_n^{-j}\right]' \qquad (A.8)$$

(for X_n we may use the part of Lemma A1(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 (A.8): 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: $V_{11}^{(n)} = \omega^2\left(\varphi_{2n}^2-1\right)\sum_{j=1}^n \varphi_{2n}^{-2j} \to \omega^2$; $V_{22}^{(n)} = \omega^2\left(\rho_n^2-1\right)\sum_{j=1}^n \rho_n^{-2j} \to \omega^2$; $V_{12}^{(n)} = \left(\rho_n^2-1\right)^{1/2}\left(\varphi_{2n}^2-1\right)^{1/2}\varphi_{2n}^{-n-1}\sum_{j=1}^n\left(\frac{\varphi_{2n}}{\rho_n}\right)^j$. When $n\left|\rho_n-\varphi_{2n}\right| \to \infty$, evaluating the geometric progression yields $V_{12}^{(n)} = O\left(\varphi_{2n}^{-n}\right)+O\left(\rho_n^{-n}\right)$; when $\left|\rho_n-\varphi_{2n}\right| = O\left(n^{-1}\right)$, $\sum_{j=1}^n\left(\varphi_{2n}/\rho_n\right)^j \le bn$ 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 A1(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 $\mathcal{N}\left(0,\omega^2\right)$ variates. For the Lindeberg condition associated with (A.8), the bound $\max_{j\leq n}\|c_{nj}\|^2 \leq 2\lambda_n$ with $\lambda_n = (\varphi_{2n}^2 - 1) \vee (\rho_n^2 - 1)$ yields

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

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 (8) is well-defined $a.s.$ because $\eta_n := \sum_{j=1}^n \rho^{-j} u_j$ converges $a.s.$ under Assumption 4: $\mathbb{P}\left(\sup_{k\geq 1}\left\|\eta_{n+k} - \eta_n\right\|_{L_1} > \delta\right) \leq \delta^{-1}\mathbb{E}\left(\sup_{k\geq 1}\left\|\eta_{n+k} - \eta_n\right\|_{L_1}\right)$ for any $\delta > 0$ and $\mathbb{E}\left(\sup_{k\geq 1}\left\|\eta_{n+k} - \eta_n\right\|_{L_1}\right) \leq \sup_{k\geq 1}\|u_k\|_{L_1} \sum_{j=n+1}^\infty |\rho|^{-j} \to 0$. 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 $\sum_{j=1}^n \left(\rho_n^{-j} - \rho^{-j}\right) u_j \to_{L_1} 0$, 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 $n \geq n_0(\delta)$: $\phi_n > \rho - \delta$ which implies that $\sum_{j=1}^n |\rho_n^{-j} - \rho^{-j}| = |\rho_n - \rho| \sum_{j=1}^n j\phi_n^{-j-1} \leq |\rho_n - \rho| \sum_{j=1}^\infty j(\rho - \delta)^{-j-1} \to 0$ since $\rho - \delta > 1$ from the choice $\delta \in (0, \rho - 1)$. Next we show that $X_\infty \neq 0$ $a.s.$ 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 $\pi_i \neq 0$ 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 $\mathbb{E}_{\mathcal{F}_{t-1}}\left(e_t^2\right) = \sigma^2$ 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 $\mathbb{P}(X_\infty = 0) = \mathbb{P}\left(\lim_{n\to\infty} \sum_{i=1}^n \pi_i e_i = -Y_0\right) = 0$.

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 $k_n/\left(\varphi_{2n}^2 - 1\right)^{-1} \to \infty$ and let $Y_n' = \left(\varphi_{2n}^2 - 1\right)^{1/2} C(1) \sum_{t=k_n}^n \varphi_{2n}^{-(n-t)-1} e_t$. 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 A1(v) implies that $\|Y_n - Y_n'\| = o_p(1)$. Also,

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

by Assumption 3. Using the fact that $X_\infty \neq 0$ $a.s.$ 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

$$g(X_n) Y_n = g(X_{k_n-1}) Y_n' + o_p(1) = \sum_{t=0}^{n-k_n} \xi_{n,t} + o_p(1) \tag{A.9}$$

where $\xi_{n,t} = C(1)\zeta_{nt} e_{t+k_n}$ and $\zeta_{nt} = \left(\varphi_{2n}^2 - 1\right)^{1/2} \varphi_{2n}^{-(n-k_n-t)-1} g(X_{k_n-1})$. Since ζ_{nt} is an $\mathcal{F}_{k_n-1}$-measurable sequence for all n, t, $\left\{\left(\xi_{n,t}, \mathcal{F}_{n,t}\right) : 0 \leq t \leq n - k_n\right\}$ 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 (A.9) is given by

$$\begin{aligned}\sum_{t=0}^{n-k_n} \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\xi_{n,t}^2\right) &= \omega^2 \sum_{t=0}^{n-k_n} \zeta_{nt}^2 = \omega^2 g^2(X_{k_n-1})\left(\varphi_{2n}^2 - 1\right) \sum_{t=0}^{n-k_n} \varphi_{2n}^{-2(n-k_n-t+1)} \\ &\to_p \omega^2 g(X_\infty)^2. \end{aligned} \tag{A.10}$$

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 $\lambda_n(\delta) := C(1)^{-1}\left(\varphi_{2n}^2 - 1\right)^{-1/2}\delta^{1/2}$ 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

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

since both $g^2\left(X_{k_n-1}\right)$ 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 $\delta > 0$, $L_n(\delta) \leq o_p(1) + g^2\left(X_{k_n-1}\right)o_p(1) = o_p(1)$ proving the Lindeberg condition. In view of (A.10), the martingale central limit theorem applied to $\sum_{t=0}^{n-k_n}\xi_{n,t}$ in (A.9) 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 A1(v) for $Y_n(s)$ and a standard approximation for $B_n(s)$ give

$$\left[U_n(s), B_n(s), Y_n(s)\right]' = \sum_{t=1}^{\lfloor ns\rfloor}\xi_{nt} + o_p(1). \qquad \text{(A.11)}$$

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 (A.11): 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

$$\sum_{t=1}^{\lfloor ns\rfloor}\mathbb{E}\left(\xi_{3,nt}^2\mathbf{1}\left\{\xi_{3,nt}^2 > \delta\right\}\right) \leq C(1)^2\max_{t\leq n}\mathbb{E}\left(e_t^2\mathbf{1}\left\{e_t^2 > \lambda_n(\delta)^2\right\}\right)\left(\varphi_{2n}^2 - 1\right)\sum_{t=1}^{n}\varphi_{2n}^{-2t} \qquad \text{(A.12)}$$

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 (A.12) is $o(1)$. This establishes the LC for the martingale difference array ξ_{nt} in (A.11). The conditional variance matrix of the array in (A.11) 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 (A.11), $\sum_{t=1}^{\lfloor ns\rfloor}\xi_{nt} \Rightarrow \xi(s)$ 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, $\xi(s) =_d \left[U(s), B(s), Y\right]'$ 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 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 (35) we obtain, under C(i)-C(ii),

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left[\tilde{\rho}_{1n}-\rho_n,\tilde{\beta}_{1n}-\beta\right]=\left[1+o_p\left(1\right)\right]\frac{1}{\tilde{\Psi}_n}\left[\tilde{U}_n\left(1\right),\tilde{U}_n^{\varepsilon}\left(1\right)\right]\tag{A.13}$$

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.

We now prove part (i) of the theorem for $\tilde{\rho}_n$: under C(i) and Assumption 2, $u_t=e_t$ and $\Gamma=0$ so $\tilde{\Psi}\left(c\right)=\sigma^2$ and $\tilde{U}_n\left(1\right)\rightarrow_d\mathcal{N}\left(0,\sigma^4\right)$ by Lemma 3, so substituting into (A.13) yields

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left(\tilde{\rho}_{1n}-\rho_n\right)\rightarrow_d\mathcal{N}\left(0,1\right).\tag{A.14}$$

We complete the proof by combining (A.14) with Lemma 2: under C(i), Lemma 2(i) implies that $m_n\mathbf{1}_{\bar{F}_n}\rightarrow_p 0$ for $m_n:=\pi_n\left\|\tilde{\rho}_{2n}-\tilde{\rho}_{1n}\right\|_{L_1}$, so (20) yields $\pi_n\left\|\tilde{\rho}_n-\tilde{\rho}_{1n}\right\|_{L_1}=\pi_n\left\|\tilde{\rho}_{2n}-\tilde{\rho}_{1n}\right\|_{L_1}\mathbf{1}_{\bar{F}_n}\rightarrow_p 0$ and the proof of $\pi_n\left(\tilde{\rho}_n-\rho_n\right)\rightarrow_d\mathcal{N}\left(0,1\right)$ under C(i) and Assumption 2 follows from (A.14). 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 $\tilde{U}_n^{\varepsilon}\left(1\right)\rightarrow_d\mathcal{N}\left(0,\left(\sigma^2+2\rho\Gamma\right)\sigma_{\varepsilon}^2\right)$ by Lemma 3(iii) with the martingale difference e_t replaced by ε_t, giving

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left(\tilde{\beta}_{1n}-\beta\right)\rightarrow_d\mathcal{N}\left(0,\sigma_{\varepsilon}^2/\left(\sigma^2+2\rho\Gamma\right)\right).\tag{A.15}$$

Choosing $m_n=\pi_n\left\|\tilde{\beta}_{2n}-\tilde{\beta}_{1n}\right\|_{L_1}$ in Lemma 2(i) shows that $\pi_n\left\|\tilde{\beta}_n-\tilde{\beta}_{1n}\right\|_{L_1}=o_p\left(1\right)$, so the left side of (A.15) is asymptotically equivalent to $\pi_n(\tilde{\beta}_n-\beta)$, completing the proof of part (i).

Under C(ii)-C(iii) of Assumption 1b, $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left|n\bar{z}_{2n-1}\bar{u}_n-\bar{u}_n\sum_{t=1}^n z_{2t-1}\right|=o_p\left(1\right)$ from $R_{1n}=o_p\left(1\right)$ in Lemma A1(iv). 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 $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left(n\bar{z}_{2n-1}\bar{u}_n\right)=o_p\left(1\right)$. For $\tilde{\beta}_{2n}$, $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left(n\bar{z}_{2n-1}\bar{\varepsilon}_n\right)=o_p\left(1\right)$ by a similar argument. The above and Lemma 4(i) imply that the numerators of $\pi_n\left(\tilde{\rho}_{2n}-\rho_n\right)$ and $\pi_n(\tilde{\beta}_{2n}-\beta)$ are asymptotically equivalent to

$$\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left[\sum_{t=1}^n\tilde{z}_{2t-1}u_t,\sum_{t=1}^n\tilde{z}_{2t-1}\varepsilon_t\right]=\left[Y_nZ_n,Y_n^{\varepsilon}Z_n\right]+o_p\left(1\right).\tag{A.16}$$

The approximation for R_{1n} in Lemma A1(iv) and (B.12) give

$$n\left(\rho_n\varphi_{2n}-1\right)\nu_n^{-1}\nu_{n,z}^{-1}\bar{x}_{n-1}\bar{z}_{2n-1}=\left[1+o_p\left(1\right)\right]\frac{\kappa_n}{n}\nu_{n,z}^{-1}\left(\rho_n\varphi_{2n}-1\right)\sum_{t=1}^n z_{2t-1}\frac{\nu_n^{-1}}{\kappa_n}\sum_{j=1}^n x_{j-1}\tag{A.17}$$

which is $o_p\left(1\right)$ under C(iii): $O_p\left(\kappa_n/n\right)$ if $\left(\rho_n-1\right)/\left(\varphi_{2n}-1\right)\rightarrow 0$ and $O_p(\left(\left(\varphi_{2n}-1\right)^{-1}/n\right))$ if $\left(\varphi_{2n}-1\right)/\left(\rho_n-1\right)=O\left(1\right)$. Under C(ii), (A.17) becomes $Z_nn^{-3/2}\sum_{j=1}^n x_{j-1}+o_p\left(1\right)$ by (B.11), showing that (A.17) 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 $\pi_n\left(\tilde{\rho}_{2n}-\rho_n\right)$ and $\pi_n\left(\tilde{\beta}_{2n}-\beta\right)$ satisfies

$$s_n^{-1}\sum_{t=1}^n\underline{x}_{t-1}z_{2t-1}=Z_n\underline{X}_n+o_p\left(1\right),\quad\underline{X}_n:=X_n-n^{-3/2}\sum_{j=1}^n x_{j-1}\tag{A.18}$$

under C(ii)-C(iii) and Assumption 4, where Z_n and X_n are defined in (38) and (39). Recalling the definition of s_n in (37) and noting that $\rho_n\varphi_{2n}-1\sim\varphi_{2n}-1$ under C(ii), the normalisation under

C(ii)-C(iii) becomes

$$\frac{s_n}{\left(\varphi_{2n}^2-1\right)^{-1}\varphi_{2n}^n}=\frac{\nu_n\left(\varphi_{2n}^2-1\right)^{1/2}}{\rho_n\varphi_{2n}-1}\sim\begin{cases}2n^{1/2}\left(\varphi_{2n}^2-1\right)^{-1/2} & \text{under C(ii)}\\ \rho_n^n\left(\varphi_{2n}^2-1\right)^{1/2}\left(\rho_n^2-1\right)^{-1/2}\left(\rho_n\varphi_{2n}-1\right)^{-1} & \text{under C(iii)}\end{cases}\tag{A.19}$$

which is π_n under C(iii) and $2\pi_n$ under C(ii). Combining (A.18) and (A.16), we obtain

$$\frac{\left(\varphi_{2n}^2-1\right)^{1/2}\nu_n}{\rho_n\varphi_{2n}-1}\left[\tilde{\rho}_{2n}-\rho_n,\tilde{\beta}_{2n}-\beta\right]=\frac{1}{\underline{X}_n}\left[Y_n,Y_n^{\varepsilon}\right]+o_p\left(1\right)\tag{A.20}$$

under C(ii)-C(iii) and Assumption 4.

We now prove part (iii) of Theorem 3: under C(iii), $\underline{X}_n=X_n+o_p\left(1\right)$ and applying parts (ii) and (iii) of Lemma 4 and the continuous mapping theorem to (A.20) we obtain

$$\pi_n\left(\tilde{\rho}_{2n}-\rho_n\right)\to_d Y/X \;\text{ and }\; \pi_n\left(\tilde{\beta}_{2n}-\beta\right)\to_d Y^{\varepsilon}/X\tag{A.21}$$

where $X=_d\mathcal{N}\left(0,\omega^2\right)$ 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 $\left(Y,Y^{\varepsilon}\right)$ and $Y=_d\mathcal{N}\left(0,\sigma^2\right)$, $Y^{\varepsilon}=_d\mathcal{N}\left(0,\sigma_{\varepsilon}^2\right)$ by Lemma 4. Under Assumption 2, $\omega^2=\sigma^2$, so $Y/X=_d\mathcal{MN}\left(0,\sigma^2/X^2\right)$; under Assumption 4, $Y^{\varepsilon}/X=_d\mathcal{MN}\left(0,\sigma_{\varepsilon}^2/X^2\right)$. Thus, (A.21) gives the correct limit distributions for part (iii) of the theorem and it is enough to show that $\pi_n\left(\tilde{\rho}_n-\tilde{\rho}_{2n}\right)=o_p\left(1\right)$ and $\pi_n\left(\tilde{\beta}_n-\tilde{\beta}_{2n}\right)=o_p\left(1\right)$ under Assumption C(iii). By (20) and (21) $\pi_n\left\|\tilde{\rho}_n-\tilde{\rho}_{2n}\right\|_{L_1}=\pi_n\left\|\tilde{\rho}_{1n}-\tilde{\rho}_{2n}\right\|_{L_1}\mathbf{1}_{F_n}$ and $\pi_n\left\|\tilde{\beta}_n-\tilde{\beta}_{2n}\right\|_{L_1}=\pi_n\left\|\tilde{\beta}_{1n}-\tilde{\beta}_{2n}\right\|_{L_1}\mathbf{1}_{F_n}$ the right side being $o_p\left(1\right)$ by applying Lemma 2(ii) with the choices $m_n=\pi_n\left\|\tilde{\rho}_{1n}-\tilde{\rho}_{2n}\right\|_{L_1}$ and $m_n=\pi_n\left\|\tilde{\beta}_{1n}-\tilde{\beta}_{2n}\right\|_{L_1}$. Combined with (A.21), this shows part (iii) of the theorem.

We proceed to prove part (ii) of the theorem under Assumption C(ii). In the notation of (A.13) and Lemma 5, $\left|\tilde{U}_n\left(1\right)-U_n\left(1\right)\right|=o_p\left(1\right)$ by the approximation for r_{1n} of Lemma A1(ii) and Lemma 3.2(i) of MP(2020). Combining, (20), (A.13) and (A.20) and recalling the normalisation in (A.19) and the above approximation for $\tilde{U}_n\left(1\right)$, we obtain

$$\begin{aligned}\pi_n\left(\tilde{\rho}_n-\rho_n\right)&=n^{1/2}\left(1-\varphi_{1n}^2\right)^{-1/2}\left(\tilde{\rho}_{1n}-\rho_n\right)\mathbf{1}_{F_n}+n^{1/2}\left(\varphi_{2n}^2-1\right)^{-1/2}\left(\tilde{\rho}_{2n}-\rho_n\right)\mathbf{1}_{\bar{F}_n}\\&=\frac{U_n\left(1\right)}{\tilde{\Psi}_n}\mathbf{1}_{F_n}+\frac{1}{2}\frac{Y_n\left(1\right)}{\underline{X}_n}\mathbf{1}_{\bar{F}_n}+o_p\left(1\right)\end{aligned}\tag{A.22}$$

$U_n\left(\cdot\right)$ and $Y_n\left(\cdot\right)$ are defined in Lemma 5 (with $u_t=e_t$ under Assumption 2). $\tilde{\Psi}_n$ in (35), $n^{-1/2}x_n$, $\mathbf{1}_{F_n}$ and $\mathbf{1}_{\bar{F}_n}$ are functionals of $B_n\left(s\right)=n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor}u_t$, on $D\left[0,1\right]$, so the functional CLT of Lemma 5 on $\left[U_n\left(s\right),B_n\left(s\right),Y_n\left(s\right)\right]$ and the continuous mapping theorem imply that

$$\frac{U_n\left(1\right)}{\tilde{\Psi}_n}\mathbf{1}_{F_n}+\frac{1}{2}\frac{Y_n\left(1\right)}{\underline{X}_n}\mathbf{1}_{\bar{F}_n}\to_d\frac{U\left(1\right)}{\omega^2\Psi_-\left(c\right)}\mathbf{1}_{F_c}+\frac{Y}{\omega\Psi_+\left(c\right)}\mathbf{1}_{\bar{F}_c}\tag{A.23}$$

since, by Lemma 3(i), $\tilde{\Psi}_n\to_d\tilde{\Psi}\left(c\right)$ with $\sigma^2+2\rho\Gamma=\omega^2$ under C(ii), $\tilde{\Psi}\left(c\right)=\omega^2\Psi_-\left(c\right)$ on the event F_c and $2\left(J_c\left(1\right)-\int_0^1 J_c\left(r\right)dr\right)=\omega\Psi_+\left(c\right)$ on the event $\bar{F}_c$. The continuous mapping theorem is applicable to (A.23) because $x=0$ is the only discontinuity point of the function $x\mapsto\mathbf{1}_{\left(-\infty,0\right]}\left(x\right)$ and $\mathbb{P}\left(K_c+c=0\right)=0$ since K_c in (34) is a continuously distributed random variable for all $c\in\mathbb{R}$. Denoting $\zeta:=\left[\sigma^{-2}U\left(1\right),\sigma^{-1}Y\right]'$, Lemma 5 implies that ζ is independent of $\mathcal{F}_B=\sigma\left(B\left(s\right):s\in\left[0,1\right]\right)$ and $\zeta=_d\mathcal{N}\left(0,I_2\right)$. Since the random variables $J_c\left(1\right)$, $\Psi\left(c\right)$, $\mathbf{1}_{F_c}$ and $\mathbf{1}_{\bar{F}_c}$ are $\mathcal{F}_B$-measurable (as non-stochastic functionals of $B\left(r\right)$ on $D\left[0,1\right]$) the independence of ζ and $\mathcal{F}_B$ implies the independence of the random vectors ζ and $\left[J_c\left(1\right),\tilde{\Psi}\left(c\right),\mathbf{1}_{F_c},\mathbf{1}_{\bar{F}_c}\right]'$. Under Assumption 2, $\omega^2=\sigma^2$ and we conclude that the limit in (A.23) 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.22); 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 of part (ii).

It remains to prove that $\pi_n\left(\tilde{\beta}_n - \beta_n^*\right) \to_p 0$; since $\beta_n^* = \tilde{\beta}_n = \tilde{\beta}_{2n}$ on the event $\bar{F}_c$ 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 (21) and (25)

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

so it is enough to show that $\pi_n^{-1}x_n\bar{z}_{1n-1} \to_p 0$. By Lemma A1(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. By using (18) and the fact that $\hat{\sigma}_n \to_p \sigma$ and $\hat{\sigma}_\varepsilon \to_p \sigma_\varepsilon$ we obtain that $T_n(\tilde{\rho}_n) = [1 + o_p(1)]T_n$, and $T_n\left(\tilde{\beta}_n\right) = [1 + o_p(1)]T_n$ where

$$T_n = T_{1n}\mathbf{1}_{F_n} + T_{2n}\mathbf{1}_{\bar{F}_n}, \;\; T_{in} = \frac{|\Psi_{in}|}{\Psi_{in}}\zeta_{in}, \; \zeta_{in} = \frac{\sum_{t=1}^n \tilde{z}_{it-1}v_t}{\left(\sum_{t=1}^n \tilde{z}_{it-1}^2\right)^{1/2}}, \; \Psi_{in} = \textstyle\sum_{t=1}^n \underline{x}_{t-1}\tilde{z}_{it-1} \tag{A.24}$$

for $i \in \{1, 2\}$, where $v_t := u_t/\sigma$ for $T_n(\tilde{\rho}_n)$ and $v_t := \varepsilon_t/\sigma_\varepsilon$ for $T_n\left(\tilde{\beta}_n\right)$. 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 $T_n(\tilde{\rho}_n)$ under Assumption 2 and $T_n\left(\tilde{\beta}_n\right)$ 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. Since $T_{1n} = O_p(1)$ and $T_{2n} = O_p(1)$ by Lemmata 3 and 4, $|T_n - T_{1n}| = |T_{2n} - T_{1n}|\mathbf{1}_{\bar{F}_n} = o_p(1)$ under C(i) by Lemma 2(i) and $|T_n - T_{2n}| = |T_{2n} - T_{1n}|\mathbf{1}_{F_n} = o_p(1)$ under C(iii) by Lemma 2(ii). Under C(i), $\Psi_{1n} \to_p \sigma^2 + 2\rho\Gamma$ (Lemma 3(i)) and $\zeta_{1n} \to_d \mathcal{N}(0, 1)$ (by Lemma 3(iii) with $\sigma_e^2 = \mathbb{E}_{\mathcal{F}_{t-1}}(v_t^2) = 1$), so $T_{1n} = (1 + o_p(1))\zeta_{1n} \to_d \mathcal{N}(0, 1)$ as required. Under Assumption C(iii), Lemma 4 implies that $T_{2n} = (1 + o_p(1))(|X_n|/X_n)Y_n(v)$ with

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

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, $T_{2n} \to_d sign(X)\zeta_2 =_d \mathcal{N}(0, 1)$. 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 3, 4 and 5 give

$$T_n = \omega^{-1}\left(\left|\tilde{\Psi}_n\right|/\tilde{\Psi}_n\right)\hat{U}_n(1)\mathbf{1}_{F_n} + \left(|\underline{X}_n|/\underline{X}_n\right)\hat{Y}_n(1)\mathbf{1}_{\bar{F}_n} + o_p(1) \to_d T_c \tag{A.25}$$

where $T_c := sign(\Psi_1)\zeta_1\mathbf{1}_{F_c} + sign(\Psi_2)\zeta_2\mathbf{1}_{\bar{F}_c}$, $\Psi_1 = \omega^2 + J_c(1)^2 - 2J_c(1)\int_0^1 J_c(r)\,dr$, $\Psi_2 = J_c(1) - \int_0^1 J_c(r)\,dr$ 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

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

The above argument also gives $\mathbb{P}\left(\zeta_2 sign\left(\Psi_2\right) \leq x, \bar{F}_c\right) = \Phi\left(x\right)\mathbb{P}\left(\bar{F}_c\right)$, so the distribution function of the limit T_c in (A.25) is given by
$\mathbb{P}\left(T_c \leq x\right) = \mathbb{P}\left(\zeta_1 sign\left(\Psi_1\right) \leq x, F_c\right) + \mathbb{P}\left(\zeta_2 sign\left(\Psi_2\right) \leq x, \bar{F}_c\right) = \Phi\left(x\right)\left[\mathbb{P}\left(F_c\right) + \mathbb{P}\left(\bar{F}_c\right)\right] = \Phi\left(x\right).$
The above argument proves that T_n in (A.24) satisfies $T_n \to_d \mathcal{N}\left(0,1\right)$ for any $\left(v_t\right)$ satisfying Assumption 2 with $\mathbb{E}_{\mathcal{F}_{t-1}}\left(v_t^2\right) = 1$ *a.s.* under Assumption 1b, when $\left(\rho_n\right)_{n\in\mathbb{N}}$ in (1) belongs to one of the autoregressive classes C(i)-C(iii) of Assumption 1b.

Now suppose that $\left(\rho_n\right)_{n\in\mathbb{N}}$ in (1) satisfies Assumption 1a and consider an arbitrary subsequence $\left(\rho_{k_n}\right)_{n\in\mathbb{N}}$ of $\left(\rho_n\right)_{n\in\mathbb{N}}$ and $\left(T_{k_n}\right)_{n\in\mathbb{N}}$ of $\left(T_n\right)_{n\in\mathbb{N}}$. 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 $T_{m_n} \to_d \mathcal{N}\left(0,1\right)$. We conclude that for any subsequence $\left(T_{k_n}\right)_{n\in\mathbb{N}}$ of $\left(T_n\right)_{n\in\mathbb{N}}$ there exists a further subsequence $\left(T_{m_n}\right)_{n\in\mathbb{N}}$ of $\left(T_{k_n}\right)_{n\in\mathbb{N}}$ such that $T_{m_n} \to_d \mathcal{N}\left(0,1\right)$; but this implies that the entire sequence $\left(T_n\right)_{n\in\mathbb{N}}$ satisfies $T_n \to_d \mathcal{N}\left(0,1\right)$.

Using the argument following (A.24), we conclude that $T_n\left(\tilde{\rho}_n\right) \to_d \mathcal{N}\left(0,1\right)$ under Assumptions 1a, 2 and 3 and $T_n\left(\tilde{\beta}_n\right) \to_d \mathcal{N}\left(0,1\right)$ under Assumptions 1a, 3 and 4.

For the confidence interval $I_n\left(\tilde{\rho}_n, \alpha\right)$, we verify Assumptions A1 and S of Andrews, Cheng and Guggenberger (2020), abreviated to ACG (2020). Given the parameter space Θ in (7), and $\theta = \left(\rho, F, X_0\right) \in \Theta$, the coverage probability of $I_n\left(\tilde{\rho}_n, \alpha\right)$ is $CP_n\left(\theta\right) = \mathbb{P}_\theta\left(\left|T_n\left(\tilde{\rho}_n\right)\right| \leq \Phi^{-1}\left(1-\alpha/2\right)\right)$ in the notation of ACG (2020). Consider a sequence $\left(\theta_n\right)_{n\in\mathbb{N}} = \left(\rho_n, F_n, X_0\left(n\right)\right)_{n\in\mathbb{N}} \subseteq \Theta$ and an arbitrary subsequence $\left(w_n\right)_{n\in\mathbb{N}} \subseteq \mathbb{N}$. Since $\left(\rho_{w_n}\right)_{n\in\mathbb{N}} \subseteq \left[-1+\delta, M\right]$ is bounded, there exists a subsequence $\left(k_n\right)_{n\in\mathbb{N}} \subseteq \left(w_n\right)_{n\in\mathbb{N}}$ such that $\rho_{k_n} \to \rho \in \left[-1+\delta, M\right]$, so $\left(\rho_{k_n}\right)_{n\in\mathbb{N}}$ satisfies Assumption 1a. Since $\left(F_{w_n}, X_0\left(w_n\right)\right)_{n\in\mathbb{N}} \subseteq \mathcal{A}$, $\left(F_{w_n}, X_0\left(w_n\right)\right)_{n\in\mathbb{N}} \subseteq \mathcal{A}_n$ for all but finitely many n; since $k_n \geq w_n$, $\left(F_{k_n}, X_0\left(k_n\right)\right) \in \mathcal{A}_{k_n}$ for all n. We conclude that there exists a subsequence $\left(k_n\right)_{n\in\mathbb{N}}$ of $\left(w_n\right)_{n\in\mathbb{N}}$ such that $\left(\theta_{k_n}\right)_{n\in\mathbb{N}} = \left(\rho_{k_n}, F_{k_n}, X_0\left(k_n\right)\right)_{n\in\mathbb{N}}$ satisfies Assumptions 1a, 2 and 3 which implies that $\lim_{n\to\infty}\mathbb{P}_{\theta_{k_n}}\left(T_{\theta_{k_n}}\left(\tilde{\rho}_n\right) \leq x\right) = \Phi\left(x\right)$ and

$$\lim_{n\to\infty} CP_{k_n}\left(\theta_{k_n}\right) = \lim_{n\to\infty}\mathbb{P}_{\theta_{k_n}}\left(\left|T_{\theta_{k_n}}\left(\tilde{\rho}_n\right)\right| \leq \Phi^{-1}\left(1-\alpha/2\right)\right) = 1-\alpha. \tag{A.26}$$

Convergence in (A.26) proves simultaneously the validity of Assumptions A1 and S of ACG(2020) and the claim $\lim_{n\to\infty}\inf_{\theta\in\Theta}\mathbb{P}\left[\rho \in I_n\left(\tilde{\rho}_n, \alpha\right)\right] = 1-\alpha$ follows from Theorem 2.1(e) of ACG(2020).

It remains to prove that $T_n\left(\beta_n^*\right) \to_d \mathcal{N}\left(0,1\right)$. We first show that $T_n\left(\tilde{\beta}_n\right) - T_n\left(\beta_n^*\right) \to_p 0$ under Assumption 1b. Since $T_n\left(\beta_n^*\right) = T_n\left(\tilde{\beta}_n\right)$ on the event $\bar{F}_n$, it is enough to show that $T_n\left(\tilde{\beta}_{1n}\right) - T_n\left(\beta_{1n}^*\right) \to_p 0$ 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 A1(iii), the denominator of $T_n\left(\beta_{1n}^*\right)$ in (26), $v_{1n}^2 := \sum_{t=1}^n \tilde{z}_{t-1}^2 - n\bar{z}_{1,n-1}^2\left(1-\hat{\rho}_{\varepsilon u}^2\right)\mathbf{1}_{F_n}$, satisfies $\pi_n^{-2}\left(\sum_{t=1}^n \tilde{z}_{1t-1}^2 - v_{1n}^2\right) \to_p 0$. Hence,

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

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

Supplementary Online Appendix B

This online Appendix contains: (i) a collection of auxiliary results (Lemma B1) and its proof as well as the proofs of Lemma A1 and Corollary 1 of the main paper in Section 1.1 and (ii) some additional simulation results in Section 2.2 below.

1.1 Additional mathematical results

Lemma B1 is concerned with the limit distribution of the normalised and centred OLS estimator $\hat{\rho}_n$ in (15) obtained from the autoregression (2)/(14) under weakly dependent errors. The result is well-known for the near-nonstationary class C(ii), so we concentrate on the near-stationary/explosive classes C(i) and C(iii).

Lemma B1. *Consider the autoregressions x_t in (2)/(14) and x_{0t} in (14) and the stochastic sequences X_n in (39) 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 and 3, the following hold:

(i) *Under C(i) and Assumption 4, $\bar{x}_{n-1} = [1 + o_p(1)](\mu + \bar{x}_{0n-1})$,*

$$n^{-1/2}\left(1-\rho_n^2\right)^{1/2}\left|\sum_{t=1}^n \underline{x}_{t-1}\underline{u}_t - \sum_{t=1}^n x_{0t-1}u_t\right| = o_p(1)$$
$$n^{-1}\left(1-\rho_n^2\right)\left|\sum_{t=1}^n \underline{x}_{t-1}^2 - \sum_{t=1}^n x_{0t-1}^2\right| = o_p(1)$$

and $\left(1-\rho_n^2\right)^{-1}(\hat{\rho}_n - \rho_n) \to_p \Gamma/\left(\sigma^2 + 2\rho\Gamma\right)$.

(ii) *Under C(i) and Assumption 2, $\left(n\left(1-\rho_n^2\right)^{-1}\right)^{1/2}(\hat{\rho}_n - \rho_n) \to_d \mathcal{N}(0,1)$.*

(iii) *Under C(iii), $\left(\rho_n^2-1\right)^{1/2}(\rho_n - 1)\rho_n^{-n} n\bar{x}_n = X_n + o_p(1)$,*

$$\left(\rho_n^2-1\right)^2 \rho_n^{-2n}\sum_{t=1}^n x_{t-1}^2 = X_n^2 + o_p(1),$$

$\left(\rho_n^2-1\right)^{-1}\rho_n^n(\hat{\rho}_n - \rho_n) = \Upsilon_n/X_n + o_p(1)$ 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.

Proof. Denote $x_0 = X_0(n)$ for brevity. By employing (3), we obtain

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

Under C(i), $\bar{x}_{0n} = O_p\left(n^{-1/2}(1-\rho_n)^{-1}\right)$ and, by Assumption 3, the order of the last term of (B.1) is given by $O_p\left(n^{-1}(1-\rho_n)^{-3/2}\right)$. Since

$$\frac{n^{-1}(1-\rho_n)^{-3/2}}{n^{-1/2}(1-\rho_n)^{-1}} = (1-\rho_n)^{-1/2} n^{-1/2} \to 0$$

the last term of (B.1) is dominated by $\bar{x}_{0n}$. In turn, $\bar{x}_{0n} = o_p(1)$ when $n^{1/2}(1-\rho_n) \to \infty$ (the half of the C(i) region closer to stationarity), in which case μ is the dominant term in (B.1). This proves the claim $\bar{x}_n = [1 + o_p(1)](\mu + \bar{x}_{0n})$ of part (i). For the second claim of part (i), (3) and $\bar{x}_{0n} = O_p\left(n^{-1/2}\kappa_n\right)$ and $\bar{u}_n = O_p\left(n^{-1/2}\right)$ give

$$\begin{aligned}
\sum_{t=1}^n \underline{x}_{t-1}\underline{u}_t - \sum_{t=1}^n x_{0t-1}u_t &= \mu\sum_{t=1}^n u_t + (x_0-\mu)\sum_{t=1}^n \rho_n^{t-1}u_t - [1+o_p(1)]\,n(\mu + \bar{x}_{0n-1})\bar{u}_n \\
&= (x_0-\mu)\sum_{t=1}^n \rho_n^{t-1}u_t - n\bar{x}_{0n-1}\bar{u}_n + o_p(n)(\mu\bar{u}_n + \bar{x}_{0n-1}\bar{u}_n) \\
&= o_p(\kappa_n) + O_p(\kappa_n) + o_p\left(n^{1/2}\right)
\end{aligned}$$

showing that $n^{-1/2}\left(1-\rho_n^2\right)^{1/2}\left|\sum_{t=1}^n \underline{x}_{t-1}\underline{u}_t - \sum_{t=1}^n x_{0t-1}u_t\right| = o_p(1)$. For the third claim of part (i), (3) gives

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

We know that $\sum_{t=1}^{n} x_{0t-1}^2 = O_p\left(n\left(1-\rho_n^2\right)^{-1}\right)$ under C(i). By Assumption 3, the last three terms in (B.2) are of order $o_p\left(\kappa_n^2\right)$ dominated by that of $\sum_{t=1}^{n} x_{0t-1}^2$ (for the first of these terms the inequality $\left\|\sum_{t=1}^{n} x_{0t-1}\rho_n^t\right\|_{L_1} \leq \max_{t\leq n}\left\|x_{0t}\right\|_{L_2}\sum_{j=1}^{n}\left|\rho_n\right|^j$ is used). Since $\sum_{t=1}^{n} x_{0t-1} = O_p\left(n^{1/2}\kappa_n\right) = o_p\left(n\kappa_n\right)$, the second term in (B.2) dominated by $\sum_{t=1}^{n} x_{0t-1}^2$; hence

$$n^{-1}\left(1-\rho_n^2\right)\sum_{t=1}^{n} x_{t-1}^2 = n^{-1}\left(1-\rho_n^2\right)\sum_{t=1}^{n} x_{0t-1}^2 + \left(1-\rho_n^2\right)\mu^2.$$

Using the above and the first claim, we conclude that

$$\begin{aligned} n^{-1}\left(1-\rho_n^2\right)\sum_{t=1}^{n} \underline{x}_{t-1}^2 &= n^{-1}\left(1-\rho_n^2\right)\left(\sum_{t=1}^{n} x_{t-1}^2 - n\bar{x}_{n-1}^2\right) \\ &= n^{-1}\left(1-\rho_n^2\right)\left(\sum_{t=1}^{n} x_{0t-1}^2 + n\mu^2 - \left[1+o_p(1)\right]n\left(\mu+\bar{x}_{0n-1}\right)^2\right) \\ &= n^{-1}\left(1-\rho_n^2\right)\left[\sum_{t=1}^{n} x_{0t-1}^2 + o_p(n) + O_p\left(n^{1/2}\kappa_n\right) + O_p\left(\kappa_n^2\right)\right] \\ &= n^{-1}\left(1-\rho_n^2\right)\sum_{t=1}^{n} x_{0t-1}^2 + o_p(1), \end{aligned}$$

completing the proof of the third claim.

For the OLS estimator, $n^{-1}\sum_{t=1}^{n} x_{0t-1}u_t \to_p \Gamma$ under C(i) and Assumption 4 by Lemma 2.2(i) of Magdalinos and Phillips (2020). Using the recursion for x_{0t}, we obtain the identity

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

Hence, using the previous claims we may write

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

For part (ii), using the approximations of part (i) we may write

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

and the last term converges in distribution to $\mathcal{N}(0,1)$ under Assumption 2 by Giraitis and Phillips (2006). For part (iii), (B.1) and $\sum_{t=1}^{n} x_{0t-1} = \left(\rho_n - 1\right)^{-1}\left(x_{0n} + \sum_{t=1}^{n} u_t\right)$ give

$$n\bar{x}_{n-1} = n\mu + n\bar{x}_{0n-1} + \left(x_0 - \mu\right)\frac{\rho_n^n - 1}{\rho_n - 1} = \frac{x_{0n}}{\rho_n - 1} + \left(x_0 - \mu\right)\frac{\rho_n^n}{\rho_n - 1} + n\mu + o_p\left(\left(\rho_n - 1\right)^{-3/2}\right)$$

so $\left(\rho_n^2 - 1\right)^{1/2}\left(\rho_n - 1\right)\rho_n^{-n}n\bar{x}_{n-1} = \left(\rho_n^2 - 1\right)^{1/2}\left\{\rho_n^{-n}x_{0n} + \left(x_0 - \mu\right)\right\} + O_p\left(n\left(\rho_n - 1\right)\rho_n^{-n}\right)$. Since $n\left(\rho_n - 1\right)\rho_n^{-n} \to 0$ under C(iii) by Lemma A1(i) and $\rho_n^{-n}x_{0n} = \sum_{j=1}^{n}\rho_n^{-j}u_j$, the claim follows from the above display and the definition of X_n. For the second claim, the second and fifth terms on the right of (B.2) are at most $O_p\left(\left(\rho_n - 1\right)^{-3/2}\rho_n^n\right) = o_p\left(\left(\rho_n^2 - 1\right)^{-2}\rho_n^{2n}\right)$; for the fourth term on the right of (B.2), since $x_{0,0} = 0$:

$$\begin{aligned} \sum_{t=1}^{n}\rho_n^{t-1}x_{0t-1} &= \sum_{t=1}^{n-1}\rho_n^t\sum_{j=1}^{t}\rho_n^{t-j}u_j = \sum_{j=1}^{n-1}\left(\sum_{t=j}^{n-1}\rho_n^{2t}\right)\rho_n^{-j}u_j \\ &= \frac{1}{\rho_n^2 - 1}\rho_n^{2n}\sum_{j=1}^{n-1}\rho_n^{-j}u_j - \frac{1}{\rho_n^2 - 1}\sum_{j=1}^{n-1}\rho_n^j u_j \\ &= \frac{1}{\rho_n^2 - 1}\rho_n^{2n}\sum_{j=1}^{n-1}\rho_n^{-j}u_j + O_p\left(\left(\rho_n - 1\right)^{-3/2}\rho_n^n\right). \end{aligned}$$

Using the approximation $\left(\rho_n^2 - 1\right)^2\rho_n^{-2n}\left|\sum_{t=1}^{n} x_{0t-1}^2 - x_{0n}^2\right| = o_p(1)$, established in Phillips and Magdalinos (2007), and substituting in (B.2), we conclude that

$$\begin{aligned} \left(\rho_n^2 - 1\right)^2\rho_n^{-2n}\sum_{t=1}^{n} x_{t-1}^2 &= \left(\rho_n^2 - 1\right)\left(\left(\rho_n^{-n}x_{0n}\right)^2 + \left(x_0 - \mu\right)^2 + 2\left(x_0 - \mu\right)\sum_{j=1}^{n}\rho_n^{-j}u_j\right) + o_p(1) \\ &= \left(\rho_n^2 - 1\right)\left(\sum_{j=1}^{n}\rho_n^{-j}u_j + x_0 - \mu\right)^2 + o_p(1) \end{aligned}$$

since $\rho_n^{-n}x_{0n} = \sum_{j=1}^{n}\rho_n^{-j}u_j$ and the claim follows from the definition of X_n. For the final claim, for the denominator of the OLS estimator, (3) gives

$$\left(\rho_n^2-1\right)^2\rho_n^{-2n}n\bar{x}_{n-1}^2=\left(\rho_n^2-1\right)^2\rho_n^{-2n}nO_p\left(\frac{1}{n^2}\left(\rho_n-1\right)^{-3}\rho_n^{-2n}\right)=O_p\left(\frac{1}{n}\left(\rho_n-1\right)^{-1}\right)=o_p\left(1\right)$$

showing that $\left(\rho_n^2-1\right)^2\rho_n^{-2n}\left|\sum_{t=1}^n\underline{x}_{t-1}^2-\sum_{t=1}^n x_{t-1}^2\right|=o_p\left(1\right)$. For the numerator of the OLS estimator, (3) and the approximation

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

established in Phillips and Magdalinos (2007), give

$$\begin{aligned}\left(\rho_n^2-1\right)\rho_n^{-n}\sum_{t=1}^n x_{t-1}u_t &= \left(\rho_n^2-1\right)\left(\rho_n^{-n}\sum_{t=1}^n x_{0t-1}u_t+\left(x_0-\mu\right)\sum_{t=1}^n\rho_n^{-(n-t+1)}u_t\right)+o_p\left(1\right)\\ &= \left(\rho_n^2-1\right)\left(\sum_{t=1}^n\rho_n^{-(n-t+1)}u_t\right)\left(\sum_{j=1}^n\rho_n^{-j}u_j+x_0-\mu\right)+o_p\left(1\right)\\ &= \Upsilon_nX_n+o_p\left(1\right).\end{aligned}$$

Also $\left(\rho_n^2-1\right)\rho_n^{-n}\left|\sum_{t=1}^n\underline{x}_{t-1}\underline{u}_t-\sum_{t=1}^n x_{t-1}u_t\right|=\left(\rho_n^2-1\right)\rho_n^{-n}n\left|\bar{x}_{n-1}\bar{u}_n\right|=O_p(n^{-1/2}\left(\rho_n^2-1\right)^{-1/2})=o_p\left(1\right)$. Using the above approximations, we may write

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

as required. When $\rho_n\to1$, Magdalinos (2012) shows that $\left[X_n,\Upsilon_n\right]\to_d\mathcal{N}\left(0,\sigma^2I_2\right)$ implying that $\Upsilon_n/X_n\to_d\mathcal{C}$; when $\rho_n\to\rho>1$, Lemma 4(iii) shows that $X_n\to_p X_\infty\neq0$ *a.s.*, and $\mathbb{E}\Upsilon_n^2\to\sigma^2$, so in both cases $\left|\Upsilon_n/X_n\right|=O_p\left(1\right)$ and $\left(\rho_n^2-1\right)^{-1}\rho_n^n\left(\hat{\rho}_n-\rho_n\right)=O_p\left(1\right)$ over the C(iii) range. This completes the proof of Lemma B1.

As a consequence of Lemma B1 and Phillips (1987b), the following orders of magnitude apply under C(ii)-C(iii):

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

Proof of Lemma A1. For part (i), write $\varphi_{1n}^n=e^{n\log[1-(1-\varphi_{1n}))]}=e^{-n(1-\varphi_{1n})(1+o(1))}$ since $\log\left(1-x\right)=-x+O\left(x^2\right)$ as $x\to0$; hence $\left[n\left(1-\varphi_{1n}\right)\right]^p\varphi_{1n}^n=\left[n\left(1-\varphi_{1n}\right)\right]^pO\left(e^{-n(1-\varphi_{1n})}\right)\to0$ for any $p\geq0$ since $n\left(1-\varphi_{1n}\right)\to\infty$ under C(i). Under C(iii), $n\left(\varphi_{2n}-1\right)\to\infty$ and $\varphi_{2n}^{-n}=e^{-n\log[1+(\varphi_{2n}-1)]}=O\left(e^{-n(\varphi_{2n}-1)}\right)$ shows that $\left[n\left(\varphi_{2n}-1\right)\right]^p\varphi_{2n}^{-n}\to0$ for any $p\geq0$. 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 $s=\left(1-\varphi_{1n}\right)t$

$$\begin{aligned}\sum_{t=1}^n t^p\varphi_{1n}^t &= \int_1^{n+1}\lfloor t\rfloor^p\varphi_{1n}^{\lfloor t\rfloor}dt\\ &= \left(1-\varphi_{1n}\right)^{-1-p}\int_{1-\varphi_{1n}}^{(n+1)(1-\varphi_{1n})}\left(\frac{\left\lfloor\left(1-\varphi_{1n}\right)^{-1}s\right\rfloor}{\left(1-\varphi_{1n}\right)^{-1}}\right)^p\varphi_{1n}^{\left\lfloor(1-\varphi_{1n})^{-1}s\right\rfloor}ds. \quad \text{(B.4)}\end{aligned}$$

Since $1-\varphi_{1n}\to0$, $n\left(1-\varphi_{1n}\right)\to\infty$ and

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

the dominated convergence theorem implies that the integral on the right side of (B.4) converges to $\int_0^\infty s^pe^{-s}ds=\Gamma\left(p+1\right)$, and the claim for $\sum_{t=1}^n t^p\varphi_{1n}^t$ follows from (B.4). 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 (3) to the instrument $\tilde{z}_{1t}=\sum_{j=1}^t\varphi_{1n}^{t-j}\Delta x_j$ in (19), we obtain the following decomposition:

$$\tilde{z}_{1t}=\tilde{z}_{0t}+\left(X_0\left(n\right)-\mu\right)q_{nt},\quad \tilde{z}_{0t}=\sum_{j=1}^t\varphi_{1n}^{t-j}\Delta x_{0j}, \quad \text{(B.5)}$$

where $q_{nt} = \frac{1-\rho_n}{\varphi_{1n}-\rho_n}\left(\rho_n^t - \varphi_{1n}^t\right)$ when $n\left|\varphi_{1n}-\rho_n\right| \to \infty$ and $q_{nt} = t\left(1-\varphi_{1n}\right)\varphi_{1n}^t\left[1+O\left(n^{-1}\right)\right]$ when $\left|\varphi_{1n}-\rho_n\right| = O\left(n^{-1}\right)$. We show that

$$[\epsilon_{1n}, \epsilon_{2n}] = n^{-1/2}\left[\left(1-\rho_n^2\varphi_{1n}^2\right)^{1/2}\kappa_n^{1/2}, \left(1-\rho_n^2\varphi_{1n}^2\right)o\left(\kappa_n\right)\right]\left(\sum_{t=1}^n q_{nt}^2\right)^{1/2} \to 0. \tag{B.6}$$

When $n\left|\varphi_{1n}-\rho_n\right| \to \infty$, $\epsilon_{1n} = \left(\kappa_n\left(\varphi_{1n}-\rho_n\right)\right)^{-1/2}\left(n^{-1}\sum_{t=1}^n\left(\rho_n^{2t}-\varphi_{1n}^{2t}\right)\right)^{1/2} \to 0$ and $\epsilon_{2n} = o\left(1\right)\left(n^{-1}\sum_{t=1}^n\left(\rho_n^{2t}-\varphi_{1n}^{2t}\right)\right)^{1/2} \to 0$. When $\left|\varphi_{1n}-\rho_n\right| = O\left(n^{-1}\right)$, $\sum_{t=1}^n t^2\varphi_{1n}^{2t} = O\left(\left(1-\varphi_{1n}\right)^{-3}\right)$ by part (i) and $\kappa_n^{-1} = O\left(1-\varphi_{1n}\right)$ imply that both ϵ_{1n} and ϵ_{1n} are $O\left(n^{-1/2}\left(1-\varphi_{1n}\right)^{-1/2}\right)$.

Now (B.5) and Assumption 3 give $r_{1n} = n^{-1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{1/2}o_p\left(\kappa_n^{1/2}\right)\sum_{t=1}^n q_{nt}u_t$. Since

$$\begin{aligned}\left\|\sum_{t=1}^n q_{nt}u_t\right\|_{L_2}^2 &\leq 2\sum_{s=1}^n\sum_{t=s}^n\left|q_{nt}\right|\left|q_{ns}\right|\left|\gamma_u\left(t-s\right)\right| = 2\sum_{t=0}^n\left|\gamma_u\left(t\right)\right|\sum_{s=1}^{n-t}\left|q_{nt+s}\right|\left|q_{ns}\right| \\ &\leq 2\sum_{t=0}^n\left|\gamma_u\left(t\right)\right|\left(\sum_{s=1}^{n-t}q_{ns}^2\right)^{1/2}\left(\sum_{s=1}^{n-t}q_{n,t+s}^2\right)^{1/2} \leq 2\sum_{s=1}^n q_{ns}^2\sum_{t=0}^\infty\left|\gamma_u\left(t\right)\right|\end{aligned}$$

and $\sum_{t=0}^\infty\left|\gamma_u\left(t\right)\right| < \infty$ by Assumption 4, $r_{1n} \to_p 0$ follows from the fact that $\epsilon_{1n} \to 0$ in (B.6). For r_{2n}, (B.5) 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

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

by (B.6) 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 Magdalinos and Phillips (2020). For $r_{3n} = r_{3n}' + r_{3n}''$, with

$$\left[r_{3n}', r_{3n}''\right] = n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)\left[\sum_{t=1}^n\left(\tilde{z}_{1t}-\tilde{z}_{0t}\right)x_t, \sum_{t=1}^n\left(x_t - x_{0t}\right)\tilde{z}_{0t}\right],$$

the Cauchy-Schwarz inequality and (B.5) imply that $r_{3n}' \leq O_p\left(1\right)\epsilon_{2n}\left(n^{-1}\kappa_n^{-1}\sum_{t=1}^n x_t^2\right)^{1/2} = o_p\left(1\right)$ by (B.6) and $n^{-1}\kappa_n^{-1}\sum_{t=1}^n x_t^2 = O_p\left(1\right)$. For r_{3n}'', (3) and the $\sum_{t=1}^n\tilde{z}_{0t} = O_p\left(n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1}\right)$ imply that

$$r_{3n}'' = o_p\left(1\right)r_{3n}''' + O_p\left(n^{-1/2}\right), \quad r_{3n}''' = n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)\kappa_n^{1/2}\sum_{t=1}^n\tilde{z}_{0t}\rho_n^t. \tag{B.7}$$

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 $r_{3n}''' \leq O_p\left(1\right)\left(n^{-1}\left(1-\rho_n^2\varphi_{1n}^2\right)\kappa_n^{1/2}\sum_{t=1}^n\rho_n^{2t}\right)^{1/2} = O_p\left(1\right)\left(n^{-1}\left(1-\varphi_{1n}^2\right)^{-1/2}\right)^{1/2} = o_p\left(1\right)$. When $n\left|\varphi_{1n}-\rho_n\right| \to \infty$, the definition of $\tilde{z}_{0t}$, the summation by parts formula and the Cauchy-Schwarz inequality give

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

Since $\left(1-\rho_n\right)\left(\sum_{j=1}^n\rho_n^{2j}\right)^{1/2} = O\left(\kappa_n^{-1/2}\right)$, $\left|r_{3n}'''\right| \leq O\left(1\right)\left(n^{-2}\sum_{j=1}^n x_{0j}^2\right)^{1/2} + O_p\left(\rho_n^n\right) = O_p\left(1\right)$ so $r_{3n} = o_p\left(1\right)$ follows from (B.7).

For part (iii), (B.5) gives $\left(1-\rho_n\varphi_{1n}\right)\sum_{t=1}^n q_{nt} \sim \left(1-\varphi_{1n}\right)^2\sum_{t=1}^n t\varphi_{1n}^t = O\left(1\right)$ by part (i) when $\left|\varphi_{1n}-\rho_n\right| = O\left(n^{-1}\right)$ and $\left(1-\rho_n\varphi_{1n}\right)\sum_{t=1}^n q_{nt} = O\left(\kappa_n^{-1}\right)\sum_{t=1}^n\left(\rho_n^t - \varphi_{1n}^t\right) = O(1\vee\left(\kappa_n\left(1-\varphi_{1n}\right)\right)^{-1})$ when $n\left|\varphi_{1n}-\rho_n\right| \to \infty$. Substituting into (B.5) gives $\left(1-\rho_n\varphi_{1n}\right)\sum_{t=1}^n\left(\tilde{z}_{1t}-\tilde{z}_{0t}\right) = o_p(\kappa_n^{1/2}) + o_p(\kappa_n^{-1/2}\left(1-\varphi_{1n}\right)^{-1})$, and the order of magnitude of part (iii) follows from $\left(1-\rho_n\varphi_{1n}\right)\sum_{t=1}^n\tilde{z}_{0t} = O_p\left(n^{1/2}\right)$ by Magdalinos and Phillips (2020). Under C(ii), $\kappa_n = n$ so $n^{-1/2}\left(1-\varphi_{1n}\right)\sum_{t=1}^n\left(\tilde{z}_{1t}-\tilde{z}_{0t}\right) = o_p\left(1\right)$. The recursion $\tilde{z}_{0t} = \varphi_{1n}\tilde{z}_{0t-1} + \Delta x_{0t}$ implies that $\left(1-\varphi_{1n}\right)n^{-1/2}\sum_{t=1}^n\tilde{z}_{0t-1} = n^{-1/2}\left(x_{0n}-\tilde{z}_{0n}\right)$ 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 $\tilde{z}_{2t} = \sum_{j=1}^{t} \varphi_{2n}^{t-j}\hat{u}_j$ in (17), we obtain the decomposition

$$\tilde{z}_{2t} = z_{2t} - (\hat{\rho}_n - \rho_n)\,\psi_{nt-1} + \varphi_{2n}^{t} g_{n,t}, \quad \psi_{nt-1} = \sum_{j=1}^{t} \varphi_{2n}^{t-j} x_{j-1} \tag{B.8}$$

$g_{n,t} := \left[(\hat{\rho}_n - \rho_n)\,\bar{x}_{n-1} - \bar{u}_n\right]\left(1 - \varphi_{2n}^{-t}\right)/\left(\varphi_{2n} - 1\right)$ satisfies $\max_{1\le t\le n}|g_{n,t}| = O_p\left(n^{-1/2}\left(\varphi_{2n} - 1\right)^{-1}\right)$ since $(\hat{\rho}_n - \rho_n)\,\bar{x}_{n-1} = O_p\left(\kappa_n^{1/2}/n\right)$ by Lemma B1(iii) under C(iii) and standard local to unity asymptotics under C(ii). When $n\left|\varphi_{2n} - \rho_n\right| \to \infty$, (A.1) will follow from the following identity for ψ_{nt-1} in (B.8):

$$\psi_{nt-1} = \frac{1}{\varphi_{2n} - \rho_n}\left(\varphi_{2n} z_{2t-1} - \rho_n x_{t-1} + \varphi_{2n}^{t} g'_{n,t}\right) \tag{B.9}$$

where $g'_{n,t} = X_0(n) - \mu\left(1 - \varphi_{2n}^{-(t-1)}\right)\frac{\rho_n - 1}{\varphi_{2n} - 1}$, with the order in (A.1) following from $\left(\varphi_{2n} - \rho_n\right)^{-1}(\hat{\rho}_n - \rho_n)\max_{1\le t\le n}\left|g'_{n,t}\right| = O_p\left(\kappa_n^{-1}\rho_n^{-n}\left(\varphi_{2n} - 1\right)^{-1}\right) = O_p\left(n^{-1}\left(\varphi_{2n} - 1\right)^{-1}\right)$ under C(ii)-C(iii). To prove (B.9), substituting x_t in (3) into the expression for ψ_{nt-1} in (B.8) we obtain

$$\begin{aligned}\psi_{nt-1} &= \varphi_{2n}^{t-1} X_0(n) + \mu\frac{\varphi_{2n}^{t-1} - 1}{\varphi_{2n} - 1} + \varphi_{2n}^{t}\sum_{i=1}^{t-1}\rho_n^{-i-1}u_i \sum_{j=i+1}^{t}\left(\frac{\rho_n}{\varphi_{2n}}\right)^{j} \\ &\quad + \varphi_{2n}^{t}\rho_n^{-1}\left(X_0(n) - \mu\right)\sum_{j=2}^{t}\left(\frac{\rho_n}{\varphi_{2n}}\right)^{j}.\end{aligned} \tag{B.10}$$

Evaluating the geometric progression

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

when $n\left|\varphi_{2n} - \rho_n\right| \to \infty$, we obtain

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

and using the expression for x_t in (3) and $z_{2t} = \sum_{i=1}^{t}\varphi_{2n}^{t-i}u_i$ proves (B.9). This completes the proof of (A.1) when $n\left|\varphi_{2n} - \rho_n\right| \to \infty$.

When $\left|\varphi_{2n} - \rho_n\right| = O\left(n^{-1}\right)$, $\rho_n/\varphi_{2n} = 1 + \left(\rho_n - \varphi_{2n}\right)/\varphi_{2n} = 1 + O\left(n^{-1}\right)$ so $\sum_{j=i+1}^{t}\left(\frac{\rho_n}{\varphi_{2n}}\right)^{j} \le nb$ for all $i < t \le n$ and some $b > 0$. Substituting into (B.10) we conclude that

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

by Assumption 3 and Lemma A1(i) since $\left|\varphi_{2n} - \rho_n\right| = O\left(n^{-1}\right)$ and the choice of φ_{2n} imply that $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(iii). This completes the proof of (A.1) when $\left|\varphi_{2n} - \rho_n\right| = O\left(n^{-1}\right)$.

For the remainder of part (iv), we employ (A.1) to each of $R_{1n}, \ldots, R_{4n}$. Using (19),

$$\left(\varphi_{2n} - 1\right)v_{n,z}^{-1}\sum_{t=1}^{n} z_{2t-1} = v_{n,z}^{-1}z_{2n} - v_{n,z}^{-1}\sum_{t=1}^{n}u_t = Z_n + o_p(1) \tag{B.11}$$

since $n^{1/2}v_{n,z}^{-1} = n^{1/2}\left(\varphi_{2n} - 1\right)^{1/2}\varphi_{2n}^{-n} \to 0$ by Lemma A1(i); also (3) gives

$$\sum_{t=1}^{n} x_t = n\mu + \left(X_0(n) - \mu\right)O\left(\rho_n^{n}\kappa_n\right) + \sum_{t=1}^{n} x_{0t} = O_p\left(\kappa_n\nu_n\right) \tag{B.12}$$

since $\sum_{t=1}^{n} x_{0t} = O_p\left(\kappa_n\nu_n\right)$ by Lemma B1(iii). Using (A.1) and the above orders for $\sum_{t=1}^{n} z_{2t-1}$ and $\sum_{t=1}^{n} x_t$ we obtain

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

by Lemma A1(i).

For $R_{2n}=\sum_{t=1}^{n}r_{nt-1}u_t$, the second term arising from (A.1): $(\varphi_{2n}^2-1)\,\varphi_{2n}^{-n}g_n\sum_{t=1}^{n}\varphi_{2n}^{t-1}u_t = O_p\left(n^{-1/2}\left(\varphi_{2n}^2-1\right)^{-1/2}\varphi_{2n}^{-n}\right)=o_p(1)$ by Lemma A1(i); when $n\left|\varphi_{2n}-\rho_n\right|\to\infty$, (A.1) and the triangle inequality, give

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

by (B.3). The first term above is $O_p\left(\varphi_{2n}^{-n}\right)$ and the second term is $O_p\left(\frac{\rho_n^{-n}}{\kappa_n\left|\varphi_{2n}-\rho_n\right|}\right)$ which is $O_p\left(\rho_n^{-n}\right)$ under C(iii) and $O_p\left(\frac{1}{n\left|\varphi_{2n}-1\right|}\right)$; since both terms are $o_p(1)$ we conclude that $\left|R_{2n}\right|=o_p(1)$ under C(ii)-C(iii).

For $R_{3n}=s_n^{-1}\sum_{t=1}^{n}r_{nt}x_t$, we estimate the two terms arising from (A.1): by (3)

$$
s_n^{-1}g_n\sum_{t=1}^{n}\varphi_{2n}^{t}x_t=s_n^{-1}g_n\left(\sum_{t=1}^{n}\varphi_{2n}^{t}x_{0t}+\mu\sum_{t=1}^{n}\varphi_{2n}^{t}+(X_0-\mu)\sum_{t=1}^{n}(\varphi_{2n}\rho_n)^t\right). \tag{B.13}
$$

Using the rate of g_n in (A.1), the third term on the right is $o_p\left(n^{-1/2}\left(\varphi_{2n}^2-1\right)^{-1/2}\right)$. The second term is $o_p\left(n^{-1/2}\right)$ under C(ii); under C(iii); the third term is $o_p\left(\rho_n^{-n}\kappa_n^{-1/2}\right)$ if $(\rho_n-1)/(\varphi_{2n}-1)=O(1)$ and $O_p\left(n^{-2}\left(\varphi_{2n}-1\right)^{-3/2}\left(n\left(\rho_n-1\right)\right)^{3/2}\rho_n^{-n}\right)=o_p\left(n^{-1/2}\right)$ if $(\rho_n-1)/(\varphi_{2n}-1)\to\infty$. We conclude that the second and the third terms of (B.13) are $o_p(1)$. For the first term of (B.13),

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

which shows that the left side of (B.13) is $o_p(1)$. When $n\left|\varphi_{2n}-\rho_n\right|\to\infty$, (A.1)

$$
\begin{aligned}
\left|R_{3n}\right| &\leq s_n^{-1}\frac{\left|\hat{\rho}_n-\rho_n\right|}{\left|\varphi_{2n}-\rho_n\right|}\left(\left|\sum_{t=1}^{n}z_{2t-1}x_t\right|+\left|\sum_{t=1}^{n}x_{t-1}x_t\right|\right)\\
&\leq bs_n^{-1}\frac{\left|\hat{\rho}_n-\rho_n\right|}{\left|\varphi_{2n}-\rho_n\right|}\left(\left|\sum_{t=1}^{n}z_{2t-1}x_{t-1}\right|+\sum_{t=1}^{n}x_{t-1}^2\right)
\end{aligned}\tag{B.14}
$$

for all but finitely many n for some $b>0$, because (14) gives $\left|\sum_{t=1}^{n}z_{2t-1}x_t\right|\leq\left|\rho_n\right|\left|\sum_{t=1}^{n}z_{2t-1}x_{t-1}\right|+\left|\mu\left(\rho_n-1\right)\right|\left|\sum_{t=1}^{n}z_{2t-1}\right|+\left|\sum_{t=1}^{n}z_{2t-1}u_t\right|$, 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.14) is $O_p\left(\frac{\kappa_n^{-1}\rho_n^{-n}}{\left|\varphi_{2n}-\rho_n\right|}\right)=O_p\left(\frac{1}{n\left|\varphi_{2n}-\rho_n\right|}\right)$ under C(ii)-C(iii) (under C(iii) it is $o_p\left(\frac{1}{n\left|\varphi_{2n}-\rho_n\right|}\right)$ by Lemma A1(i)). The second term on the right of (B.14) is $O_p\left(\kappa_n^{1/2}\nu_{n,z}^{-1}\right)=o_p\left(\left(n\left(\varphi_{2n}-1\right)\right)^{1/2}\varphi_{2n}^{-n}\right)=o(1)$ by the orders in (B.3) and Lemma A1(i). This proves that $R_{3n}=o_p(1)$.

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

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

because the Cauchy-Schwarz inequality gives

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

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

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

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

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

The last term is $O(n^{-1}\left(\varphi_{2n}^2-1\right)^{-1})=o(1)$; using (B.3), the second term is $O_p(\left(\frac{\varphi_{2n}^2-1}{\varphi_{2n}-\rho_n}\right)^2\varphi_{2n}^{-2n})=O_p((n(\varphi_{2n}-\rho_n))^{-2}(n(\varphi_{2n}-1))^2\varphi_{2n}^{-2n})=o_p((n(\varphi_{2n}-\rho_n))^{-2})=o_p(1)$ by LemmaA1(i). By Lemma 4(i) and (B.3), the first term is $O_p(\kappa_n^{-2}\rho_n^{-2n}(\varphi_{2n}-\rho_n)^{-2})$ which is $O_p((n(\varphi_{2n}-\rho_n))^{-2})$ under C(ii) and $O_p((n/\kappa_n)^2\rho_n^{-2n}n^{-2}(\varphi_{2n}-\rho_n)^{-2})=o_p(n^{-2}(\varphi_{2n}-\rho_n)^{-2})$ under C(iii) by Lemma A1(i). The above shows that $R'_{4n}=o_p(1)$ and $R_{4n}=o_p(1)$. This completes the proof of part (iv).

For part (v), we begin by showing that

$$\epsilon_{1n}:=(\varphi_{2n}-1)\sum_{j=1}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-t}c_{t+j}\right)^2,\ \epsilon_{2n}:=(\varphi_{2n}-1)\sum_{j=1}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-(n-t+1)}c_{t+j}\right)^2 \qquad \text{(B.15)}$$

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

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

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 we obtain

$$\begin{aligned}Z_n &= (\varphi_{2n}-1)^{1/2}\sum_{j=1}^n\varphi_{2n}^{-j}\left(\sum_{t=0}^{n-j}\varphi_{2n}^{-t}c_t\right)e_j+(\varphi_{2n}-1)^{1/2}\sum_{j=0}^\infty\left(\sum_{t=1}^n\varphi_{2n}^{-t}c_{t+j}\right)e_{-j}\\ &= \left(\sum_{t=0}^n\varphi_{2n}^{-t}c_t\right)(\varphi_{2n}-1)^{1/2}\sum_{j=1}^n\varphi_{2n}^{-j}e_j-Z_{1n}+Z_{2n} \qquad \text{(B.16)}\end{aligned}$$

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 $\mathbb{E}\left(Z_{2n}^2\right)\le\mathbb{E}\left(e_1^2\right)\epsilon_{1n}\to 0$ by (B.15) and

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

by Lemma A1(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.16). 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 $\mathbb{E}\left(Y_{2n}^2\right)\le\mathbb{E}\left(e_1^2\right)\epsilon_{2n}\to 0$ by (B.15). 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

$$p_n=\left(\varphi_{2n}^2-1\right)\sum_{j=1}^n\varphi_{2n}^{-2j}\left(\sum_{t=0}^{j-1}\left(\varphi_{2n}^t-1\right)c_t\right)^2\to 0. \qquad \text{(B.17)}$$

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

$$\varphi_{2n}^{t}-1\leq t\varphi_{2n}^{t}\log\varphi_{2n} \tag{B.18}$$

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

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

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.17) and completes the proof of part (iv).

Proof of Corollary 1. For the last two t-statistics in (29) $T_n(\tilde{\gamma}_n), T_n\left(\tilde{\delta}_n\right)\rightarrow_d\mathcal{N}(0,1)$ follow directly from Theorem 1 by putting $v_t=u_{2t}/\sigma_\gamma$ and $v_t=u_{3t}/\sigma_\delta$ in (A.24). For $T_n(\tilde{r}_n)$, write

$$\begin{aligned}
\tilde{r}_n-r_0 &= \frac{\tilde{\rho}_n-\rho_n}{\tilde{\gamma}_n+\tilde{\delta}_n}-\frac{\rho_n-1}{\left(\tilde{\gamma}_n+\tilde{\delta}_n\right)(\gamma+\delta)}\left(\tilde{\gamma}_n-\gamma+\tilde{\delta}_n-\delta\right)\\
&= \left[\frac{1}{\tilde{\gamma}_n+\tilde{\delta}_n},-\frac{\rho_n-1}{\left(\tilde{\gamma}_n+\tilde{\delta}_n\right)^2},-\frac{\rho_n-1}{\left(\tilde{\gamma}_n+\tilde{\delta}_n\right)^2}\right]\begin{bmatrix}\tilde{\rho}_n-\rho_n\\ \tilde{\gamma}_n-\gamma\\ \tilde{\delta}_n-\delta\end{bmatrix}\\
&= \hat{\upsilon}_n'\frac{\sum_{t=1}^{n}\tilde{z}_{t-1}u_t}{\sum_{t=1}^{n}I_{t-1}\tilde{z}_{t-1}}
\end{aligned}$$

where

$$\hat{\upsilon}_n=\left[\frac{1}{\hat{\gamma}_n+\hat{\delta}_n},-\frac{\hat{\rho}_n-1}{\left(\hat{\gamma}_n+\hat{\delta}_n\right)^2},-\frac{\hat{\rho}_n-1}{\left(\hat{\gamma}_n+\hat{\delta}_n\right)^2}\right]',$$

and $u_t=(u_{1t},u_{2t},u_{3t})'$ in the notation of Assumption 5. Under Assumption 5, $\hat{\Sigma}_n\rightarrow_p\Sigma>0$, $\hat{\upsilon}_n\rightarrow_p\upsilon=\left[\frac{1}{\gamma+\delta},-\frac{\rho-1}{(\gamma+\delta)^2},-\frac{\rho-1}{(\gamma+\delta)^2}\right]'$ and $\hat{\sigma}_{r_0}^2\rightarrow_p\upsilon'\Sigma\upsilon$; hence $T_n(\tilde{r}_n)=[1+o_p(1)]T_n$ with T_n given by (A.24) 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 $T_n\rightarrow_d\mathcal{N}(0,1)$ 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.24) 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 $T_n\rightarrow_d\mathcal{N}(0,1)$ by Theorem 1.

1.2 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 instument selection of Section 4.1 of the main paper. 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 mildly explosive instrument is chosen. Figure B5 is a comparison of the length of confidence intervals 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 confidence intervals of the IV 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.

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%

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%

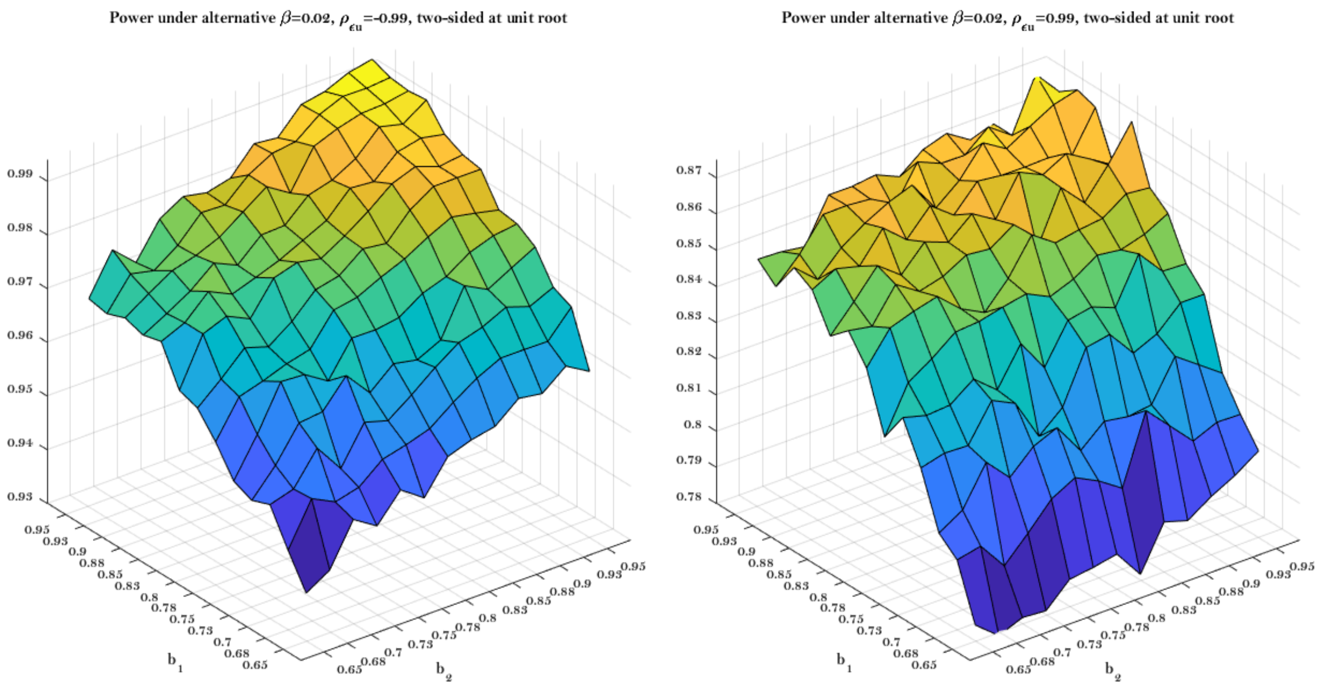


Figure B1: Power at $\rho = 1$ over a grid for b_1 and b_2

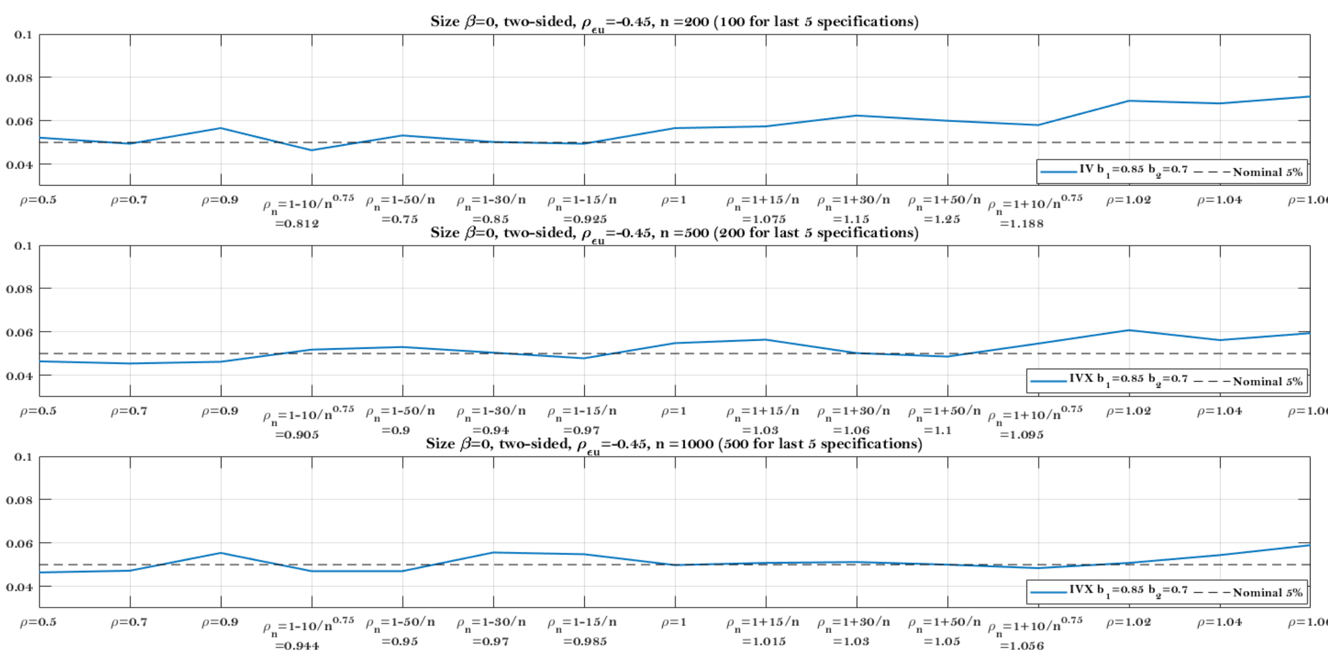


Figure B2: Empirical size of the two-sided test on β, $\rho_{\varepsilon u} = -0.45$

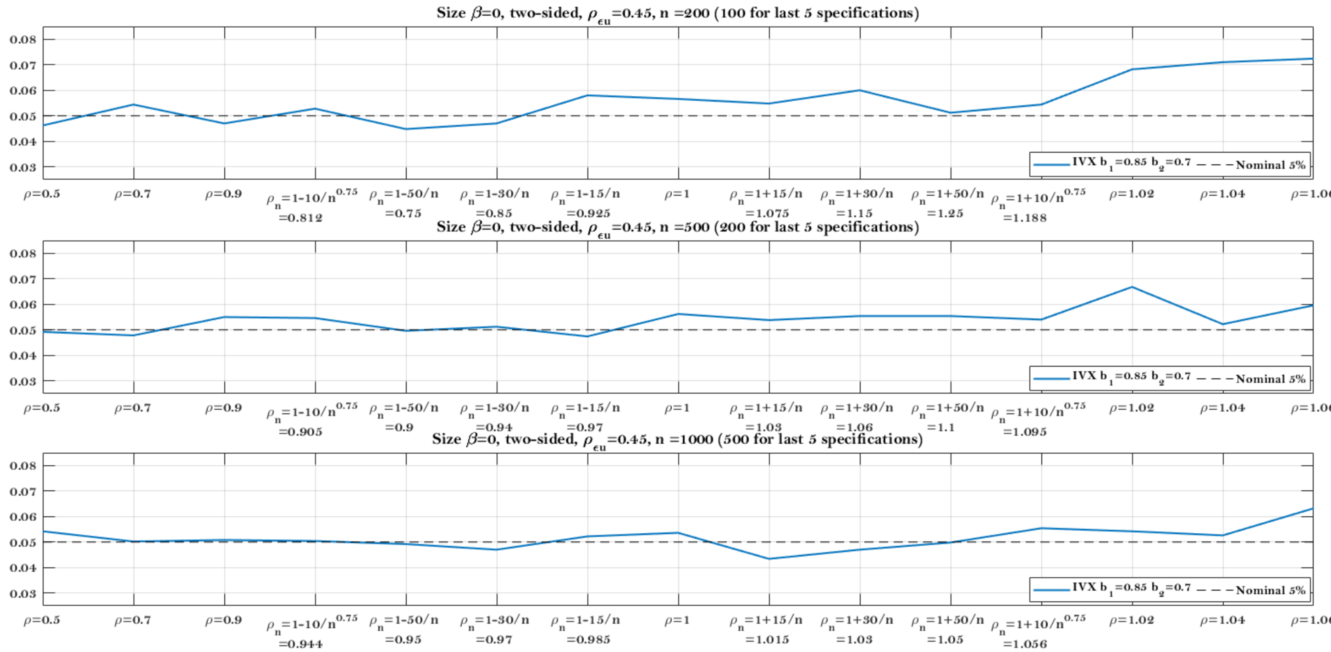


Figure B3: Empirical size of the two-sided test on β, $\rho_{\varepsilon u} = 0.45$

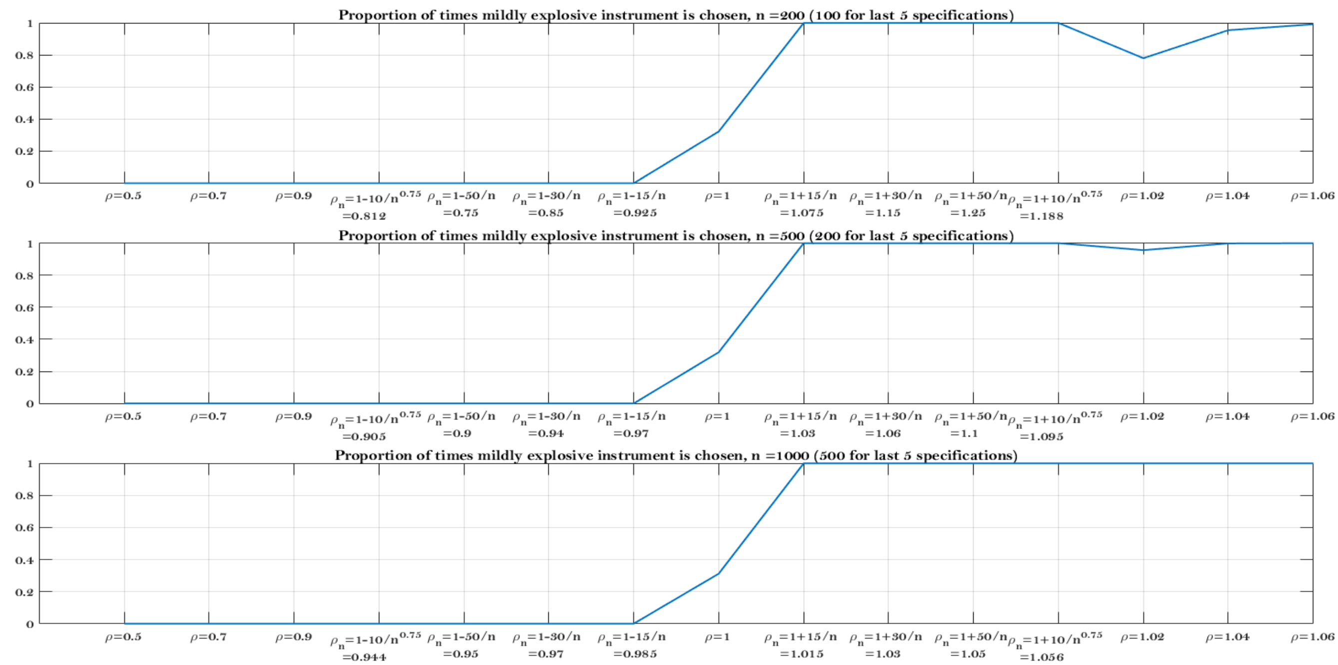


Figure B4: Proportion of times $\tilde{z}_{2t}$ is chosen

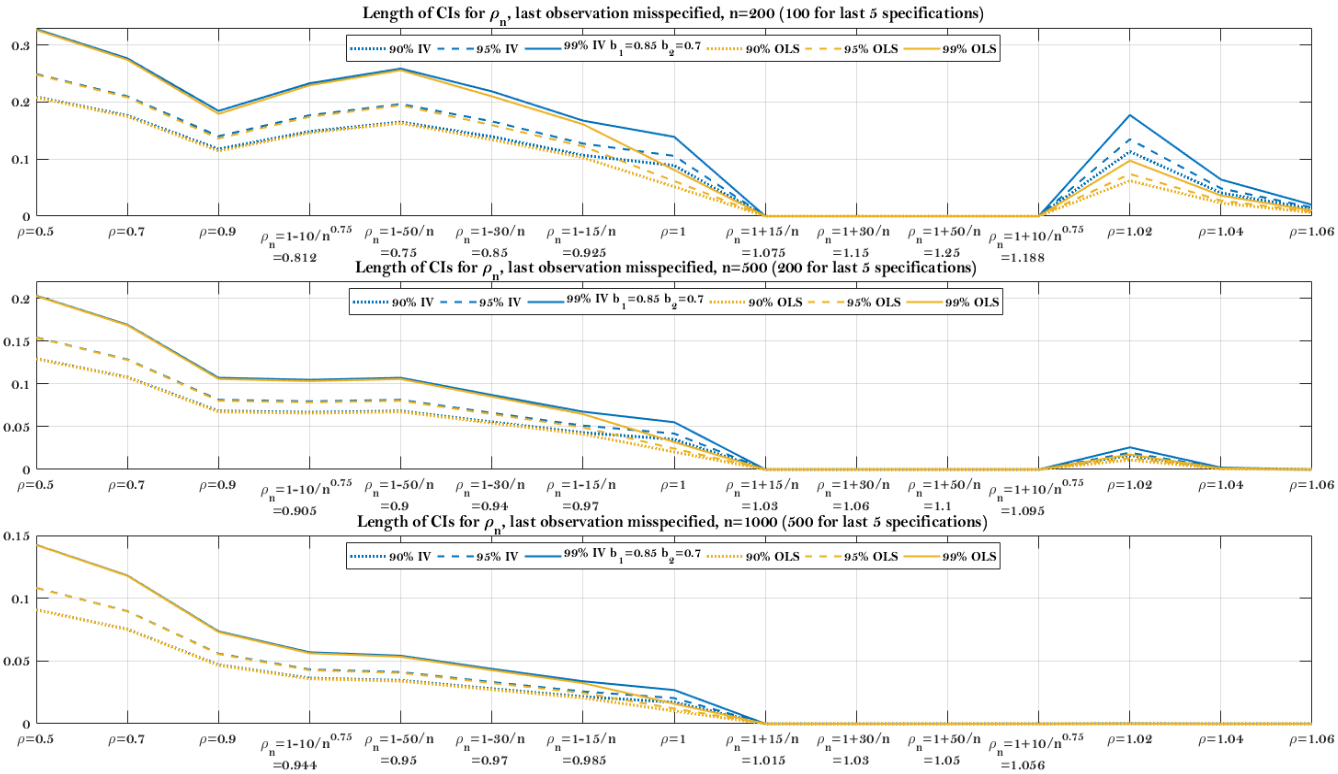


Figure B5: Length of intervals of IV and OLS under misspecification of the last observation

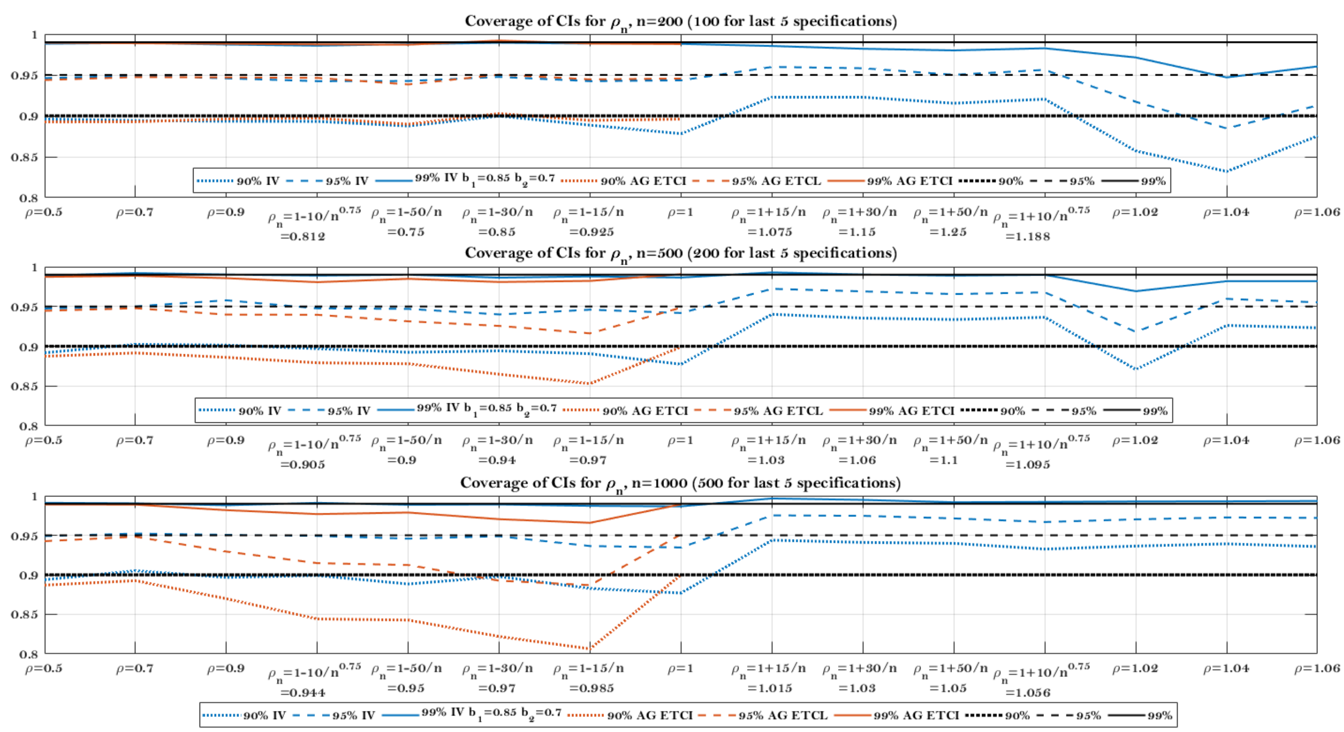


Figure B6: Coverage of confidence intervals of IV and ETCI of Andrews and Guggenberger (2014).

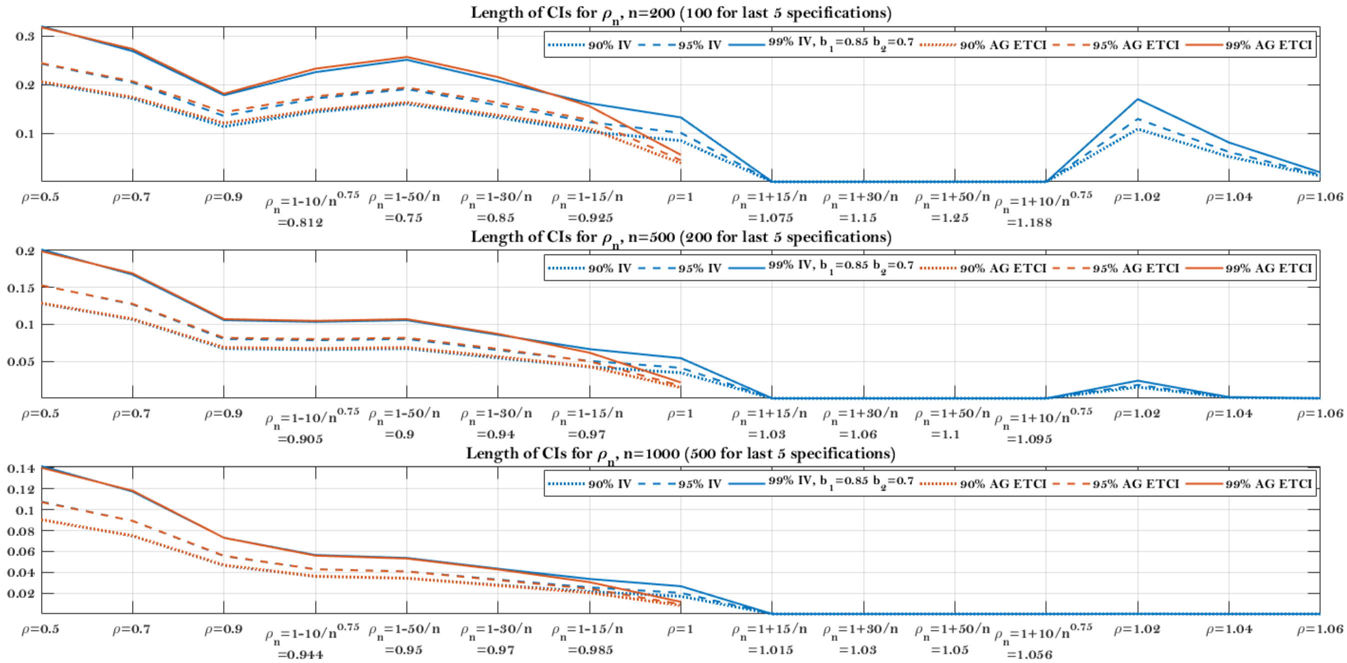


Figure B7: Length of confidence intervals of IV and ETCI of Andrews and Guggenberger (2014).

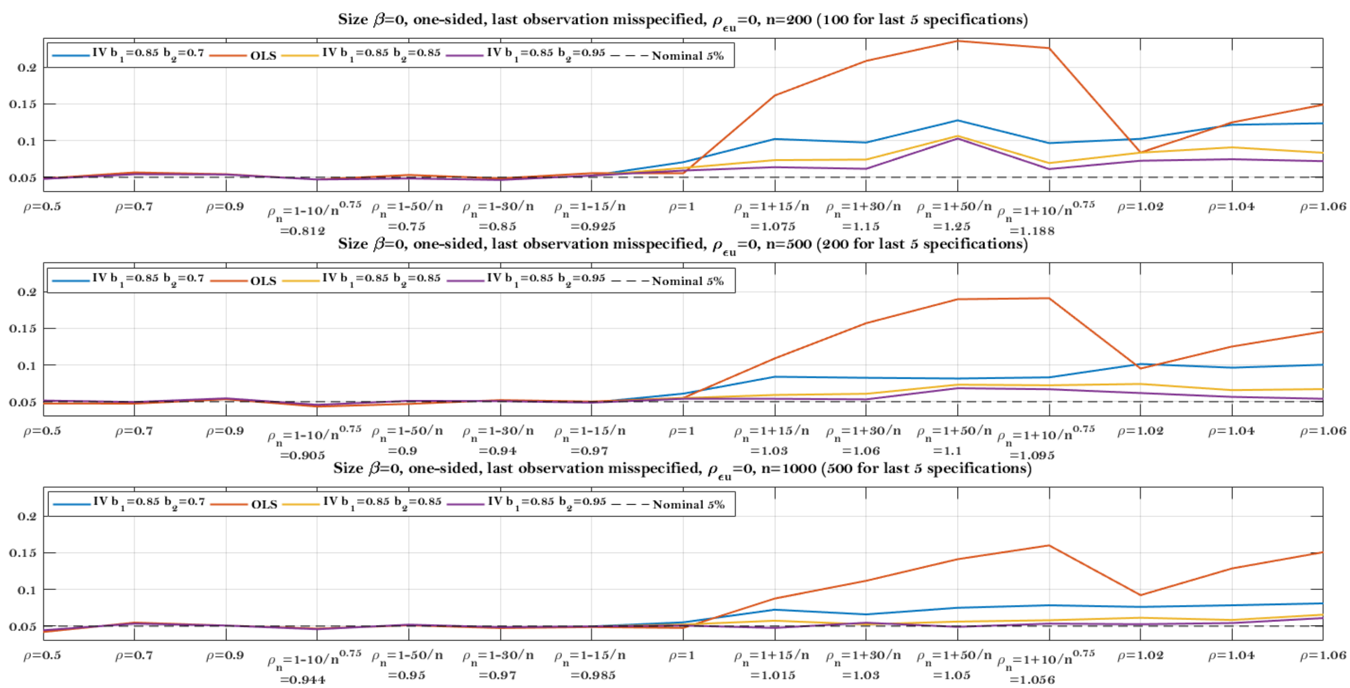


Figure B8: Size of OLS- and IV-based one-sided test under misspecification of the last observation