STAFF REPORTS

NO. 1151
APRIL 2025

# Uniform Inference with General Autoregressive Processes

Tassos Magdalinos | Katerina Petrova

FEDERAL RESERVE BANK of NEW YORK

**Uniform Inference with General Autoregressive Processes**

Tassos Magdalinos and Katerina Petrova

*Federal Reserve Bank of New York Staff Reports*, no. 1151
April 2025
https://doi.org/10.59576/sr.1151

**Abstract**

A unified theory of estimation and inference is developed for an autoregressive process with root in $(-\infty, \infty)$ that includes the stationary, local-to-unity, explosive and all intermediate regions. The discontinuity of the limit distribution of the t-statistic outside the stationary region and its dependence on the distribution of the innovations in the explosive regions $(-\infty, -1) \cup (1, \infty)$ are addressed simultaneously. A novel estimation procedure, based on a data-driven combination of a near-stationary and a mildly explosive artificially constructed instrument, delivers mixed-Gaussian limit theory and gives rise to an asymptotically standard normal t-statistic across all autoregressive regions. The resulting hypothesis tests and confidence intervals are shown to have correct asymptotic size (uniformly over the space of autoregressive parameters and the space of innovation distribution functions) in autoregressive, predictive regression and local projection models, thereby establishing a general and unified framework for inference with autoregressive processes. Extensive Monte Carlo simulation shows that the proposed methodology exhibits very good finite sample properties over the entire autoregressive parameter space $(-\infty, \infty)$ and compares favorably to existing methods within their parametric $(-1, 1]$ validity range. We demonstrate how our procedure can be used to construct valid confidence intervals in standard epidemiological models as well as to test in real-time for speculative bubbles in the price of the Magnificent Seven tech stocks.

JEL classification: C12, C22
Key words: uniform inference, central limit theory, autoregression, predictive regression, instrumentation, mixed-Gaussianity, t-statistic, confidence intervals

---

Petrova: Federal Reserve Bank of New York (email: katerina.petrova@ny.frb.org). Magdalinos: University of Southampton (email: a.magdalinos@soton.ac.uk). The authors thank Peter Phillips, Richard Crump, Kirill Evdokimov, Adam McCloskey, Bo Zhou, Gabriele La Spada, and Nina Boyarchenko for valuable comments and discussions. They are also grateful to Hannah O'Keefe for invaluable research assistance.

# 1 Introduction

Imposing short memory assumptions in macroeconomic and financial models is convenient since it delivers standard econometric inference on the models' parameters with conventional asymptotic distributions. However, such stationarity assumptions are often empirically unrealistic and a variety of stochastic trends has been found in many macroeconomic and financial time series.

Whenever a nonstationary regressor is included in a regression model, the additional signal from the strong time dependence present in the regressor, while facilitating more precise point estimates, usually makes the series non-ergodic, thus invalidating standard central limit theory (CLT). Consequently, OLS-based test statistics have different distributional limits depending on the persistence degree of the regressor and the critical values required by practitioners to construct confidence intervals (CIs) or run hypothesis tests on the model's parameters are vastly different. As a result, misspecifying the type of regressor invalidates inference resulting in size distortions which do not improve with the sample size and lead to erroneous empirical conclusions.

In this paper, we focus on a single regressor generated by the prototypical time series model, a first-order autoregressive (AR) process with root in $(-\infty, \infty)$. We consider inference in several regression models where this regressor enters on the right-hand side: a pure autoregression, a predictive regression and a local projection model. Even in these simple setups, different persistence degrees of the regressor create discontinuities in the OLS limit distributions with limits involving stochastic integrals in the vicinity of (negative and positive) unit root as well as limits of unknown distributional form which depend on the distributions of the innovations in explosive regions. The discontinuities in the limits of OLS-based test statistics, well-documented in a series of classical papers (e.g. Mann and Wald (1943) for autoregressions with stationary roots, Anderson (1959) for explosive roots, Phillips (1987a) for unit-roots and Chan and Wei (1987) for local-to-unity roots) present a major challenge for inference. Even robust standard procedures such as bootstrap have been shown to be invalid in presence of unit roots (Basawa et al. (1991)). Given the inference challenges arising from nonstationarity, a strand of the literature is dedicated to designing screening procedures for researchers to detect possible nonstationarities in the data (e.g. unit root and cointegration testing) and account for or remove them, e.g. by differencing or detrending. Pre-testing not only leads to size pile-up but, as Cavanagh et al. (1995) show, two-step procedures with highly persistent regressors are invalid unless the latter have an exact unit root. Such fundamental issues with pre-testing have led to the design of new inference procedures in order to achieve valid inference in these nonstandard setups. Early work on obtaining CIs for an AR coefficient in $(-1, 1]$, thereby accommodating stationary and unit-root processes, includes Stock (1991), Andrews (1993), Hansen (1999), Romano and Wolf (2001), Mikusheva

(2007) and Andrews and Guggenberger (2009, 2014). The literature on inference in predictive regressions was developed in parallel proposing different solutions to the inference problem with a nonstationary regressor. Notable contributions include Campbell and Yogo (1996), Jansson and Moreira (2006), Elliott, Müller and Watson (2015) and Cavaliere and Georgiev (2020). One common feature of most existing inference procedures is that their proposed solution is designed for a model-specific nonstandard environment by modifying existing test statistics and designing new and often complicated procedures which perform well in the presence of the particular type of persistence considered and are often invalid and work poorly outside the region considered.

In this paper, we propose a new inference procedure based on instrumental variable (IV) estimators which makes use of usual IV-based standard errors and standard normal ($\mathcal{N}(0,1)$) critical values. The new methodology approaches the classical inference problem from a new angle: rather than providing new test statistics designed to work in the various nonstandard setups, the novel approach constructs a new process from the data, designed to automatically adjust the regressor's persistence to a level where standard (mixed) Gaussian central limit theory applies. We establish a unified asymptotic inference framework with regressor covering the entire autoregressive spectrum (including stationary, local-to-unity, explosive processes and all intermediate and oscillating regions) and simultaneously providing a solution for autoregressive, predictive regressive and local projection models. Our novel procedure thus places all these potentially nonstandard processes and models under a common econometric inference framework which delivers $\mathcal{N}(0,1)$ t-statistic regardless of the regressor's stochastic properties, with uniform validity over the AR parameter space and the space of distribution functions of the innovations.

The key idea of the approach is to filter the regressor through a process that acts as a proxy, artificially constructed from the data as a combination of moderately stationary and mildly explosive process and is employed as an instrumental variable. In the case of a stationary regressor, this instrument reduces to the process itself, so that our IV estimators are asymptotically equivalent to OLS, thereby inheriting all optimality properties of OLS in the stationary region. Outside the stationary region, the instrument is designed to automatically adjust the persistence of the original regressor to a level where CLT applies. Using the same instrument, we can provide a solution to the inference problem in all three regression models considered. In particular, we show that the resulting IV estimators are asymptotically mixed-Gaussian regardless of the true unknown stochastic integration order of the regressor and, consequently, self-normalised test statistics based on these IV estimators deliver standard asymptotic inference. Moreover, we show that the resulting hypothesis tests and confidence intervals have correct asymptotic size uniformly over the space of AR parameters, the distribution functions of the innovations and the initial condition.

Developing the limit theory in full generality and establishing the theoretical properties of the new IV estimators and the uniformity of the resulting critical regions is dense and technical and is deferred to Section 3; the inference procedure that we propose, however, is based on simple closed-form linear estimators and standard test statistics and requires nothing else than $\mathcal{N}(0,1)$-critical values, rendering the procedure extremely easy and straightforward to implement.

While a univariate AR(1) setup for the regressor is simple and stylised, the inference problem arising from the discontinuities of the OLS limit distribution is fundamentally similar in models with more complex dynamics and larger dimensions. The central idea of controlling the regressor's persistence through filtering rather than modifying existing test procedures is general and can be applied to regressors with more complex dynamics. Moreover, the procedure's linearity and tractability make it easily scalable to multivariate setups, which is in contrast to existing approaches. In addition to providing the first uniform and distribution-free procedure for models with a general AR regressor, the current paper, therefore, also lays the basic foundation for the application of this novel methodology to processes with richer dynamics and larger dimensions.

The remainder of the paper is organised as follows: Section 2.1 presents the regression models considered, Section 2.2 introduces the novel IV procedure and informally discusses its properties. Sections 2.3 and 2.4 provide a discussion of the advantages of the proposed methodology relative to existing approaches in the literature and its empirical relevance. Development of the asymptotic theory appears in Section 3. Theorem 1 establishes the asymptotic mixed-Gaussianity property of the IV estimators and the $\mathcal{N}(0,1)$ limit distribution of the associated t-statistics. The main results of the paper on uniform asymptotic inference in the autoregressive, predictive regressive and local projection models appear in Theorems 2, 3 and Corollary 1 respectively. Section 4 discusses practical implementation of the procedure and conducts Monte Carlo experiments to assess the finite sample properties of our CIs and hypothesis tests in comparison to the leading existing inference procedures. Section 5.1 applies the results of Theorems 2 and 3 to construct CIs for the parameters of an SIR model and Section 5.2 demonstrates how our procedure can be used to test in real-time for speculative bubbles in asset prices of popular tech stocks and Section 6 concludes. Some auxiliary results as well as all mathematical proofs are provided in the online Appendix, which also contains additional simulation comparisons.

# 2 The IV Methodology

## 2.1 The Regression Models

Let $x_t$ be a first order AR process with an intercept of the form:

$$x_t - \mu = \rho\left(x_{t-1} - \mu\right) + u_t, \quad t \in \{1, ..., n\}, \quad \rho \in \mathbb{R}. \tag{1}$$

We consider three regression models where the process for $x_t$ in (1) enters as a regressor. The first model is the pure AR model in (1) in which the key parameter of interest is $\rho \in \mathbb{R}$. The second model is a predictive regression (PR) of the form:

$$y_t = \gamma + \beta x_{t-1} + \varepsilon_t, \tag{2}$$

where $x_t$ is generated by (1). While the object of interest in this model is $\beta$ and $\rho$ is a nuisance parameter, the asymptotics and hence, the validity of inference procedures on $\beta$ is driven by the degree of persistence of $x_t$. Finally, the third model we consider is a local projection (LP) model of the form

$$x_t = r_h x_{t-h} + v_{h,t} \tag{3}$$

with $x_t$ generated by (1) and the object of interest is the impulse response at horizon $h$ given by $r_h = \rho^h$.

A complete set of assumptions for the models (1), (2) and (3) is provided in Section 3 (Assumptions 1-4). Here, we simply mention that the autoregressive parameter $\rho$ may take any real value (Assumption 1), $(u_t)$ and $(\varepsilon_t)$ are (possibly conditionally heteroskedastic) martingale difference sequences and that $(u_t)$ may exhibit linear short memory autocorrelation when $x_{t-1}$ is a predictive regressor in (2) (Assumptions 2 and 4). We employ a drifting sequence of parameters approach to establishing uniform asymptotic size properties of confidence intervals and critical regions. To this end, we consider sample-size dependent sequences of AR roots $(\rho_n)_{n\in\mathbb{N}}$ and innovations $(u_{n,t})_{t\in\mathbb{N}}$, $(\varepsilon_{n,t})_{t\in\mathbb{N}}$ (Assumptions 5 and 6). Autoregressive processes with roots satisfying $\rho_n \to \rho \in \mathbb{R}$, may be categorised into three broad classes according to their stochastic properties.

**Definition 1 (AR classification).** *Let $\rho_n \to \rho \in \mathbb{R}$ and $c := \lim_{n\to\infty} n\left(|\rho_n| - 1\right)$ exist in $[-\infty, \infty]$. The AR process $x_t$ in (1) belongs to one of the following classes:*

C(i) *near-stationary processes if* $c = -\infty$,

C(ii) *near-unit-root processes if* $c \in \mathbb{R}$,

C(iii) *near-explosive processes if* $c = \infty$.

Each class C$(\cdot)$ above may be further partitioned into a *regular* subclass $\mathrm{C}_+(\cdot)$ when $\rho \geq 0$ and an *oscillating* subclass $\mathrm{C}_-(\cdot)$ when $\rho < 0$. We further denote by $\mathrm{C}_1(\cdot)$ and $\mathrm{C}_{-1}(\cdot)$ the subclasses of C$(\cdot)$ where $\rho = 1$ and $\rho = -1$ respectively.

Unit root and local-to-unity processes are included in class C(ii). Purely stationary processes satisfying $\rho_n \to \rho \in (-1, 1)$ are included in C(i) and purely explosive processes $\rho_n \to \rho \in (-\infty, -1) \cup (1, \infty)$ are included in C(iii). In all three regression models (1), (2), (3) considered, the limit distributions of the OLS estimators for $\rho$, $\beta$ and $r_h$ are vastly different depending on the true unknown persistence level of the regressor $x_t$. Consequently, OLS-based inference requires different critical values for each autoregressive category of Definition 1: (i) the OLS-based t-statistic is standard normal only if $x_t$ belongs to C(i); (ii) if $x_t$ has an exact (positive or negative) unit root, i.e. $x_t$ belongs to C(ii) with $c = 0$, the t-statistic has a limit distribution that takes the form of a Dickey-Fuller ratio; (iii) if $x_t$ is local but not exact unit root ($c \neq 0$ in C(ii)), the limiting distribution of the t-statistic features a ratio of stochastic integrals which depend on $c$; and (iv) if $x_t$ is an explosive process with $|\rho| > 1$ then the distributional limit of the OLS t-statistic (when it exists) is of unknown form entirely driven by the distribution of the innovations[1], rendering OLS-based inference infeasible.

The processes in $\mathrm{C}_1$(i) and $\mathrm{C}_1$(iii) (usually referred to as moderately stationary and mildly explosive respectively) and their oscillating counterparts $\mathrm{C}_{-1}$(i) and $\mathrm{C}_{-1}$(iii) provide a crucial building block in the construction of our instrumentation procedure. Processes in the above classes satisfy a CLT[2], in stark contrast to local to unity processes in C(ii) and to purely explosive processes. Our instrument process is constructed to belong approximately to one of the classes $\mathrm{C}_1$(i), $\mathrm{C}_{-1}$(i), $\mathrm{C}_1$(iii), $\mathrm{C}_{-1}$(iii), thereby inheriting their desirable asymptotic properties.

## 2.2 The IV Approach

The idea behind our novel data-generated IV estimation procedure is to filter the regressor $x_t$ in (1) through a time series that acts as an instrument constructed to behave asymptotically as: a moderately stationary process belonging to $\mathrm{C}_1$(i) ($\mathrm{C}_{-1}$(i)) when $x_t$ belongs to $\mathrm{C}_+$(i) ($\mathrm{C}_-$(i)); a mildly explosive process belonging to $\mathrm{C}_1$(iii) ($\mathrm{C}_{-1}$(iii)) when $x_t$ belongs to $\mathrm{C}_+$(iii) ($\mathrm{C}_-$(iii)); a random linear combination of the two when $x_t$ belongs to the near-unit-root class $\mathrm{C}_+$(ii) ($\mathrm{C}_-$(ii)). The choice of instrument process is motivated by two considerations: (i) it is built to mimic closely the actual process $x_t$, rendering the instrument relevant, (ii) the instrument's persistence is artificially controlled so that the instrument inherits the desirable asymptotic properties of moderately stationary/mildly explosive processes $\mathrm{C}_1$(i) and $\mathrm{C}_1$(iii) and their oscillating counterparts.

We denote the OLS estimator for $\rho$ and the resulting residuals by

$$\hat{\rho}_n = \left(\sum_{t=1}^n \underline{x}_{n,t-1}^2\right)^{-1} \sum_{t=1}^n \underline{x}_{n,t}\underline{x}_{n,t-1} \quad \text{and} \quad \hat{u}_{n,t} = \underline{x}_{n,t} - \hat{\rho}_n \underline{x}_{n,t-1}, \tag{4}$$

where $\underline{x}_{n,t} := x_t - n^{-1}\sum_{t=1}^n x_t$. Next, we define the event

$$F_n = \{n\left(|\hat{\rho}_n| - 1\right) \leq 0\}, \tag{5}$$

its complement $\bar{F}_n$, and the events $F_n^+ = F_n \cap \{\hat{\rho}_n \geq 0\}$, $F_n^- = F_n \cap \{\hat{\rho}_n < 0\}$, $\bar{F}_n^+ = \bar{F}_n \cap \{\hat{\rho}_n \geq 0\}$ and $\bar{F}_n^- = \bar{F}_n \cap \{\hat{\rho}_n < 0\}$. We employ the disjoint events $\left\{F_n^+, F_n^-, \bar{F}_n^+, \bar{F}_n^-\right\}$ as an automatic data-

[1] For purely explosive regressors with i.i.d. innovations, no CLT applies and sample moments converge as $L_2$-bounded martingales to random variables whose distribution depends on the distribution of the innovations. For non-identically distributed innovations, a limit of the OLS t-statistic may not exist; see Anderson (1959).

[2] Class $C_1(i)$ was introduced by Phillips and Magdalinos (2007) and generalised by Giraitis and Phillips (2006) and Andrews and Guggenberger (2012). While persistent, processes in $C_1(i)$ are ergodic and obey a CLT. Although processes in $C_1(iii)$ are not ergodic, sample moments satisfy CLT to a mixed-Gaussian distribution (established by Phillips and Magdalinos (2007) and extended in various directions by Aue and Horvath (2007), Magdalinos (2012) and Arvanitis and Magdalinos (2019)).

driven selection of the instrument process from one of the classes[3] $\left\{C_1(i),\, C_{-1}(i),\, C_1(iii),\, C_{-1}(iii)\right\}$. Letting $\nabla v_t := v_t + v_{t-1}$, we define quasi-innovations $\tilde{u}_{n,t}$ and an AR root $\rho_{nz}$ as:

$$\tilde{u}_{n,t} = \Delta x_{n,t}\mathbf{1}_{F_n^+} + \nabla \underline{x}_{n,t}\mathbf{1}_{F_n^-} + \hat{u}_{n,t}\mathbf{1}_{\bar{F}_n^+} + \hat{u}_{n,t}\mathbf{1}_{\bar{F}_n^-} \tag{6}$$

$$\rho_{nz} = \varphi_{1n}\mathbf{1}_{F_n^+} + \varphi_{1n}^-\mathbf{1}_{F_n^-} + \varphi_{2n}\mathbf{1}_{\bar{F}_n^+} + \varphi_{2n}^-\mathbf{1}_{\bar{F}_n^-} \tag{7}$$

where $(\varphi_{1n})_{n\in\mathbb{N}}$ and $\left(\varphi_{1n}^-\right)_{n\in\mathbb{N}}$ are *chosen* sequences in $C_1(i)$ and $C_{-1}(i)$ (so that $\varphi_{1n} \uparrow 1$ and $\varphi_{1n}^- \downarrow -1$ with the rate of C(i)); $(\varphi_{2n})_{n\in\mathbb{N}}$ and $\left(\varphi_{2n}^-\right)_{n\in\mathbb{N}}$ are *chosen* sequences in $C_1(iii)$ and $C_{-1}(iii)$ (so that $\varphi_{2n} \downarrow 1$ and $\varphi_{2n}^- \uparrow -1$ with the rate of C(iii)). The choice for $\rho_{nz}$ and the practical implementation of the procedure are discussed in detail in Section 4.1. Finally, we construct our instrument process by accumulating the stochastic sequence $\tilde{u}_{n,t}$ in (6) according to an AR(1) process initialised at $\tilde{z}_0 = 0$ with artificial root $\rho_{nz}$ set automatically by (7):

$$\tilde{z}_{n,t} = \rho_{nz}\tilde{z}_{n,t-1} + \tilde{u}_{n,t} = \sum_{j=1}^{t} \rho_{nz}^{t-j}\tilde{u}_{n,j}. \tag{8}$$

We employ the process for $\tilde{z}_{n,t}$ as an instrumental variable in the three regression models (1), (2) and (3). In particular, the resulting IV estimator for $\rho_n$ in (1) after instrumenting $x_{t-1}$ by $\tilde{z}_{n,t-1}$ takes the form of a standard IV estimator:

$$\tilde{\rho}_n = \left(\sum_{t=1}^{n} \underline{x}_{n,t-1}\tilde{z}_{n,t-1}\right)^{-1} \sum_{t=1}^{n} \underline{x}_{n,t}\tilde{z}_{n,t-1}. \tag{9}$$

For the PR model (2), instrumenting $x_{t-1}$ by $\tilde{z}_{n,t-1}$ gives rise to the estimator

$$\tilde{\beta}_n = \left(\sum_{t=1}^{n} \underline{x}_{n,t-1}\tilde{z}_{n,t-1}\right)^{-1} \sum_{t=1}^{n} \underline{y}_{n,t}\tilde{z}_{n,t-1}. \tag{10}$$

In the LP model, we instrument $x_{t-h}$ by $\tilde{z}_{n,t-h}$ and the resulting IV estimator for the impulse response at horizon $h$ is given by

$$\tilde{r}_{h,n} = \left(\sum_{t=1}^{n} x_{t-h}\tilde{z}_{n,t-h}\right)^{-1} \sum_{t=1}^{n} x_t\tilde{z}_{n,t-h}. \tag{11}$$

The underlying reason why such an instrument process $\tilde{z}_{n,t}$ works in all these models is three-fold: (i) it is designed to always be relevant since it is constructed to track the series $x_t$ through the accumulation of the quasi-innovations $\tilde{u}_{n,t}$ and the data-driven root $\rho_{nz}$ (e.g. when $x_t$ oscillates, the instrument is automatically designed to oscillate), hence the denominators of all three IV estimators converge in distribution to *a.s.* nonzero random variables; (ii) it inherits the desirable asymptotic properties of moderately stationary/mildly explosive processes, so the resulting numerators of the IV estimators always satisfy a martingale CLT; (iii) the possibly random limits of the denominators are independent from those of the numerators. These statements are formally established in Lemmata 3-5 of the Appendix. Consequently, the IV estimators in all three models are asymptotically mixed-Gaussian along all classes C(i)-C(iii) and independently of the distribution of the innovations, a result established in Theorem 1 of Section 3. While the data-driven combination of moderately stationary and mildly explosive instrument processes in (8) to unify

[3] Successful instrumentation requires asymptotic separation between processes for $x_t$ in C(i) and C(iii), this is formally established in Lemma 2 in the Appendix.

inference is intuitively appealing, the asymptotic validity of such an approach is not obvious and is justified in Lemma 2. The asymptotic mixed-Gaussianity (AMG) property of the estimators, established in Theorem 1 requires the asymptotic instrument selection result of Lemma 2 and the limit distribution theory of Lemmata 3-6 of the Appendix as building blocks.

Having the limit distributions of the IV estimators, we consider the usual IV-based t-statistics (with or without heteroskedasticity-consistent standardisation); the precise formulae for the test statistics of the AR, PR and LP models can be found in Section 3.2 in equations (16), (23) and (18) respectively. Crucially, the asymptotic mixed-Gaussianity of the IV estimators in Theorem 1 guarantees that these self-normalised t-statistics are asymptotically standard normal along all AR classes C(i)-C(iii). Moreover, we show that the proposed IV-based test statistics have correct size, uniformly over the parameter space of AR roots and sequences of innovation distribution functions, and give rise to CIs with uniformly correct asymptotic coverage. This is the fundamental result of the paper, and it is established in Theorem 2 for the AR model in (1), Theorem 3 for the PR model in (2) and Corollary 1 for the LP model in (3). The main practical implication of this result is that critical values from $\mathcal{N}(0,1)$ can be used to construct valid confidence intervals and hypothesis tests regardless of the true unknown stochastic nature of the regressor $x_t$ and without the need to pre-test or restrict the parameter space of admissible processes. To our knowledge, this is the first work to consider econometric inference for a general AR(1) process with unrestricted root, providing unified and distribution-free procedure for processes exhibiting arbitrary stochastic characteristics ranging from stationarity to (oscillating) explosivity and anything in-between.

Before the technical development of Section 3, we provide a brief discussion of how the approach compares to existing approaches and its empirical relevance in Sections 2.3 and 2.4 respectively.

## 2.3 Comparison with Existing Approaches

Early work on obtaining CIs for the AR model (1) with root $\rho \in (-1, 1]$, thereby accommodating the stationary region C(i) and the positive pure unit root part ($\rho = 1$, $c = 0$) of C(ii), includes Stock (1991), Andrews (1993), Hansen (1999) and Romano and Wolf (2001). Mikusheva (2007) develops the first general methodology for establishing uniform asymptotic validity of CIs in the AR model (1) with root in $(-1, 1]$ thereby accommodating C(i) and the entire left half of the positive local-to-unity ($c \leq 0,\ \rho = 1$) region C(ii) by providing a correction of Stock (1991)'s method. Subsequent work by Andrews and Guggenberger (2014) establishes methodology for CI construction with correct asymptotic size uniformly over the above regions under the potential presence of conditional heteroskedasticity of unknown form.

The literature on obtaining CIs and testing procedures that achieve valid inference for $\beta$ in the PR model (2) was developed in parallel. Early work that allows the regressor in (2) to fall in the positive local-to-unity region C(ii) include Campbell and Yogo (1996), Jansson and Moreira (2006) and Phillips and Magdalinos (2009). Procedures that can accommodate regressor's persistence in the stationary region C(i) as well as (parts of) the positive local-to-unity region C(ii) include Elliott et al. (2015), Kostakis et al. (2015), Magdalinos and Phillips (2020), Hu et al. (2021) as well as the bootstrap methods of Cavaliere and Georgiev (2020). None of these inference procedures for $\beta$ in the PR model (2) have been shown to have uniform validity over the considered regions of persistence degree $\rho$.

The literature on local projections and direct forecasting has also considered inference issues stemming from having a persistent regressor. For the LP model (3), Olea and Plagborg-Moller (2021) propose a lag-augmentation approach for inference on the impulse response parameter $r_h$ when the regressor $x_{t-h}$ may belong to C(i) and the positive part of C(ii). The lag-augmentation approach, in similar spirit to our IV approach, modifies the regressor (by adding an additional lag) rather than the test procedures; however, it achieves valid inference at a considerable power loss (the rate of convergence in C(ii) is $\sqrt{n}$ compared to a rate close to the $n$-OLS rate for our procedure; see (26) and Section 4.3.3 for Monte Carlo comparisons). Moreover, adding an additional lag is a device which cannot provide a solution for regressors in C(iii).

It is worth providing a brief comparison of our approach with the IVX procedure[4] of Phillips and Magdalinos (2009). Filtering in (8) is similar in spirit to the IVX procedure and one of the pieces of our instrument process $\tilde{z}_{n,t}$, assigned to the event $F_n^+ = \{n(|\hat{\rho}_n| - 1) \leq 0\} \cap \{\hat{\rho}_n \geq 0\}$ is precisely the IVX instrument of the aforementioned paper. However, the IVX instrument only achieves robust inference in C(i)-C(ii) for the case $\rho > 0$; it is invalid for oscillating processes in C(i) and C(ii) when $\rho = -1$. For such cases, our novel instrument in (8) is designed to use $\nabla \underline{x}_{n,t}$ instead of $\Delta x_{n,t}$ as 'residuals' for the instrument construction[5] and employs an oscillating root $\varphi_{1n}^- \to -1$ so that the instrument process emulates the oscillation of the original process $x_{n,t}$. Moreover, the IVX procedure of Phillips and Magdalinos (2009) is not suited for inference in near-explosive classes $\mathrm{C}_+$(iii) and $\mathrm{C}_-$(iii), and the two new mildly explosive pieces of our instrument process are designed for conducting inference in those regions. This is particularly relevant for one-sided unit root tests routinely used for detection of speculative bubbles in asset prices (see Section 5.2 for power comparisons of our novel IV procedure with the IVX). The mildly explosive part of the IV estimators in this paper differ from the IVX estimator in two important ways: first, the instrument construction is based on the OLS residuals $\hat{u}_{n,t}$ which (unlike $\Delta x_{n,t}$ and $\nabla \underline{x}_{n,t}$) approximate well the true innovation process in (1) in class C(iii); second, mildly explosive roots are employed in

[4]The IVX procedure has been extended in a number of directions by Breitung and Demetrescu (2015), Yang, Long, Peng and Cai (2020), Magdalinos and Phillips (2020), Demetrescu, Georgiev, Rodrigues and Taylor (2022).

[5]If $\rho = -1$, $\nabla \underline{x}_t = u_t + O_p\left(n^{-1/2}\right)$ behaves asymptotically as an innovation.

the instrument generation. Finally, the IVX approach has not been shown to achieve uniform inference; in fact, we are the first to show that with a modification of the parameterisation for $\rho_n$ to allow for arbitrary drifting sequences, the IVX-based inference can be made uniform (this is an immediate corollary to Theorem 3, see Remark 3 of Section 3.3). But the genuine novelty of our approach relative to IVX lies not so much in the construction of the three new instrument pieces, but in the data-driven combination of the novel mildly explosive instruments for regions C(ii) and C(iii) with the moderately stationary instruments for C(i) and C(ii) in order to achieve correct asymptotic inference for an AR root anywhere on the real line without *a priori* knowledge of which persistence region the true process belongs to.

One disadvantage that all approaches discussed above share is that they impose restrictions on the parameter space and hence the permitted degree of persistence: the negative local-to-unity region C(ii) (around $\rho = -1$) and the whole near-explosive region C(iii) are not considered; the right side of the positive local-to-unity ($\rho = 1$, $c > 0$) region C(ii) is also ruled out in most of the literature. A serious difficulty stems from the fact that OLS-based inference on the parameters in all three models considered with regressor in the purely explosive region only applies under i.i.d. Gaussian innovations[6]. In particular, the limit distributions of the OLS estimator and the associated t-statistic are not invariant to deviations from the assumptions of i.i.d. Gaussian errors and zero initial condition; in general, they are of unknown form driven by the distribution of the innovations in the autoregression (see Anderson (1959)).

The IV approach of this paper removes any restrictions imposed by the literature on the parameter space of the AR root in all three models and can accommodate the entire spectrum of AR roots $\rho \in \mathbb{R}$. In particular, we include: the negative part of local-to-unity region C(ii), the right side of the positive local-to-unity C(ii) and the entire near-explosive C(iii) region. To our knowledge, we provide the first procedure with general asymptotic validity in the purely explosive region[7]. These advances are achieved with a minimum loss of rate relative to OLS (for example in the purely explosive region our procedure preserves the exponential OLS rate of convergence).

Another disadvantage of existing procedures (with the exception of Mikusheva (2007) and Andrews and Guggenberger (2009, 2014)) is that they provide *pointwise* and not uniform asymptotic validity over their (restricted) parameter spaces. Our procedure has uniform asymptotic validity and we achieve this without any parameter space restriction. In addition, we also prove uniformity of our IV t-statistics over the space of a wide class of innovation distribution functions, to which

[6] Only under i.i.d. Gaussian innovations and zero initial condition, the normalised and centred OLS estimators for all three models considered have Cauchy limits and the corresponding t-statistics are standard normal.

[7] See Remark 5 in Section 3.3 for details on how large sample distributional invariance through CLT is achieved in $C_0$(iii) for our IV estimators.

OLS inference is not invariant in the purely explosive region.

A further drawback of existing approaches is that they impose restrictions not only on the parameter space but also on the regression model considered, with existing approaches on inference in the AR model (1) not applicable for the PR and LP models (2) and (3) and vice versa. One advantage of the IV approach proposed in this paper is that it can resolve the inference problem of a regressor with arbitrary persistence in all three regression models considered by employing the same instrument process, thereby removing the disconnect between different methods and literatures in these seemingly different model setups.

Finally, practical implementation for most existing approaches in the literature involves numerical approximations, rendering available approaches either computationally intensive or difficult for practitioners to understand and implement. It is therefore unsurprising that these robust approaches are rarely employed in the empirical time series literature. A major advantage of the proposed IV procedure is that it is linear and closed-form and it does not require programmes for numerical approximations or tables with nonstandard critical values. Beyond the simple construction of the instrument process in (8), it requires nothing other than standard IV estimation, t-statistics and $\mathcal{N}(0,1)$ critical values.

Extensive Monte Carlo experimentation reveals good finite sample properties for the proposed IV-based hypothesis tests and CIs that provide correct inference in $(-\infty,-1]\cup(1,\infty)$ where no existing alternative approach has general asymptotic validity, while also, in most cases, outperforming (in terms of power) existing leading procedures for inference in: the AR model (Andrews and Guggenberger (2014)), the PR model (Elliott et al. (2015)) and the LP model (Olea and Plagborg-Moller (2021)) in their parametric validity range $(-1,1]$.

## 2.4 Empirical Relevance

The AR(1) regressor process considered in this paper is the baseline time series model employed for empirical analysis in macroeconomics and finance and the empirical relevance of having a simple inference procedure with general applicability for such a regressor without restricting its stochastic order of integration cannot be overstated. The generality of our methodology makes it suitable for a variety of empirical applications; we provide a brief discussion of its empirical relevance below.

Many macroeconomic series, while persistent, maybe near-stationary and, in this baseline case, our procedure is asymptotically efficient because it is equivalent to OLS; this is in contrast to most existing robust procedures which in this case suffer from considerable power loss (see Section 4 for power comparisons). This is empirically important, since our procedure can provide sharper inference (shorter CIs and more powerful tests) than alternative robust procedures in the baseline case of any ergodic time series linear regression, while, in contrast to OLS, remaining correct as we approach the nonstationary regions.

Local-to-unity processes in C(ii) have played a fundamental role in the development of the theory of cointegration[8] and causal inference in systems of macroeconomic and financial variables. In applied macroeconomics, unit root (as well as cointegration) tests have been employed to test for the existence of long-run equilibrium relationships as well as a variety of macroeconomic theory-implied hypotheses involving unit roots. Examples of such hypotheses include: the Permanent Income Hypothesis, the Uncovered Interest Rate Parity Hypothesis, the Expectation Hypothesis of the Term Structure, as well as Efficient Market Hypothesis. See Rossi (2007) for a detailed review on the how these hypotheses can be formulated in terms of the tests on the slope parameter $\beta$ in the PR model in (25), as well as long-horizon versions of it.

Local projections have been utilised to recover effects of fundamental shocks. Allowing for a general process $x_t$ in the framework of (3) permits for the effect of shocks to be short-lived, longer-lived but temporary, permanent or near-permanent, or even continuously increasing, without restricting the parameter space of permitted IRFs. Moreover, once we leave the mean-reverting framework of purely stationary processes, conducting correct inference at longer horizons becomes relevant, since the IRF $r_h$ may not be zero as $h \to \infty$.

In finance, the PR model in (2) is the workhorse reduced-form model for testing for risk premia or time-varying expected returns. In that literature, the interest has been on testing for predictability of financial returns: for example: (i) stock returns by the dividend yield (Fama and French (1988)), (ii) bond returns by bond yield spreads (Keim and Stambaugh (1986)), (iii) changes in spot exchange rates by exchange rates spreads (Fama (1984)). These are the classic contributions that have generated a vast empirical literature that continues to the present day (e.g. see Goyal et al. (2024) for a recent study on equity risk premia). The predictability of returns has been interpreted as evidence for either market inefficiency or time-varying expected returns (see Fama (1991)). While returns are not highly autocorrelated, the candidate lagged explanatory variables considered are highly persistent series. As a result, including a persistent stochastic regressor leads to spuriously discovering evidence of strong predictability (i.e. OLS-based tests on $\beta$ incorrectly reject the null $\beta = 0$ with large size distortions, a problem which does not improve with the sample size). Consequently, more persistent series are more likely to be found significant in the absence of actual predictability leading to wrong empirical conclusions. More recently, Fama-type predictive regressions have been considered in the presence of possibly

[8] While the setup of the PR model in (2) is of a predictive regression, our approach can easily handle the simple cointegrating regression (by amending the innovation assumption). We choose to abstract from this, since our assumptions on $\varepsilon_t$ rule out the spurious regression case. In fact, the instrumental procedure of this paper can be amended to also provide correct inference in the spurious regression case, for breivity, we leave this for future work.

mildly explosive regressors, see for example, Pavlidis et al. (2017). Our procedure provides an inference solution for predictability testing in such asset returns applications: it does not suffer from size distortions (see Section 4 for empirical size), while remaining uniformly valid when there is uncertainty surrounding the regressor's order of integration.

In asset pricing models, the debate on whether asset prices follow a random walk dates back to Kendall (1953). The Efficient Market Hypothesis implies they do, where explosive AR(1) processes have been considered to allow for the formation of speculative bubbles (e.g. Blanchard 1979). The literature on testing and dating episodes of bubbles in financial and commodity prices during periods of market exuberance includes, among others, Phillips and Yu (2011), Phillips et al. (2011), Phillips et al. (2015a,b), Harvey et al. (2016), where under the null, asset prices follow random walks, while under bubbly episodes, they exhibit explosivity. Existing approaches in this literature model bubbles as mildly explosive episodes but assume away the purely explosive case $\rho > 1$, due to the lack of asymptotic validity of existing test procedures in this region. There is also a literature on testing for speculative bubbles in foreign exchange markets (Evans (1986), Pavlidis et al. (2017)). Our test statistic for $\rho$ in Theorem 2 is robust to this criticism and can be directly utilised[9] to deliver tests which remain valid even if the bubble period arises in the case $\rho > 1$. In addition, it provides considerable power gains relative to the IVX procedure for one-sided bubble alternative tests (see Section 5.2).

Finally, AR processes with coefficients in the explosive region $(1, \infty)$ have also been employed for the modelling of other phenomena whose temporal evolution exhibits stochastic exponential growth, from the rate of epidemic infection to modelling climate dynamics. For example, the most widely used models in the epidemiological literature on disease transmission are versions of the classic susceptible-infected-removed (SIR) model for disease transmission. Upon linearisation, these models imply that the number of active infections evolves as an AR(1) process with an explosive (stable) root whenever the basic reproduction number $r_0$ is above (below) unity. In this setup, our procedure can be employed to construct CIs and make correct probability statements for $r_0$ and other epidemiological parameters of SIR models without *a priori* knowledge on whether the epidemic is in a controllable or uncontrollable stage, i.e. without restricting the parameter space, which has important policy implications (see Section 5.1).

At this point, it is important to highlight that our procedure has an advantage over existing procedures not only because it adds the missing part of C(ii) and the near-explosive regions C(iii) to the analysis but, crucially, because our data-driven instrument removes all limit discontinuities in the null limit distribution and achieves uniform inference. It is not possible to distinguish an

[9] It is straightforward to extend our procedure to recursive schemes for bubble detection which explicitely model the switching between bubble and non-bubble periods, as in Phillips et al. (2011).

exact unit root from a local-to-unity process (see Cavanagh et al. (1995)), but it is also virtually impossible in small samples to differentiate between an AR root exactly one and a root close, but not equal, to unity, for example $\rho$ in $(1-\delta, 1+\delta)$ for a small $\delta > 0$. Since most macroeconomic and financial time series exhibit persistence close, but not necessarily exactly equal, to that of a unit root, not restricting the true value of $\rho$ to be contained inside the interval $(-1, 1]$, as is the case with all existing approaches, is essential in practice and useful beyond considering purely explosive processes *per se*, since it allows the construction of CIs for processes with any root close to unity. In contrast, *all* existing approaches can only provide truncated CIs that remove the part of the parameter space not considered. To our knowledge, no existing procedure can achieve valid inference when the true value for $\rho$ lies in $(1-\delta, 1+\delta)$ even for small $\delta > 0$, since the asymptotics in this case could be driven by either stationary, local-to-unity or explosive behaviour. Having a robust procedure that can achieve correct inference without restricting the parameter space of $\rho$ and without prior knowledge on the exact stochastic integration order of the process is invaluable for econometric analysis with any persistent economic series.

Oscillating processes with roots in $[-1, 0]$ arise naturally in series which exhibit seasonality at certain frequency. Seasonal unit root tests are routinely used to test hypotheses on whether shocks have permanent effect on the seasonal pattern of the series (e.g. Hylleberg et al. (1990), Chambers et al. (2014)). Oscillating explosive processes with $\rho \in (-\infty, -1)$ are of limited empirical relevance for economics, we choose to include them in the analysis in order to cover the entire parameter space $(-\infty, \infty)$.

# 3 Theoretical Development

## 3.1 Assumptions

We consider the AR(1) process $x_t$ in (1) with root $\rho$, innovation sequence $(u_t)_{t\in\mathbb{N}}$ and initialisation $X_0 := x_0 - \mu$. For the PR model, we consider $y_t$ in (2) with slope $\beta$ and innovation sequence $(\varepsilon_t)_{t\in\mathbb{N}}$. Assumptions 1-3 specify the AR(1) model (1). Assumption 4 relaxes Assumption 2 on the AR innovations $(u_t)$ when (1) is employed as a predictor in the PR model (2). Definitions 2 and 3 provide explicit expressions for the parameter spaces of the AR and PR models respectively.

**Assumption 1 (AR root).** *$\rho \in \Theta_\rho := [-M, M]$ for some (arbitrarily large) $M > 0$.*

**Assumption 2 (AR innovation sequence).** *Given a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, $u_t$ in (1) is an $\mathcal{F}_t$-martingale difference sequence such that for some fixed* [10] *$\delta, B \in (0, \infty)$*

$$\liminf_{t\to\infty} \mathbb{E}_{\mathcal{F}_{t-1}} |u_t| \geq \delta \ \ a.s. \tag{12}$$

*and $\sigma_t^2 := \mathbb{E}_{\mathcal{F}_{t-1}}(u_t^2)$ and $\sigma^2 := \mathbb{E}\sigma_t^2$ satisfy one of the following conditions:*

[10] We denote by $\delta$ and $B$ small and large positive constants, not necessarilty the same across equations.

(i) $\sigma_t^2 = \sigma^2$ *a.s. for all* $t$ *and* $(u_t^2)_{t\in\mathbb{Z}}$ *is a uniformly integrable (UI) sequence.*

(ii) $\sigma_t^2$ *is generated by a stationary ARCH*$(\infty)$ *process:* $u_t = \eta_t\sigma_t$, $\mathcal{F}_t = \sigma\left(\eta_t, \eta_{t-1}, ...\right)$, $(\eta_t)$ *is an i.i.d.* $(0,1)$ *sequence with* $\mathbb{E}\left(\eta_1^4\right) \leq B$, $\mathbb{E}\left(\sigma_1^4\right) \leq B$, $\sup_{t\geq 1}\sigma_t^2 < \infty$ *a.s. and*

$$\sigma_t^2 = \varpi + \sum_{i=1}^{\infty}\alpha_i u_{t-i}^2, \ \ \alpha_i \geq 0, \ \varpi \geq \delta, \ \sum_{i=1}^{\infty}\alpha_i \leq 1-\delta, \ \sum_{i=1}^{\infty} i^{\delta}\alpha_i \leq B. \tag{13}$$

**Assumption 3 (IC).** $X_0$ *is* $\mathcal{F}_0$*-measurable and* $\mathbb{E}\left|X_0\right|^{\delta} \leq B < \infty$ *for some* $\delta > 0$.

**Assumption 4 (PR innovation sequence).** *The innovation sequence* $(u_t)_{t\in\mathbb{N}}$ *in (1) is a linear process of the form* $u_t = \sum_{j=0}^{\infty} c_j e_{t-j}$, *where, for some fixed* $\delta, B \in (0,\infty)$, $(c_j)_{j\geq 0}$ *is a sequence of constants satisfying* $\sum_{j=0}^{\infty} j^{1+\delta}c_j^2 \leq B$, $c_0 = 1$ *and* $C_\rho(1) = \sum_{j=0}^{\infty}\rho^{-j}c_j$ *satisfies* $|C_\rho(1)| \geq \delta$ *for* $|\rho| \geq 1$. *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 (12) holds with* $u_t$ *replaced by* $e_t$, $\mathbb{E}\left(v_t v_t'\right) = \Sigma$ *for all* $t$ *with minimal eigenvalue* $\lambda_{\min}(\Sigma) \geq \delta > 0$, *and one of the following conditions is satisfied:*

**(i)** $\mathbb{E}_{\mathcal{F}_{t-1}}\left(v_t v_t'\right) = \Sigma$ *a.s. for all* $t$ *and* $\left(\|v_t\|^2\right)_{t\in\mathbb{Z}}$ *is a UI sequence;*

**(ii)** $\left(\varepsilon_t^2, \mathbb{E}_{\mathcal{F}_{t-1}}\left(\varepsilon_t^2\right), e_t^2, \mathbb{E}_{\mathcal{F}_{t-1}}\left(e_t^2\right)\right)_{t\in\mathbb{Z}}$ *is a strictly stationary process satisfying* $\mathbb{E}\varepsilon_0^4 \leq B$, $\mathbb{E}e_0^4 \leq B$, $\sup_{t\in\mathbb{Z}}\mathbb{E}_{\mathcal{F}_{t-1}}\left(e_t^2\right) < \infty$ *a.s., and (13) is satisfied with* $(u_t^2, \sigma_t^2)$ *replaced by* $\left(\varepsilon_t^2, \mathbb{E}_{\mathcal{F}_{t-1}}\left(\varepsilon_t^2\right)\right)$.

**Definition 2 (AR parameter space).** *Let* $\Theta_\rho$ *be defined in Assumption 1 and consider the sets* $\Theta_u = \left\{(F_t)_{t\in\mathbb{Z}} : F_t(x) = \mathbb{P}(u_t \leq x) \text{ for each } t\in\mathbb{Z} \text{ and } (u_t)_{t\in\mathbb{Z}} \text{ satisfies Assumption 2}\right\}$ *and* $\Theta_{X_0} = \left\{F_{X_0} : F_{X_0}(x) = \mathbb{P}(X_0 \leq x) \text{ and } X_0 \text{ satisfies Assumption 3}\right\}$. *Denote the restriction of* $\Theta_u$ *under conditional homoskedasticity by* $\Theta_u^{\text{hom}} = \left\{(F_t)_{t\in\mathbb{Z}} \in \Theta_u : \sigma_t^2 = \sigma^2 \text{ for each } t\in\mathbb{Z}\right\}$. *The parameter space of the AR model in (1) is the Cartesian product* $\Theta = \Theta_\rho \times \Theta_u \times \Theta_{X_0}$. *We denote by* $\Theta^{\text{hom}} = \Theta_\rho \times \Theta_u^{\text{hom}} \times \Theta_{X_0}$ *the restriction of* $\Theta$ *under conditional homoskedasticity of* $(u_t)$.

**Definition 3 (PR parameter space).** *The parameter space for the PR model (1)-(2) is given by the Cartesian product* $\bar{\Theta} = \Theta_\rho \times \Theta_{\varepsilon u} \times \Theta_{X_0}$ *where:*
$\Theta_{\varepsilon u} = \left\{\left(\bar{F}_t\right)_{t\in\mathbb{Z}} : \bar{F}_t(x,y) = \mathbb{P}(\varepsilon_t \leq x, u_t \leq y) \text{ for each } t \text{ and } (\varepsilon_t, u_t)_{t\in\mathbb{Z}} \text{ satisfies Assumption 5}\right\}$ *with* $\Theta_\rho$, $\Theta_{X_0}$ *as in Definition 1. Denote by* $\Theta_{\varepsilon u}^{\text{hom}} = \left\{(F_t)_{t\in\mathbb{Z}} \in \Theta_{\varepsilon u} : \mathbb{E}_{\mathcal{F}_{t-1}}\varepsilon_t^2 = \sigma_\varepsilon^2\right\}$ *and* $\bar{\Theta}^{\text{hom}} = \Theta_\rho \times \Theta_{\varepsilon u}^{\text{hom}} \times \Theta_{X_0}$ *the restrictions of* $\Theta_{\varepsilon u}$ *and* $\bar{\Theta}$ *under conditional homoskedasticity of* $(\varepsilon_t)$.

**Assumption 5 (Drifting sequence, AR).** *Each element* $\theta_n = \left(\rho_n, (F_{n,t})_{t\in\mathbb{Z}}, F_{n,X_0}\right)$ *of* $\Theta$ *satisfies the following:* $\rho_n \to \rho \in \mathbb{R}$; $(u_{nt}, \mathcal{F}_{nt})_{t\in\mathbb{Z}}$ *in (24) is a martingale difference array such that* $\liminf_{n\to\infty}\liminf_{t\to\infty}\mathbb{E}_{\mathcal{F}_{n,t-1}}|u_{n,t}| > 0$ *a.s.,* $\sigma_n^2 := \mathbb{E}\left(\sigma_{n,t}^2\right) \to \sigma^2 > 0$ *and* $\sigma_{n,t}^2 := \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(u_{n,t}^2\right)$ *satisfies one of the following conditions:*

(i) $\sigma_{n,t}^2 = \sigma_n^2$ *a.s. for all* $t$ *and* $\max_{1\leq t\leq n}\mathbb{E}\left(u_{n,t}^2\mathbf{1}\left\{u_{n,t}^2 > \lambda_n\right\}\right) \to 0$ *when* $\lambda_n \to \infty$.

(ii) *For each $n$, the process $\left(u_{n,t},\sigma_{n,t}^2\right)_{t\in\mathbb{Z}}$ is strictly stationary with $\sigma_{n,t}^2>0$ a.s., $\sup_{n\geq 1}\mathbb{E}\left(\sigma_{n,1}^4\right)<\infty$, $\limsup_{n\to\infty}\sup_{t\in\mathbb{N}}\sigma_{n,t}^2<\infty$ a.s. and there exist $b>0$ and sequences of positive numbers $(\psi_m)_{m\in\mathbb{N}}$ and $\left(\tilde{\psi}_n\right)_{n\in\mathbb{N}}$ satisfying $\psi_m\to 0$ and $\tilde{\psi}_n\to 0$ such that* [11]

$$\sup_{t\in\mathbb{N}}\left\|\mathbb{E}_{\mathcal{F}_{n,t-1-m}}\left(\sigma_{n,t}^2-\sigma_n^2\right)\right\|_{L_2}\leq b\left(\psi_m+\tilde{\psi}_n\right)\quad \text{for all } m,n\geq 1. \tag{14}$$

*When $\rho\in(-1,1)$, $v_1(\rho):=\lim_{n\to\infty}\mathbb{E}u_{n,1}^2\left(\sum_{j=0}^{\infty}\rho^j u_{n,-j}\right)^2$ exists in $(0,\infty)$.*

*$X_{n,0}\to_d X_0$ where $X_0$ is a $\mathcal{F}_0$-measurable random variable with $\mathcal{F}_0=\sigma\left(\cup_{n\in\mathbb{N}}\mathcal{F}_{n,0}\right)$. When $|\rho|>1$, the sequence $(U_n)_{n\in\mathbb{N}}$ defined by $U_n:=\sum_{j=1}^{n}\rho^{-j}u_{n,j}$ converges in distribution jointly with $X_{n,0}$.*

**Assumption 6 (Drifting sequence, PR).** *Each element $\bar{\theta}_n=\left(\rho_n,\left(\bar{F}_t\right)_{t\in\mathbb{Z}},F_{n,X_0}\right)$ of $\bar{\Theta}$ satisfies the following: $\rho_n\to\rho\in\mathbb{R}$; the sequence $(u_{n,t})_{t\in\mathbb{N}}$ in (24) is a linear process of the form $u_{n,t}=\sum_{j=0}^{\infty}c_{n,j}e_{n,t-j}$, where $\left(c_{n,j}\right)_{j\geq 0}$ is a sequence of constants satisfying $\sup_{n\in\mathbb{N}}\sum_{j=0}^{\infty}j^{1+\delta}c_{n,j}^2<\infty$ for some $\delta>0$ and $\sum_{j=0}^{\infty}\rho_n^{-j}c_{n,j}\to\sum_{j=0}^{\infty}\rho^{-j}c_j\neq 0$ for $|\rho|\geq 1$. The autocovariance function $\gamma_{u_n}(h)=\mathbb{E}\left(u_{n,t}u_{n,t-h}\right)$ of $u_{n,t}$ satisfies $\lim_{n\to\infty}\gamma_{u_n}(h)=\gamma(h)$, $h\in\mathbb{Z}$. Given a filtration $(\mathcal{F}_{n,t})_{t\in\mathbb{Z}}$, the sequence $v_{n,t}:=\left(\varepsilon_{n,t},e_{n,t}\right)'$ is an $\mathcal{F}_{n,t}$-martingale difference array such that (12) holds with $u_{n,t}$ replaced by $e_{n,t}$, $\Sigma_n:=\mathbb{E}\left(v_{n,1}v_{n,1}'\right)\to\Sigma>0$ and one of the following conditions is satisfied:*

(i) *$\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(v_{n,t}v_{n,t}'\right)=\Sigma_n$ a.s. for all $t$, $\max_{1\leq t\leq n}\mathbb{E}\left\|v_{n,t}\right\|^2\mathbf{1}\left\{\left\|v_{n,t}\right\|^2>\lambda_n\right\}\to 0$ when $\lambda_n\to\infty$.*

(ii) *$\left(e_{n,t},\mathbb{E}_{\mathcal{F}_{n,t-1}}e_{n,t}^2,\varepsilon_{n,t},\mathbb{E}_{\mathcal{F}_{n,t-1}}\varepsilon_{n,t}^2\right)_{t\in\mathbb{Z}}$ is strictly stationary for each $n$ with $\sup_{n\in\mathbb{N}}\mathbb{E}e_{n,0}^4<\infty$, $\sup_{n\in\mathbb{N}}\mathbb{E}\varepsilon_{n,0}^4<\infty$, $\limsup_{n\to\infty}\sup_{t\in\mathbb{N}}\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(e_{n,t}^2\right)<\infty$ a.s. and (14) is satisfied with $\left(\sigma_{n,t}^2,\sigma_n^2\right)$ replaced by $\left(\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\varepsilon_{n,t}^2\right),\mathbb{E}\varepsilon_{n,0}^2\right)$; $v_2(\rho):=\lim_{n\to\infty}\mathbb{E}\varepsilon_{n,1}^2\left(\sum_{j=0}^{\infty}\rho^j u_{n,-j}\right)^2$ exists in $(0,\infty)$ when $\rho\in(-1,1)$.*

*Assumption 4 holds for the sequences $(X_{n,0})_{n\in\mathbb{N}}$ and $(U_n)_{n\in\mathbb{N}}$.*

**Assumption 7 (AR categories).** *The limit $c:=\lim_{n\to\infty}n\left(|\rho_n|-1\right)$ exists in $\mathbb{R}\cup\{-\infty,\infty\}$.*

**Remarks on Assumptions 1-7.**

**1.** Assumption 1 accommodates all types of AR stochastic behaviour, including stationary ($\rho\in(-1,1)$), unit root ($\rho=1$) and explosive ($\rho>1$) processes as well as their oscillating unit root and explosive ($\rho=-1$ and $\rho<-1$) counterparts. The AR specification (1), designed to introduce an intercept while maintaining the stochastic structure of a nonstationary autoregression by reducing the contribution of the intercept as $\rho$ approaches unity, is standard in the literature: see Andrews (1993), Mikusheva (2007), Andrews and Guggenberger (2009, 2014). It is well-known that a process of the form $x_t=\mu+\rho x_{t-1}+u_t$ behaves asymptotically as a linear deterministic trend

[11] Condition (14) is a mixingale array condition: since $\sigma_{n,t}^2-\sigma_n^2$ is $\mathcal{F}_{n,t-1}$-adapted, condition (14) implies that $\left(\sigma_{n,t}^2-\sigma_n^2,\mathcal{F}_{n,t-1}\right)$ is a mixingale sequence; see p.19 of Hall and Heyde (1980) or Andrews (1988).

when $\rho = 1$[12]. Our procedure can accommodate such degeneracies of AR stochastic behaviour, in the sense that Theorems 1-3 continue to hold (which is in contrast to existing robust procedures), but we omit the details as such deterministic trends have limited relevance for economic modelling.

**2.** The martingale difference (m.d.) condition of Assumption 2 on $(u_t)$ is standard for inference with AR processes; when $\rho$ is not the parameter of interest (as is the case for the PR framework), $(u_t)$ is extended to a short memory linear process (Assumption 4).

**3.** The moment conditions imposed on $(u_t)$ vary according to its conditional variance with uniform integrability of $(u_t^2)$ required under conditional homoskedasticity (Assumption 2(i) for the AR and LP models and 5(i) for the PR model respectively) and $\mathbb{E}u_t^4 < \infty$ required under conditional heteroskedasticity (Assumption 2(ii) for the AR and LP models and 5(ii) for the PR model respectively). For the homoskedastic case, (12) and the Chebyshev inequality ensure $\sigma^2 > 0$; for the heteroskedastic case, Theorem 2.1 of Giraitis et al. (2000) guarantees that (13) has a unique strictly stationary causal solution with $\sigma^2 = \varpi/(1 - \sum_{i=1}^{\infty} \alpha_i)$. This ARCH($\infty$) solution encompasses the solution of any stationary GARCH$(p, q)$ process with finite second moment as a special case: the summability condition $\sum_{i=1}^{\infty} i^{\delta} \alpha_i \leq B$ for some $\delta > 0$ does not impose a restriction on the ARCH($\infty$) representation of a stationary GARCH$(p, q)$ process, as the latter is known to have exponentially decaying autocovariance function whereas the autocovariance function of an ARCH($\infty$) process satisfying $\sum_{i=1}^{\infty} i^{\delta} \alpha_i < \infty$ has an autocovariance function $\gamma(\cdot)$ that decays at hyperbolic rate: $|\gamma(k)| = O\left(k^{-1-\delta}\right)$, see Proposition 3.2 of Giraitis et al. (2000). Consequently, our heteroskedasticity conditions in Assumptions 2(ii) and 4(ii) hold when the conditional variance is any stationary GARCH$(p, q)$ process[13] satisfying $\sup_{t \geq 1} \sigma_t^2 < \infty$ *a.s.* and $\mathbb{E}\left(\sigma_1^4\right) \leq B$.

**4.** Condition (12) in Assumption 2 and $\sup_{t \geq 1} \sigma_t^2 < \infty$ *a.s.* imply that the m.d. sequence $(u_t, \mathcal{F}_t)$ satisfies local Marcinkiewicz-Zygmund (MZ) conditions that ensure that, in the explosive case $|\rho| > 1$, the denominator of the OLS estimator is asymptotically non-zero *a.s.*: see Corollary 2 of Lai and Wei (1983).

**5.** Assumption 3 accommodates $\sigma(u_0, u_{-1}, ...)$-measurable random variables as initial conditions and removes the condition $X_0 = 0$ typically imposed in the explosive case: see Wang and Yu (2015) for the effect of $X_0$ on OLS asymptotic inference when $\rho > 1$.

**6.** The parameter spaces of Definitions 1 and 2 include the marginal distribution functions

[12] Since $\rho = -1$ does not increase the order of magnitude of the non-stochastic component of $x_t$ that is driven by the intercept, no adjustment is required in the oscillating case.

[13] Within the ARCH($\infty$) or GARCH$(p, q)$ framework, Assumption 2(ii) is weaker than Assumption INNOV(iii) of Andrews and Guggenberger (2012) which requires finite 6 moments for $(u_t)$ with further moment conditions imposed for the GARCH(1,1) example of their equation (5). On the other hand, the results of Andrews and Guggenberger (2012) hold under a strong mixing condition with conditional heteroskedasticity not necessarily generated by an ARCH($\infty$) recursion.

of the innovation sequences $(u_t)$ and $(\varepsilon_t)$ and the distribution function of the initial condition as (infinite dimensional) nuisance parameters. Such analysis provides insight on the sensitivity of autoregressive and predictive regression inference to the distributional characteristics of the innovations and the intitialisation, an issue of particular relevance for explosive processes where these characteristics are known to affect standard OLS asymptotic inference.

**7**. Our development of critical regions (CRs) and CIs with uniform asymptotic properties over the parameter spaces of Definitions 2 and 3 employs a *drifting sequence* approach, see Andrews, Cheng and Guggenberger (2020) for a general discussion. This requires the derivation of the limit distribution of the t-statistics proposed in Section 3.2 over an arbitrary sequence of drifting parameters from the parameter spaces of Definitions 2 and 3. For presentational and notational economy, convergence properties that hold only subsequentially are typically assigned to drifting sequences in assumptions designed to prove intermediate results and we follow this convention in Assumptions 5 and 6 used to establish Theorem 1. The existence of a limit for $(\sigma_n^2)_{n\in\mathbb{N}}$ and of $v_1(\rho)$ and $v_2(\rho)$ is ensured along a subsequence by the Bolzano-Weierstrass theorem; similarly, the existence of a limit in distribution for $(X_{n,0})_{n\in\mathbb{N}}$ and $(U_n)_{n\in\mathbb{N}}$ is ensured along a subsequence by tightness (implied by $\sup_{n\geq 1}\mathbb{E}\left|X_{n,0}\right|^{\eta}<\infty$ for some $\eta>0$ and $\sup_{n\geq 1}\mathbb{E}\left|U_n\right|<\infty$). The same holds for Assumption 7 that strengthens[14] the requirement $\rho_n\to\rho\in\mathbb{R}$ so that $x_t$ can be classified according to Definition 1 (see Lemma 1(i) in the Appendix). It is important to note Assumptions 5-7 are only needed to establish Theorem 1, an intermediate result on the limit distribution of IV estimators and t-statistics under drifting sequences of parameters, and are not used in the main results of the paper on the uniform asymptotic validity of CRs and CIs (Theorems 1 and 2 and Corollary 1).

**8.** Assumption 6 implies a short memory array condition: $\sup_{n\in\mathbb{N}}\sum_{j=0}^{\infty}j^q\left|c_{n,j}\right|<\infty$ for some $q>0$, see (A.43) in the online Appendix. This together with the lack of negative memory condition $\sum_{j=0}^{\infty}\rho^{-j}c_j\neq 0$ for $|\rho|\geq 1$ and the mixingale array condition (14) of Assumption 6, imply a LLN for the sample autocovariance of the array $(u_{n,t})$; see Lemma 1 of Magdalinos and Petrova (2024). Under Assumption 6, we denote the autocovariance function and long-run variance of $(u_{n,t})$ by $\gamma_{u_n}(\cdot)$ and $\omega^2=\lim_{n\to\infty}\sum_{k=-\infty}^{\infty}\gamma_{u_n}(k)=C(1)^2\sigma^2$ respectively. When $|\rho|\leq 1$ we denote

$$\Gamma_n=\sum_{k=1}^{\infty}\rho_n^{k-1}\gamma_{u_n}(k) \quad \text{and} \quad \Gamma=\lim_{n\to\infty}\Gamma_n=\sum_{k=1}^{\infty}\rho^{k-1}\gamma(k)\,, \tag{15}$$

with $\omega^2,\Gamma\in\mathbb{R}$ by Assumption 6, $\rho_n\to\rho\in[-1,1]$ and the dominated convergence theorem.

[14] When $|\rho|=1$ Assumption 7 is more restrictive than $\rho_n\to\rho\in\mathbb{R}$: e.g. $\rho_n=1+(-1)^n/k_n$, $k_n\to\infty$ and $k_n/n=O(1)$ satisfies $\rho_n\to 1$ but $c$ does not exist. However, Assumption 7 holds subsequentially; see Lemma 1(i).

### 3.2 IV-Based Inference Procedures

Correct studentisation of the IV estimators $\tilde{\rho}_n$, $\tilde{r}_{h,n}$ and $\tilde{\beta}_n$ depends on the conditional variance of the innovations $u_t$ in (1) and $\varepsilon_t$ in (2), with conventional studentisation employed under conditional homoskedasticity (Assumptions 2(i) and 4(i)) and Eicker-White standard errors employed under conditional heteroskedasticity (Assumptions 2(ii) and 4(ii)). Denoting $X = (x_1, ..., x_{n-1})'$, $\tilde{Z} = (\tilde{z}_1, ..., \tilde{z}_{n-1})'$ and $\underline{X} = (x_1 - \bar{x}_{n-1}, ..., x_{n-1} - \bar{x}_{n-1})'$, we define t-statistics under the null based on the IV estimator $\tilde{\rho}_n$ and conventional/Eicker-White standard errors as follows:

$$T_n = \hat{\sigma}_n^{-1} \left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{1/2} (\tilde{\rho}_n - \rho) \quad \text{and} \quad T_n^{EW} = Q_n^{-1/2} (\tilde{\rho}_n - \rho) \tag{16}$$

where $P_{\tilde{Z}} = \tilde{Z}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'$, $Q_n = \left(\underline{X}'\tilde{Z}\right)^{-2} \sum_{t=1}^n \tilde{z}_{n,t-1}^2 \left(\hat{u}_{n,t}^2 \mathbf{1}_{F_n} + \hat{\sigma}_n^2 \mathbf{1}_{\bar{F}_n}\right)$, $\hat{\sigma}_n^2 = n^{-1} \sum_{t=1}^n \hat{u}_{n,t}^2$ and $\hat{u}_{n,t}$ are the OLS residuals of (1). For the PR model in (1) and (2), we employ a similar studentisation to (16) based on the IV estimator $\tilde{\beta}_n$:

$$\bar{T}_n = \hat{\sigma}_\varepsilon^{-1} \left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{1/2} \left(\tilde{\beta}_n - \beta\right) \quad \text{and} \quad \bar{T}_n^{EW} = Q_{n,\varepsilon}^{-1/2} \left(\tilde{\beta}_n - \beta\right) \tag{17}$$

where $Q_{n,\varepsilon} = \left(\underline{X}'\tilde{Z}\right)^{-2} \sum_{t=1}^n \tilde{z}_{n,t-1}^2 \left(\hat{\varepsilon}_{n,t}^2 \mathbf{1}_{F_n} + \hat{\sigma}_\varepsilon^2 \mathbf{1}_{\bar{F}_n}\right)$, $\hat{\sigma}_\varepsilon^2 = n^{-1} \sum_{t=1}^n \hat{\varepsilon}_{n,t}^2$ and $\hat{\varepsilon}_{n,t}$ are the OLS residuals of (2). For the LP model in (3), we define the t-statistic based on the IV estimator $\tilde{r}_{h,n}$:

$$T_{n,h} = Q_{n,h}^{-1/2} (\tilde{r}_{h,n} - r_h) \quad \text{and} \quad T_{n,h}^{EW} = \left(Q_{n,h}^{EW}\right)^{-1/2} (\tilde{r}_{h,n} - r_h), \tag{18}$$

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

$$Q_{n,h} = v_n \left(X'\tilde{Z}\right)^{-2} \mathbf{1}_{F_n} + \hat{\sigma}_n^2 \left(X' P_{\tilde{Z}} X\right)^{-1} \left(|\phi_{2n}|^{-h} \sum_{j=0}^{h-1} \left(|\hat{\rho}_n| \, |\phi_{2n}|\right)^j\right)^2 \mathbf{1}_{\bar{F}_n} \tag{19}$$

with $v_n = \hat{\sigma}_n^2 \sum_{i=1}^{n-2h+1} \left(\sum_{t=0}^{h-1} \hat{\rho}_n^t \tilde{z}_{n,t+i}\right)^2$ and $Q_{n,h}^{EW}$ is defined in the same way as $Q_{n,h}$ with $v_n$ replaced by $v_n^{EW} = \sum_{i=1}^{n-2h+1} \hat{u}_i^2 \left(\sum_{t=0}^{h-1} \hat{\rho}_n^t \tilde{z}_{n,t+i}\right)^2$. When $h = 1$, $T_{n,1} = T_n$ and $T_{n,1}^{EW} = T_n^{EW}$, so the LP t-statistics reduce to the t-statistics for the AR model.

The t-statistics in (16) may be used to construct critical regions (CRs):

$$\mathcal{R}_{n,\alpha} = \left\{|T_n| > \Phi^{-1}(1 - \alpha/2)\right\} \quad \text{and} \quad \mathcal{R}_{n,\alpha}^{EW} = \left\{\left|T_n^{EW}\right| > \Phi^{-1}(1 - \alpha/2)\right\} \tag{20}$$

where $\Phi$ denotes the $\mathcal{N}(0,1)$ cdf, and $(1 - \alpha)\%$ confidence intervals (CIs) are given by:

$$I_n(\tilde{\rho}_n, \alpha) = [\tilde{\rho}_n - c_n(\alpha), \tilde{\rho}_n + c_n(\alpha)] \quad \text{and} \quad I_n^{EW}(\tilde{\rho}_n, \alpha) = \left[\tilde{\rho}_n - c_n^{EW}(\alpha), \tilde{\rho}_n + c_n^{EW}(\alpha)\right] \tag{21}$$

with $c_n(\alpha) = \left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{-1/2} \Phi^{-1}(1 - \alpha/2) \hat{\sigma}_n$ and $c_n^{EW}(\alpha) = Q_n^{1/2} \Phi^{-1}(1 - \alpha/2)$.

Similarly, for the PR and LP models, we denote by $\bar{\mathcal{R}}_{n,\alpha} = \left\{\left|\bar{T}_n\right| > \Phi^{-1}(1 - \alpha/2)\right\}$ and $\mathcal{R}_{n,h,\alpha} = \left\{|T_{n,h}| > \Phi^{-1}(1 - \alpha/2)\right\}$ the CRs of the t-tests in (17) and (18) based on conventional studentisation and by $(\bar{\mathcal{R}}_{n,\alpha}^{EW}, \mathcal{R}_{n,h,\alpha}^{EW})$ their heteroskedasticity consistent counterparts with $(\bar{T}_n, T_{n,h})$ replaced by $(\bar{T}_n^{EW}, T_{n,h}^{EW})$. Further denote by $I_n(\tilde{\beta}_n, \alpha)$, $I_n(\tilde{r}_{h,n}, \alpha)$, $I_n^{EW}(\tilde{\beta}_n, \alpha)$ and $I_n^{EW}(\tilde{r}_{h,n}, \alpha)$ the CIs corresponding to $\bar{\mathcal{R}}_{n,\alpha}$, $\mathcal{R}_{n,h,\alpha}$, $\bar{\mathcal{R}}_{n,\alpha}^{EW}$ and $\mathcal{R}_{n,h,\alpha}^{EW}$ respectively.

Finally, it will be useful to consider a finite sample correction to $\tilde{\beta}_n$ based on the fully-modified (FM) transformation of Phillips and Hansen (1990) that orthogonalises the innovations $\varepsilon_{n,t}$ of (2) with respect to the innovations $u_{n,t}$ of (1): see the discussion in Remark 6 in Section 3.3. The

FM-corrected IV estimator $\tilde{\beta}_n$ in (10) takes the form

$$\beta_n^* = \left(\sum_{t=1}^n \underline{y}_{n,t}\tilde{z}_{t-1} + \hat{\delta}_{\varepsilon u}\left(\hat{\sigma}_\varepsilon/\hat{\omega}_n\right) x_{n,n}\bar{z}_{n-1}\mathbf{1}_{F_n^+}\right)\left(\sum_{t=1}^n \underline{x}_{n,t-1}\tilde{z}_{t-1}\right)^{-1} \tag{22}$$

where $\hat{\sigma}_\varepsilon^2$, $\hat{\omega}_n^2$ and $\hat{\delta}_{\varepsilon u}$ are consistent estimators of $\sigma_\varepsilon^2$, $\omega^2$ and $\delta_{\varepsilon u} = corr\left(\varepsilon_{n,t}, u_{n,t}\right)$. A computation of the standard error of the estimator in (22) gives rise to the following finite sample corrected t-statistics using conventional and EW standard errors: letting $\Delta_n = n\bar{z}_{1,n-1}^2(1-\hat{\delta}_{\varepsilon u}^2)\mathbf{1}_{F_n^+}$,

$$T_n^* = \Sigma_n^{-1/2}\left(\beta_n^* - \beta_n\right) \text{ and } T_n^{*EW} = Q_{n,\varepsilon}^{*-1/2}\left(\beta_n^* - \beta_n\right), \tag{23}$$

where $\Sigma_n = \left(\underline{X}'\tilde{Z}\right)^{-2}\left(\tilde{Z}'\tilde{Z} - \Delta_n\right)\hat{\sigma}_\varepsilon^2$ and $Q_{n,\varepsilon}^* = \left(\underline{X}'\tilde{Z}\right)^{-2}\left[\sum_{t=1}^n \tilde{z}_{n,t-1}^2\left(\hat{\varepsilon}_{n,t}^2\mathbf{1}_{F_n} + \hat{\sigma}_\varepsilon^2\mathbf{1}_{\bar{F}_n}\right) - \Delta_n\hat{\sigma}_\varepsilon^2\right]$. We denote by $(\mathcal{R}_{n,\alpha}^*, I_n(\beta_n^*, \alpha))$ and $(\mathcal{R}_{n,\alpha}^{*EW}, I_n^{EW}(\beta_n^*, \alpha))$, the CR and CI based on $T_n^*$ and $T_n^{*EW}$.

## 3.3 Main Theoretical Results

Writing (1) and (2) along sequences of drifting parameters in $\Theta$ and $\bar{\Theta}$ yields

$$x_{n,t} = \mu\left(1-\rho_n\right) + \rho_n x_{n,t-1} + u_{n,t}, \quad x_{n,0} = \mu + X_{n,0} \tag{24}$$

$$y_{n,t} = \gamma + \beta x_{n,t-1} + \varepsilon_{n,t}. \tag{25}$$

We first derive the limit distribution of the IV estimators of Section 2.2 and the t-statistics of Section 3.2 under the above triangular array specification. Theorem 1 shows the consistency rate of the IV estimators is given by

$$\pi_n = \begin{cases} n^{1/2}\left(1-\rho_n^2\phi_{1n}^2\right)^{-1/2}, & \text{under C(i)} \\ n^{1/2}\left(1-\phi_{1n}^2\right)^{-1/2}\mathbf{1}_{F_n} + 2n^{1/2}\left(\phi_{2n}^2-1\right)^{-1/2}\mathbf{1}_{\bar{F}_n}, & \text{under C(ii)} \\ \left(\phi_{2n}^2-1\right)^{1/2}\left(|\rho_n|\,|\phi_{2n}|-1\right)^{-1}\left(\rho_n^2-1\right)^{-1/2}|\rho_n|^n, & \text{under C(iii)} \end{cases} \tag{26}$$

for the events $F_n$ in (5), $\phi_{1n} = \varphi_{1n}\mathbf{1}\{\rho \geq 0\} + \varphi_{1n}^-\mathbf{1}\{\rho < 0\}$, $\phi_{2n} = \varphi_{2n}\mathbf{1}\{\rho \geq 0\} + \varphi_{2n}^-\mathbf{1}\{\rho < 0\}$.

Let $W(t)$ denote a standard Brownian motion on $[0,1]$ and $B(t) = \omega W(t)$. When $c \in \mathbb{R}$ in Assumption 7, define the Ornstein-Uhlenbeck processes

$$W_c(t) = \textstyle\int_0^t e^{c(t-s)}dW(s) \text{ and } J_c(t) = \int_0^t e^{c(t-s)}dB(s), \tag{27}$$

the random variables $K_c = \int_0^1 J_c(r)\,dB(r)/\int_0^1 J_c(r)^2\,dr$, $\Psi_2(c) = W_c(1) - \int_0^1 W_c(r)\,dr\mathbf{1}\{\rho = 1\}$ and $\Psi_1(c) = \left(\sigma^2 + 2\rho\Gamma\right)/\omega^2 + W_c(1)^2 - 2W_c(1)\int_0^1 W_c(r)\,dr\mathbf{1}\{\rho = 1\}$, the event $F_c = \{K_c + c \leq 0\}$ and its complement $\bar{F}_c$.

Define by $\Theta_0$ the restriction of $\Theta$ when $F_t(x) = \Phi_{\sigma^2}(x)$ and by $\bar{\Theta}_0$ the restriction of $\bar{\Theta}$ when $F_t(x,y) = \Phi_\Sigma(x,y)$ for all $t \in \mathbb{Z}$ where $\Phi_{\sigma^2}(\cdot)$ and $\Phi_\Sigma(\cdot,\cdot)$ denote the $\mathcal{N}(0,\sigma^2)$ and $\mathcal{N}(0,\Sigma)$ cdfs with $\sigma^2$ and $\Sigma$ defined in Assumptions 5 and 6.

**Theorem 1.** *Consider the AR model (24), the PR model (25), the IV estimators $\tilde{\rho}_n$, $\tilde{\beta}_n$ and $\beta_n^*$ in (9), (10) and (22), the t-statistics in (16) (17) and (23), and the normalisation sequence $\pi_n$ in (26). For arbitrary sequences $(\theta_n)_{n\in\mathbb{N}}$ in $\Theta$ and $(\theta_n')_{n\in\mathbb{N}}$ in $\Theta^{\text{hom}}$ satisfying Assumptions 5 and 7 and $\left(\bar{\theta}_n\right)_{n\in\mathbb{N}}$ in $\bar{\Theta}$ and $(\bar{\theta}_n')_{n\in\mathbb{N}}$ in $\bar{\Theta}^{\text{hom}}$ satisfying Assumptions 6 and 7, the following hold:*

**(i)** $\lim_{n\to\infty}\mathbb{P}_{\theta_n}\left(\pi_n\left(\tilde{\rho}_n - \rho_n\right) \leq x\right) = \mathbb{P}_{\theta_0}\left(\mathcal{L}_1 \leq x\right)$, $\lim_{n\to\infty}\mathbb{P}_{\bar{\theta}_n}\left(\pi_n(\tilde{\beta}_n - \beta) \leq x\right) = \mathbb{P}_{\bar{\theta}_0}\left(\mathcal{L}_2 \leq x\right)$ *$x \in \mathbb{R}$ with $\theta_0 \in \Theta_0$, $\bar{\theta}_0 \in \bar{\Theta}_0$, $\mathcal{L}_1 =_d \mathcal{MN}(0, V_1)$, $\mathcal{L}_2 =_d \mathcal{MN}(0, V_2)$ and $V_1$ and $V_2$ given by:*

**(a)** *Under C(i):* $V_1 = 1$ *and* $V_2 = \sigma_\varepsilon^2 / (\sigma^2 + 2\rho\Gamma)$ *with* $\Gamma$ *defined in (15), when* $|\rho| = 1$ *or when* $(u_{n,t})$ *and* $(v_{n,t})$ *satisfy Assumptions 5(i) and 6(i).*

**(b)** *Under C(i) with* $\rho \in (-1, 1)$*:* $[V_1, V_2] = (1-\rho^2)^{-1} \upsilon(\rho)^{-2} [\upsilon_1(\rho), \upsilon_2(\rho)]$, $\upsilon(\rho) = \sum_{j,i=0}^{\infty} \rho^{j+i} \gamma(i-j)$ *when* $(u_{n,t})$ *and* $(v_{n,t})$ *satisfy Assumptions 5(ii) and 6(ii).*

**(c)** *Under C(ii),* $V_1 = \Psi(c)^{-2}$ *and* $V_2 = (\sigma_\varepsilon^2/\omega^2)\Psi(c)^{-2}$, *where* $\Psi(c) = \Psi_1(c)\mathbf{1}_{F_c} + \Psi_2(c)\mathbf{1}_{\bar{F}_c}$.

**(d)** *Under C(iii),* $\mathcal{L}_1 = Y/X$, $\mathcal{L}_2 = \tilde{Y}/X$, $V_1 = \sigma^2/X^2$, *and* $V_2 = \sigma_\varepsilon^2/X^2$, *where* $Y =_d \mathcal{N}(0, \sigma^2)$, $\tilde{Y} =_d \mathcal{N}(0, \sigma_\varepsilon^2)$, $X$ *is independent of* $\left(Y, \tilde{Y}\right)$ *and* $X \neq 0$ *a.s.; when* $|\rho| = 1$, $X =_d \mathcal{N}(0, \omega^2)$.

**(ii)** $\lim_{n\to\infty} \mathbb{P}_{\theta_n'}(|T_n| \leq x) = \Phi(x)$ *and* $\lim_{n\to\infty} \mathbb{P}_{\theta_n}\left(\left|T_n^{EW}\right| \leq x\right) = \Phi(x)$, $x \in \mathbb{R}$.

**(iii)** $\lim_{n\to\infty} \mathbb{P}_{\bar{\theta}_n'}\left(\left|\bar{T}_n\right| \leq x\right) = \Phi(x)$ *and* $\lim_{n\to\infty} \mathbb{P}_{\bar{\theta}_n}\left(\left|\bar{T}_n^{EW}\right| \leq x\right) = \Phi(x)$, $x \in \mathbb{R}$, $\pi_n(\tilde{\beta}_n - \beta_n^*) \to_{\mathbb{P}_{\bar{\theta}_n}} 0$, $\left|T_n^{*EW} - \bar{T}_n^{EW}\right| = o_{\mathbb{P}_{\bar{\theta}_n}}(1)$ *and* $\left|T_n^* - \bar{T}_n\right| = o_{\mathbb{P}_{\bar{\theta}_n'}}(1)$.

**Theorem 2.** *Consider the AR process (1), the parameter spaces* $\Theta$ *and* $\Theta^{\text{hom}}$ *of Definition 2, the critical regions in (20) and confidence intervals in (21). For any* $\alpha \in (0, 1)$*:*

**(i)** $\lim_{n\to\infty} \sup_{\theta\in\Theta} \mathbb{P}_\theta\left(\mathcal{R}_{n,\alpha}^{EW}\right) = \alpha$ *and* $\lim_{n\to\infty} \sup_{\theta\in\Theta^{\text{hom}}} \mathbb{P}_\theta(\mathcal{R}_{n,\alpha}) = \alpha$

**(ii)** $\lim_{n\to\infty} \inf_{\theta\in\Theta} \mathbb{P}_\theta\left[\rho \in I_n^{EW}(\tilde{\rho}_n, \alpha)\right] = 1 - \alpha$ *and* $\lim_{n\to\infty} \inf_{\theta\in\Theta^{\text{hom}}} \mathbb{P}_\theta[\rho \in I_n(\tilde{\rho}_n, \alpha)] = 1 - \alpha$.

**Theorem 3.** *Consider the PR model (1)-(2), the parameter spaces* $\bar{\Theta}$ *and* $\bar{\Theta}^{\text{hom}}$ *in Definition 3, the critical regions* $\bar{\mathcal{R}}_{n,\alpha}$, $\bar{\mathcal{R}}_{n,\alpha}^{EW}$, $\mathcal{R}_{n,\alpha}^*$ *and* $\mathcal{R}_{n,\alpha}^{*EW}$ *and their associated confidence intervals* $I_n(\tilde{\beta}_n, \alpha)$, $I_n^{EW}(\tilde{\beta}_n, \alpha)$, $I_n(\beta_n^*, \alpha)$ *and* $I_n^{EW}(\beta_n^*, \alpha)$. *For any* $\alpha \in (0, 1)$*:*

**(i)** *the sequences* $\sup_{\theta\in\bar{\Theta}^{\text{hom}}} \mathbb{P}_\theta(\bar{\mathcal{R}}_{n,\alpha})$, $\sup_{\theta\in\bar{\Theta}} \mathbb{P}_\theta(\bar{\mathcal{R}}_{n,\alpha}^{EW})$, $\sup_{\theta\in\bar{\Theta}^{\text{hom}}} \mathbb{P}_\theta(\mathcal{R}_{n,\alpha}^*)$ *and* $\sup_{\theta\in\bar{\Theta}} \mathbb{P}_\theta(\mathcal{R}_{n,\alpha}^{*EW})$ *all converge to* $\alpha$ *as* $n \to \infty$.

**(ii)** *the sequences* $\inf_{\theta\in\bar{\Theta}^{\text{hom}}} \mathbb{P}_\theta[\beta \in I_n(\tilde{\beta}_n, \alpha)]$, $\inf_{\theta\in\bar{\Theta}} \mathbb{P}_\theta[\beta \in I_n(\tilde{\beta}_n^{EW}, \alpha)]$, $\inf_{\theta\in\bar{\Theta}^{\text{hom}}} \mathbb{P}_\theta[\beta \in I_n(\beta_n^*, \alpha)]$ *and* $\inf_{\theta\in\bar{\Theta}} \mathbb{P}_\theta[\beta \in I_n(\beta_n^{EW*}, \alpha)]$ *all converge to* $1 - \alpha$ *as* $n \to \infty$.

**Corollary 1.** *Consider the AR(1) process (1) with* $\mu = 0$ *and* $r_h = \rho^h$. *The following hold for the local projection CRs and CIs for any* $\alpha \in (0, 1)$ *when* $h/n \to 0$*:*

**(i)** $\lim_{n\to\infty} \sup_{\theta\in\Theta} \mathbb{P}_\theta\left(\mathcal{R}_{n,h,\alpha}^{EW}\right) = \alpha$ *and* $\lim_{n\to\infty} \sup_{\theta\in\Theta^{\text{hom}}} \mathbb{P}_\theta(\mathcal{R}_{n,h,\alpha}) = \alpha$

**(ii)** $\lim_{n\to\infty} \inf_{\theta\in\Theta} \mathbb{P}_\theta\left[r_h \in I_{n,h}^{EW}(\alpha)\right] = 1 - \alpha$ *and* $\lim_{n\to\infty} \inf_{\theta\in\Theta^{\text{hom}}} \mathbb{P}_\theta[r_h \in I_{n,h}(\alpha)] = 1 - \alpha$.

**Remarks.**

**1.** Theorem 1 establishes the asymptotic mixed Gaussianity (AMG) property of the IV estimators in (9), (10) and (22) under drifting sequences for each AR regime C(i)-C(iii) arising from Assumptions 5, 6 and 7. In particular, the normalised and centred IV estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ satisfy

$$\pi_n(\tilde{\rho}_n - \rho_n) \to_d \mathcal{MN}(0, V_1) \quad \text{and} \quad \pi_n\left(\tilde{\beta}_n - \beta\right) \to_d \mathcal{MN}(0, V_2)$$

along arbitrary sequences $(\theta_n)_{n\in\mathbb{N}}$ in $\Theta$ and $\left(\bar{\theta}_n\right)_{n\in\mathbb{N}}$ in $\bar{\Theta}$, i.e. along: (a) the entire spectrum of AR regressor processes, including stationary, non-stationary, explosive processes, all intermediate regimes and their oscillating counterparts; (b) the space of distribution functions of $(u_t)$ for the

AR model and the space of distribution functions of $(\varepsilon_t, u_t)$ for the PR model which allow the innovations to be non-Gaussian, non-identically distributed, conditionally heteroskedastic and, in the case of the PR model, to exhibit short memory linear autocorrelation for $(u_t)$; (c) the space of distribution functions of an $\mathcal{F}_0$-measurable initial condition. The AMG property is derived via central limit theory and does not depend on the cdf of the innovation sequences $(u_t)$ and $(\varepsilon_t)$. The only component that depends on the distribution of $(u_t)$ is the mixing variate $X_\infty$ in the explosive case $|\rho| > 1$ which does not the affect the AMG property and, upon studentisation of $\tilde{\rho}_n$, $\tilde{r}_{h,n}$ and $\tilde{\beta}_n$, is scaled out of the limit distribution of the corresponding t-statistics. This AMG property of the proposed estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ is in sharp contrast to the OLS estimators which do not have an AMG limit distribution in the local-to-unity case C(ii) and in the explosive case $|\rho| > 1$ when limit distribution theory is entirely driven by the cdf of the innovations $(u_t)$ and $(\varepsilon_t)$. The main implication of the AMG property is that, upon studentisation, inference based on $\tilde{\rho}_n$ and $\tilde{\beta}_n$ in (9) and (10) is asymptotically standard normal, uniform and distribution-free.

**2.** Theorems 2 and 3 and Corollary 1 show that the t-statistics based on the IV estimators for $\rho$, $\beta$ and $r_h$ give rise to critical regions and confidence intervals with uniform asymptotic properties over the parameter spaces $\Theta$ for the AR and LP models and $\bar{\Theta}$ for the PR model. Theorem 2 shows that the critical regions $\mathcal{R}_n^{EW}$ and $\mathcal{R}_n$ have correct asymptotic size uniformly over $\Theta$ and $\Theta^{\text{hom}}$ respectively and that the corresponding confidence intervals $I_n^{EW}(\tilde{\rho}_n, \alpha)$ and $I_n(\tilde{\rho}_n, \alpha)$ have correct asymptotic coverage uniformly over $\Theta$ and $\Theta^{\text{hom}}$. In the terminology of Andrews, Cheng and Guggenberger (2020), $\mathcal{R}_n$ and $I_n(\tilde{\rho}_n, \alpha)$ are uniformly asymptotically similar over $\Theta^{\text{hom}}$ and $\mathcal{R}_n^{EW}$ and $I_n^{EW}(\tilde{\rho}_n, \alpha)$ are uniformly asymptotically similar over $\Theta$. This is the first procedure that provides a CR and CI with uniform asymptotic size and coverage rate for an arbitrary autoregressive root. Theorem 3 and Corollary 1 establish the corresponding uniform inference results for the PR and LP models respectively. The uniformity of Theorems 2 and 3 and Corollary 1 extends over the space of marginal cdfs $(F_t)$ and $(\bar{F}_t)$ of the innovations $(u_t)$ and $(\varepsilon_t)$ and the initial condition $X_0$. To our knowledge, our inferential framework constitutes the first procedure that achieves autoregressive inference with general asymptotic validity and, at the same time, provides the first solution to the long-standing problem of lack of distribution-free inference in the purely explosive region.

**3.** Theorems 2 and 3 provide, as an immediate corollary, the first proof of the uniformity property of the IVX method of Phillips and Magdalinos (2009) and Kostakis et al. (2015) over a parameter space consisting of the restrictions $\bar{\Theta}\big|\,\Theta_\rho = [-1+\delta, 1]$ and $\bar{\Theta}^{\text{hom}}\big|\,\Theta_\rho = [-1+\delta, 1]$ for some $\delta > 0$; it also shows that with an Eicker-White standard error adjustment, this uniformity property may be extended over the restriction of $\Theta$ when $\Theta_\rho = [-1+\delta, 1]$. Additionally, Theorems

2 and 3 establish the uniformity of the procedures of the current paper over the entire parameter AR space and are the first to accommodate this level of generality in AR and PR inference. It is worth noting that the uniformity properties of leading procedures in PR analysis, e.g. Jansson and Moreira (2006) and Elliott et al. (2015), are unknown even over the restricted parameter space $\Theta_\rho = [-1+\delta, 1]$. Finally, the uniformity of the CRs and CIs of Theorem 3 over the space $\Theta_{\varepsilon u}$ of cdfs of $(\varepsilon_t, u_t)$ implies their uniformity over $corr\,(\varepsilon_t, u_t)$, a nuisance parameter that is well-documented to affect the finite sample performance of all known PR model tests.

**4.** The key element of the procedure that delivers the AMG property and the distributional invariance to the innovations across the AR classes C(i)-C(iii) is the newly proposed combined instrument $\tilde{z}_t$ in (6)-(8). The instrumentation of the procedure employs information from a non-AMG procedure (OLS) to construct estimators that enjoy the AMG property across all AR classes. In particular, our instrument employs information from the OLS estimator (through the events $F_n$ in (5)) to distinguish between the C(i) and C(iii) autoregressive classes, i.e. to separate the cases $c = -\infty$ and $c = \infty$ in Assumption 7. Lemma 2 in the Appendix shows that this asymptotic separation is achieved at arbitrary rate with probability tending to 1. When $c = -\infty$, $\tilde{z}_{n,t}$ takes the form of a moderately stationary instrument which is: (i) regular $\tilde{z}_{1t} := \tilde{z}_{n,t}\mathbf{1}_{F_n^+}$ when $\rho = 1$; (ii) oscillating $\tilde{z}_{1t}^- := \tilde{z}_{n,t}\mathbf{1}_{F_n^-}$ when $\rho = -1$; (iii) $\tilde{z}_{1t}$ or $\tilde{z}_{1t}^-$ when $\rho \in (-1,1)$, in which case, the IV estimators $\tilde{\rho}_n$ and $\tilde{\beta}_n$ based on either are asymptotically equivalent to the (asymptotically normal) OLS estimator. When $c = \infty$, $\tilde{z}_{n,t}$ takes the form of: (i) a mildly explosive instrument $\tilde{z}_{2t} := \tilde{z}_{n,t}\mathbf{1}_{\bar{F}_n^+}$ when $\rho \geq 1$ and (ii) an oscillating mildly explosive instrument $\tilde{z}_{2t}^- := \tilde{z}_{n,t}\mathbf{1}_{\bar{F}_n^-}$ when $\rho \leq -1$. Finally, when $c \in \mathbb{R}$, the autoregression is of the local-to-unity type C(ii), in which case, $\tilde{z}_{n,t}$ takes the form of: a random linear combination of $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ when $\rho = 1$ and of $\tilde{z}_{1t}^-$ and $\tilde{z}_{2t}^-$ when $\rho = -1$. This random combination, reflected in the random normalisation $\pi_n$ of Theorem 1(i), depends on the limit distribution of the OLS estimator $\hat{\rho}_n$ through the events $F_n^+$ and $F_n^-$ in (5) which, like the limit distribution of $\pi_n^{-2}\sum_{t=1}^n \underline{x}_{n,t-1}\tilde{z}_{n,t-1}$ in the denominator of (9)-(10), can be expressed as non-stochastic functionals of the Brownian motion $B(\cdot)$ arising from the FCLT on $(u_{n,t})$; the asymptotic independence between $\pi_n^{-1}\sum_{t=1}^n \tilde{z}_{n,t-1}u_{n,t}$ and $B(\cdot)$, established by Lemma 5 in the Appendix, implies that the additional randomness introduced by the combination of $\tilde{z}_{1t}$ and $\tilde{z}_{2t}$ (and $\tilde{z}_{1t}^-$ and $\tilde{z}_{2t}^-$ when $\rho < 0$) does not affect the AMG property of $\tilde{\rho}_n$ and $\tilde{\beta}_n$.

**5.** It is worth providing a brief overview of how distribution-free asymptotic inference is achieved in the explosive case $|\rho_n| \to |\rho| > 1$. By employing the OLS residuals $\hat{u}_t$ and a (regular or oscillating) mildly explosive root $\varphi_{2n}$ or $\varphi_{2n}^-$ for the construction of the instrument $\tilde{z}_{2t}$, the instrumentation of this paper and Lemmata 2 and 4 ensure that, under C(iii), the limit distribution of $\tilde{\rho}_n$ is driven by the mildly explosive component ($z_{2t}$ in (A.5) or its oscillating counterpart $z_{2t}^-$)

and inherits the desirable AMG property of mildly explosive martingale transforms even when $x_t$ in (1) is a purely explosive process. The price paid for this asymptotic distributional invariance is a reduction in the convergence rate of our IV estimators by an order of $\left(\phi_{2n}^2 - 1\right)^{-1/2}$ compared to the $|\rho_n|^n$-OLS rate. Given that the above order satisfies $o\left(n^{1/2}\right)$ and that the exponential part $|\rho_n|^n$ of the OLS rate is maintained in the convergence rate of Theorem 1(iii), the efficiency loss associated with employing the IV estimators $\tilde{\rho}_n$, $\tilde{r}_{h,n}$ and $\tilde{\beta}_n$ is small compared to the benefit from an estimation procedure that gives rise to test statistics and CIs of general asymptotic validity. In the case when $|\rho_n| \to 1$, the limit distributions $Y/X$ and $\frac{\omega}{\sigma_\varepsilon}\tilde{Y}/X$ are Cauchy.

**6.** While $T_n^*$ and $\bar{T}_n$ have the same limit distribution, the test based on $\bar{T}_n$ may suffer from finite sample distortion due to the fact that the estimation of the intercept $\gamma$ in (2) does not feature in the first-order asymptotic theory. As documented by Kostakis et al. (2015), Hosseinkouchack and Demetrescu (2021) and Harvey, Leybourne and Taylor (2021), this becomes an issue under $C_+$(ii) where estimation of the intercept 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 1. While this contribution is $o_p(1)$, $n\pi_n^{-1}\bar{z}_{1n-1}\bar{\varepsilon}_n = O_p(n^{-1/2}(1-\varphi_{1n})^{-1/2})$ under $C_+$(ii), $n\bar{z}_{1n-1}\bar{\varepsilon}_n$ is asymptotically equivalent to $(1-\varphi_{1n})^{-1} x_{n,n}\sum_{t=1}^n \varepsilon_{n,t}$ and the correlation between $x_{n,n}$ and $\sum_{t=1}^n \varepsilon_{n,t}$ distorts mixed-Gaussianity in finite samples. As a result, the t-statistic based on $\bar{T}_n$ exhibits finite sample distortions when the following occur *jointly*: (i) the AR root of $x_t$ is very close to 1; (ii) $\delta_{\varepsilon u} = corr(\varepsilon_t, u_t)$ is close to 1 in absolute value; (iii) $\varphi_{1n}$ is chosen close to 1. The FM transformation of Phillips and Hansen (1990), $\varepsilon_{0t} = \varepsilon_{n,t} - \omega^{-1}\mathbb{E}(\varepsilon_{n,t}u_{n,t})u_{n,t}$, orthogonalises $n^{-1/2}\sum_{t=1}^n \varepsilon_{0t}$ and $n^{-1/2}\sum_{t=1}^n u_{n,t}$ asymptotically when $x_{n,t}$ 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 $\beta_n^*$ arising from employing the FM transformation and the corresponding t-statistic $T_n^*$ have significantly improved finite sample properties whenever $\rho_n$ is close to 1 with large $|\delta_{\varepsilon u}|$, while both $\bar{T}_n$ and $T_n^*$ perform equally well in all other cases. It is worth noting that the terms arising from the estimation of the intercept in both IV and OLS, $n\bar{z}_{1n-1}\bar{\varepsilon}_n$ and $n\bar{x}_{n-1}\bar{\varepsilon}_n$, have reduced order of magnitude for an autoregressive root close to $-1$, so no finite sample adjustment is necessary under $C_-$(ii).

**7.** Practical implementation of the test procedures of Theorems 1 and 2 and Corollary 1 requires a choice for $\varphi_{1n}$, $\varphi_{1n}^-$, $\varphi_{2n}$ and $\varphi_{2n}^-$ in (6) for the construction of the instrument $\tilde{z}_{n,t}$. We require a single instrument that will perform well across C(i)-C(iii) for all AR, PR and LP models. We base our choice for $\varphi_{1n}$ and $\varphi_{2n}$ on the principle of controlling the worst finite sample distortion that occurs in the case of a unit root regressor with large correlation $|\delta_{\varepsilon u}|$ (Remark 6 above). We conduct a grid search Monte Carlo to select the maximal values of $\varphi_{1n}$ and $\varphi_{2n}$ (by Theorem 1, these achieve maximal power) subject to a satisfactory test size in the above least favourable case and we set $\varphi_{1n}^{-} = -\varphi_{1n}$ and $\varphi_{2n}^{-} = -\varphi_{2n}$; a detailed analysis of the choice of $\varphi_{1n}$ and $\varphi_{2n}$ can be found in Section 4.1. We demonstrate that the proposed choice of instrument in Section 4.1 works very well (in terms of size and power) in all three regression setups and across all AR regions.

**8.** The above methodology may be extended to multivariate AR/PR/LP models where both $x_t$ and $y_t$ in (1) and (2) are vector-valued and the statistical problem consists of testing a set of $q$ restrictions on the parameters. A model along the lines of Magdalinos and Phillips (2020) (that assumes away cointegrating relationships within the VAR(1) process $x_t$) extended to account for simple real eigenvalues for the AR matrix. The asymptotically $\mathcal{N}(0,1)$ t-statistics of Theorem 1 replaced by asymptotically $\chi^2$ Wald statistics based on the combined (vector-valued) instrument (6)-(8) of Section 2.2. The fact that the methodology of this paper extends directly to multivariate systems is a major advantage over existing methods, including Campbell and Yogo (2006) and Elliott et al. (2015). A multivariate extension is not pursued here as it would be a deviation from the main focus of the paper (the construction of CIs for $\rho$, $\beta$ and $r_h$ with uniform asymptotic validity). An additional advantage of our inference procedures in Theorems 2 and 3 and Corollary 1 is their simplicity and ease of implementation: they employ closed-form linear estimators and $\mathcal{N}(0,1)$ critical values, rendering implementation of the procedures by practitioners natural and straightforward.

# 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 AR 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 regions: 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 AR and PR models. On the other hand, we demonstrate that the IV procedure of this paper continues to provide correct inference. Next, we compare the finite sample properties of our procedure to leading existing approaches: in Section 4.3.1, we provide a comparison of our CIs in (21) for the AR parameter to Andrews and Guggenburger (2014) procedure; in Section 4.3.2, we compare the size and power of our testing procedure in (23) in the PR setup to the procedure proposed by

Elliott et al. (2015) and in Section 4.3.2 we compare our CIs based on (18) to the procedure in Olea and Plogborg-Moller (2021) in the LP model. In the first two cases, we demonstrate that the IV procedure delivers: (i) correct size across all autoregressive regimes considered, and (ii) superior power in all cases of roots in $[-1, 1]$ (including local-to-unity, near- and purely stationary regions) except for the case of exact unit root, where the differences in power are very small and vanish as the sample size increases. In the LP case, our procedure is actually more powerful than the lag-augmentation approach around unity. Crucially, our procedure also provides correct inference on the right side of unity and on the left side of $-1$, in the local-to-unity, mildly and purely explosive regions, where no existing alternative approach has general asymptotic validity.

## 4.1 Practical implementation and instrument selection

Practical implementation of our procedure requires a choice for $\varphi_{1n}$, $\varphi_{1n}^{-}$, $\varphi_{2n}$ and $\varphi_{2n}^{-}$ in (6) for the instrument construction in (8). While theoretically, any values of $\varphi_{1n} \to 1$ in $\mathrm{C}_1$(i), $\varphi_{1n}^{-} \to -1$ in $\mathrm{C}_{-1}$(i), $\varphi_{2n} \to 1$ in $\mathrm{C}_1$(iii) and $\varphi_{2n}^{-} \to -1$ in $\mathrm{C}_{-1}$(iii) deliver correct asymptotic inference, finite sample performance may vary considerably with the choice for particular values. For simplicity and symmetry, we set $\varphi_{1n}^{-} = -\varphi_{1n}$ and $\varphi_{2n}^{-} = -\varphi_{2n}$. Choosing

$$\varphi_{1n} = 1 - 1/n^{b_1}, \;\; \varphi_{2n} = 1 + 1/n^{b_2} \tag{28}$$

reduces the problem to selecting values for $b_1$ and $b_2$ in $(0, 1)$. We adopt a conservative approach: (i) Remark 13 of Section 3.3 indicates that inference based on $T_n^*$ suffers the worst finite sample distortion in the predictive regression case when $\rho_n = 1$ with large correlation $\rho_{\varepsilon u}$ between the innovations $\varepsilon_t$ and $u_t$ in (1) and (2)[15]; (ii) Theorem 1 shows that the power of the t-tests $\bar{T}_n$ and $T_n^*$ is always increasing with $b_1$ and is increasing with $b_2$ in the regions C(i)-C(ii) (in C(iii) the exponential rate $\rho_n^n$ in $\pi_n$ is independent of $b_2$, so the choice of $b_2$ has only a minor effect on power). Given (i) and (ii), we base our selection of $b_1$ and $b_2$ on selecting the maximal values of $b_1$ and $b_2$ for which the size in the worst case scenario (i) is controlled. This amounts to a two-dimensional grid search problem outlined in detail in the Appendix. Imposing a 5.99% threshold on the empirical size for a nominal 5% size for these most unfavourable cases yields the following selection in (28): $b_1 = 0.85$ and $b_2 = 0.7$. We recommend this choice for the implementation of Algorithm 1 and use it throughout the Monte Carlo section and the empirical application in Section 5. In Section 4.3. we demonstrate that this choice works well for all AR specifications in all PR, AR and LP setups.

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

—— **Algorithm 1** ——

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

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

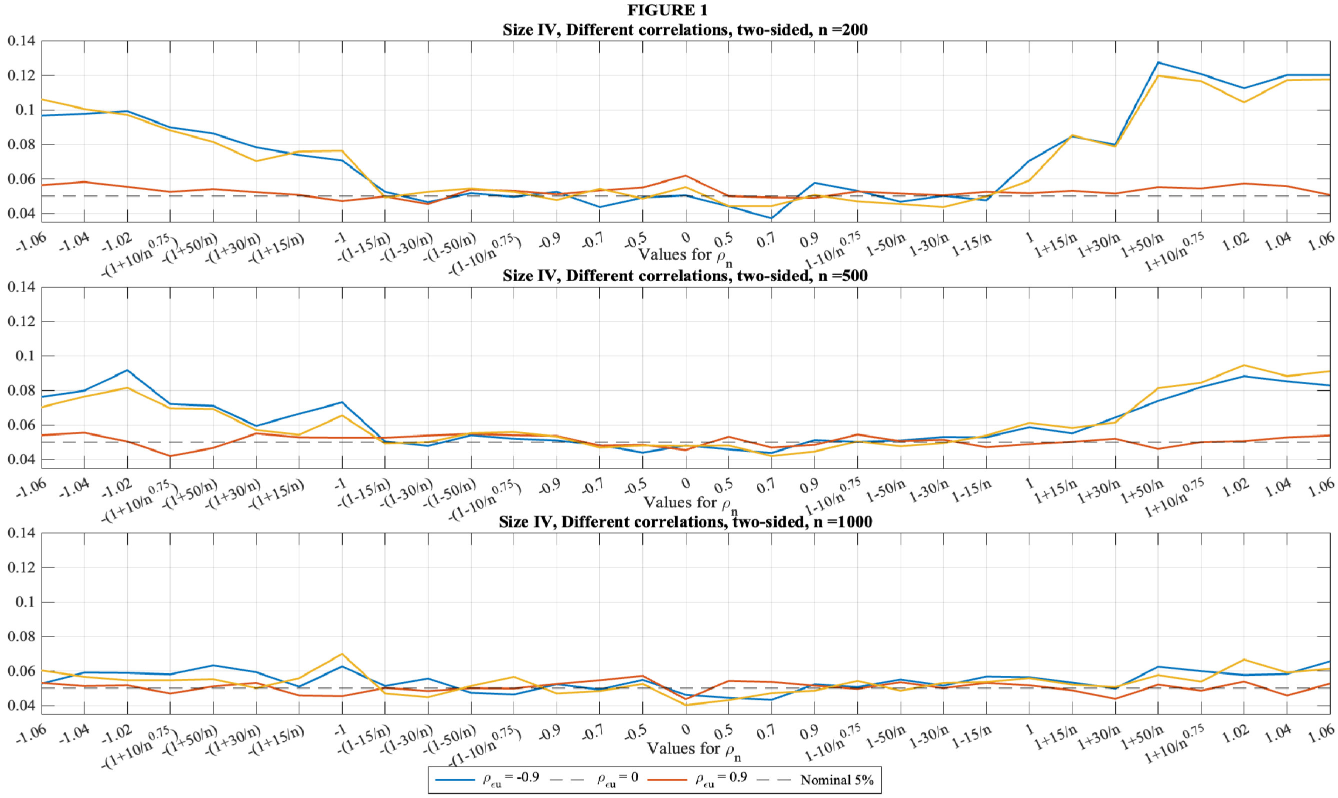
*2. Select $\varphi_{1n}$, $\varphi_{1n}^{-}$, $\varphi_{2n}$ and $\varphi_{2n}^{-}$ (e.g. from (28) with the recommended $b_1 = 0.85$ and $b_2 = 0.7$), compute $\rho_{nz}$ in (6) and build recursively the instrument $\tilde{z}_t$ in (8) initialising at $\tilde{z}_0 = 0$.*

*3. Use the constructed instrument $\tilde{z}_t$ to compute the IV estimator $\tilde{\rho}_n$ in (9) for the AR setup, the IV estimator $\tilde{r}_{h,n}$ in (11) for the LP setup, or, given a sample for $y_t$, compute the IV estimator $\beta_n^*$ in (22) for the PR setup.*

*4. Compute the IV-based t-statistic in (16) and the CI $I_n(\tilde{\rho}_n, \alpha)$ of Theorem 2 for the AR setup, IV-based t-statistic in (18) and the CI $I_n(\tilde{r}_{h,n}, \alpha)$ of Corollary 1 for the LP setup or the IV-based t-statistic in (23) and the CI $I_n(\beta_n^*, \alpha)$ of Theorem 3 for the PR setup; conduct inference using $\mathcal{N}(0,1)$ critical values.*

---

**FIGURE 1**




We first implement our choice of instrument in the PR model for AR regimes for $x_t$ in (1):

$$\rho_n \in \{\text{-1.06,-1.04,-1.02,-(1+10/n}^{0.75}\text{),-(1+50/n),-(1+30/n),-(1+15/n),-1,-(1-15/n),-(1-30/n)} \\ \text{-(1-50/n),-(1-10/n}^{0.75}\text{),-0.9,-0.7,-0.5, 0, 0.5, 0.7, 0.9, 1-10/n}^{0.75}\text{,1-50/n,1-30/n,1-15/n, 1,} \\ \text{1+15/n,1+30/n, 1+50/n, 1+10/n}^{0.75}\text{, 1.02, 1.04, 1.06}\},\ X_0 = 0,\ \mu = \mu_y = 0, \tag{29}$$

$$\varepsilon_t \sim \mathcal{N}(0,1),\ u_t \sim \mathcal{N}(0,1),\ \rho_{\varepsilon u} \in \{-0.9, -0.45, 0, 0.45, 0.9\}. \tag{30}$$

For each specification, we compute the empirical size of the 95% two-sided test statistic in (23) based on 5,000 simulated samples for sample sizes $n \in \{200, 500, 1000\}$. Throughout the entire

Monte Carlo section, we always use reduced[16] sample sizes $n \in \{100, 200, 500\}$ for the explosive specifications $\rho_n \in \pm\{1 + 50/n, 1 + 10/n^{0.75}, 1.02, 1.04, 1.06\}$. Figure 1 displays the rejection probability of our test procedure in (23) under the null $\beta = 0$ for the different AR 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 provides evidence that our procedure delivers satisfactory empirical size throughout the different persistence regions converging to the nominal 5% as the sample size increases. Section 1.3 of the Appendix contains the proportion of times each of the instruments is chosen throughout the different AR specifications. As expected, the (oscillating) mildly explosive instrument is never chosen in the stationary region C(i) even for small samples, and is chosen in the (negative) pure unit root case around 33% of the time (since the OLS distribution in this case is left-skewed with values below unity occurring with probability 2/3).

## 4.2 Invalidity of OLS in the explosive regions

In this section, we briefly discuss the relative performance of OLS and our procedure in the explosive regions $(-\infty, 1) \cup (1, \infty)$ and provide an illustration of the invalidity of OLS-based inference even in large samples. The lack of CLT for the numerator of the OLS estimators of $\rho_n$ and $\beta$ 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 AR and PR model in (1) and (2), 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 of 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 can affect the limit and invalidate inference. In Figure 2, we report the 90%, 95% and 99% coverage rates of the IV and OLS estimators of $\rho_n$ respectively for different sample sizes (as in Section 4.1, we work with the AR specifications in (29) and reduced sample sizes for the explosive processes). We compute the coverage rates as the proportion of time the true $\rho_n$ finds itself in the 90%, 95% and 99% CIs implied by the IV and OLS respectively, based on 5,000 replications. From Figure 2, it is clear that 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 for small $n$.

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

FIGURE 2

Coverage of CIs for $\rho_n$, n=200

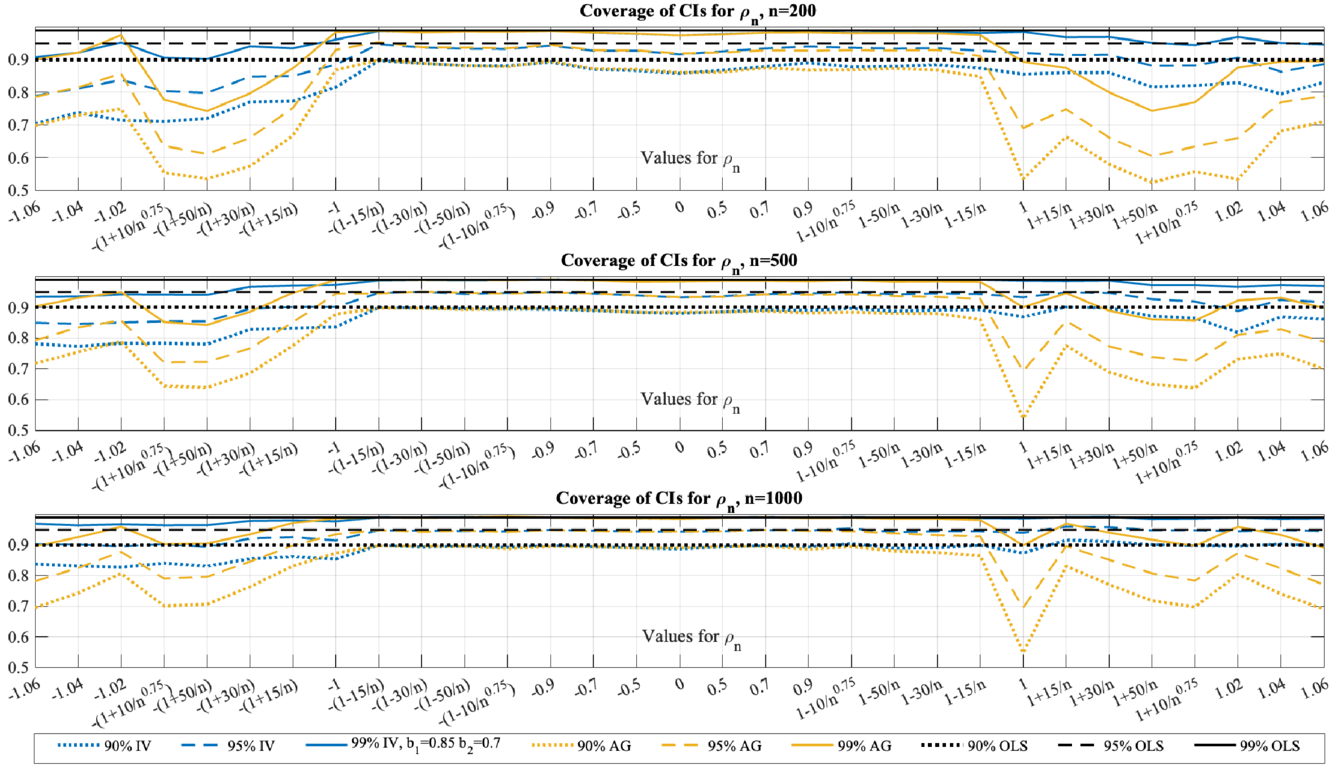
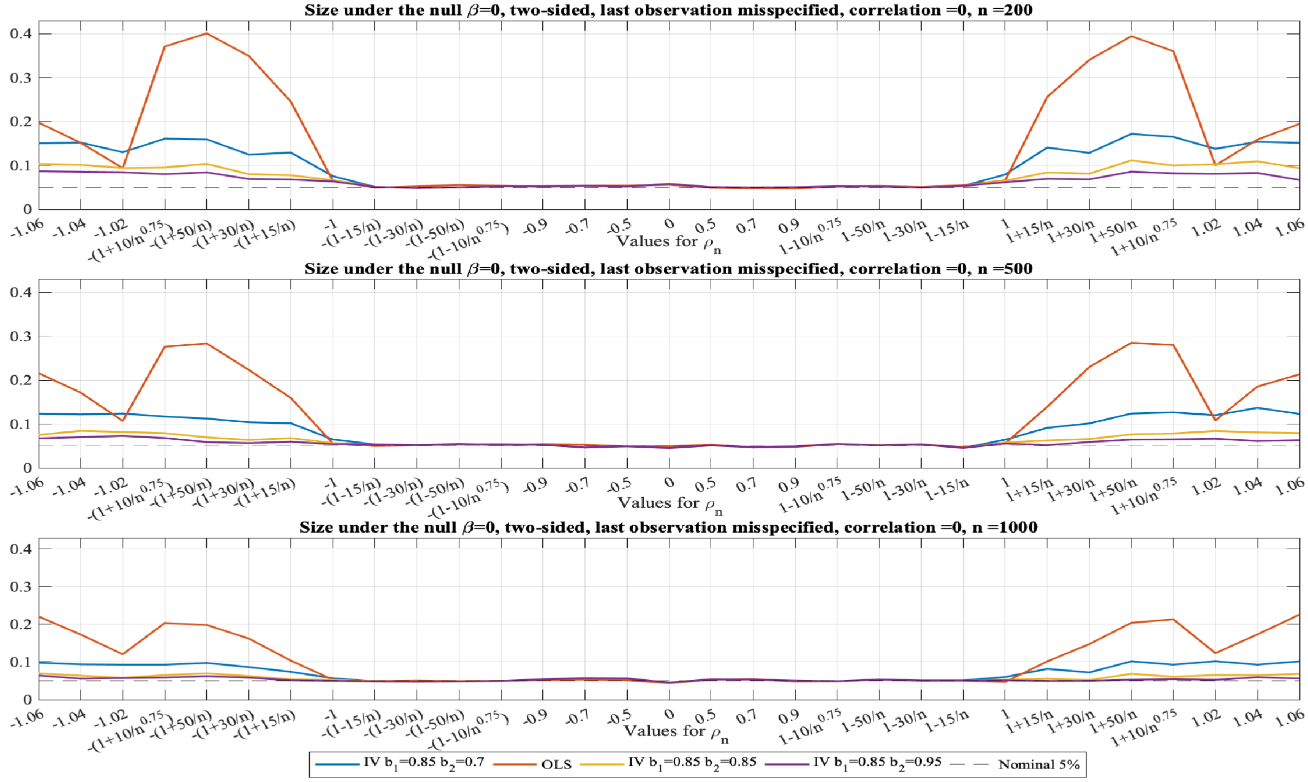



FIGURE 3

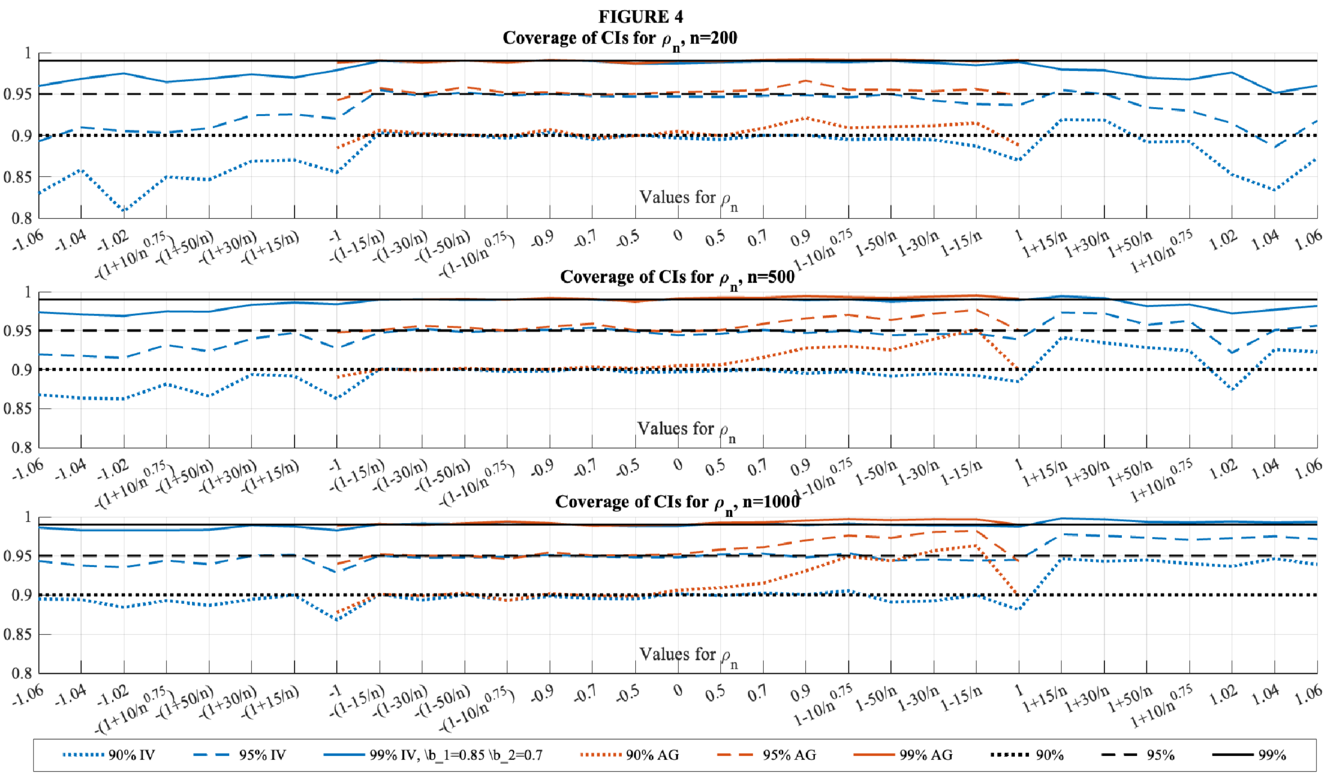
Size under the null $\beta$=0, two-sided, last observation misspecified, correlation =0, n =200




However, as $n$ increases, the coverage rates of the IV procedure converge to the nominal levels, as Theorem 2 suggests. The coverage rates of OLS for the mildly explosive specification $\rho_n = \pm(1 + 10/n^{0.75})$ also improve as expected (although very slowly). Crucially, for the purely explosive DGPs, the OLS distortions do not improve even for larger samples. For example, when $\rho_n = \pm 1.06$, the 90% OLS CI contains the truth 70% of the time irrespective of increases in the sample size. We find similar results in the PR setup. Figure 3 reports the rejection probability of the OLS under the null $\beta = 0$ against a two-sided alternative 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 $\beta_2$ to 0.85 and 0.95 respectively. From Figure 3, the empirical size of the OLS for the purely explosive regions is distorted and, crucially, the distortions deteriorate as the sample size increases; the size of our procedure on the other hand converges to the nominal size, as suggested by the theoretical results of Theorem 3.

## 4.3 Comparisons with existing methods

### 4.3.1 Inference in the AR model

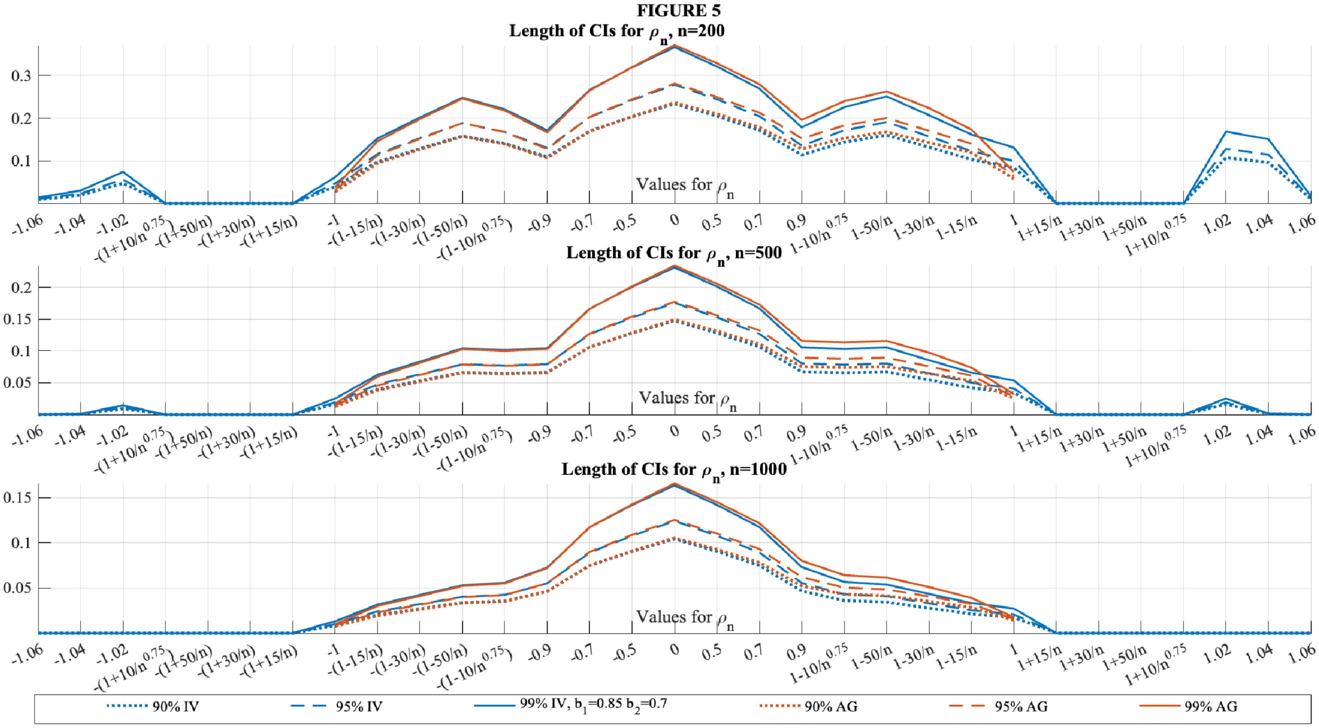



In this section, we present a comparison of our procedure to current state-of-the-art methodology in the literature of robust inference in AR, PR and LP models for $\rho \in (-1, 1]$. We first evaluate our proposed AR CIs in (21) and we compare them to the procedure by Andrews and Guggenberger (2014)[17] (henceforth AG), which constructs the intervals by inverting the OLS t-statistic, which under the null is asymptotically nuisance-parameter-free. In Figures 4 and 5, we report the coverage rates and lengths of the 90%, 95% and 99% CIs respectively for the IV estimator and AG

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

procedure for $\rho_n$ for different AR 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. Figure 4 presents evidence that our IV-based CIs are comparable to the CIs based on the AG procedure in $[-1, 1]$, while also providing correct coverage for $\rho_n$ in $(-\infty, -1] \cup [1, \infty)$ in the local-to-unity, mildly and purely explosive regions. In terms of interval length, Figure 5 shows that our intervals are shorter[18] than those of AG (translating to higher power) for all specifications except for the exact (positive and negative) unit root case $|\rho| = 1$; the differences in length when $|\rho| = 1$ are not large and become negligible for large samples.

FIGURE 5




### 4.3.2 Size and power comparison in the PR model

Next we generate data from the PR in (2) for the specifications of (29) and (30) in order to evaluate the performance of the IV-based t-statistic in (23), by comparing it to the one-sided test procedure by Elliott et al. (2015)[19], which, in the presence of a nuisance parameter, is nearly-optimal when the innovations of the model are Gaussian[20]. We found that in the one-sided test setup, our choice of instrument works well in all but one scenaria: 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

[18] This result also holds for the equal-tailed two-sided intervals of AG.

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

[20] 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.

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.

FIGURE 6-7

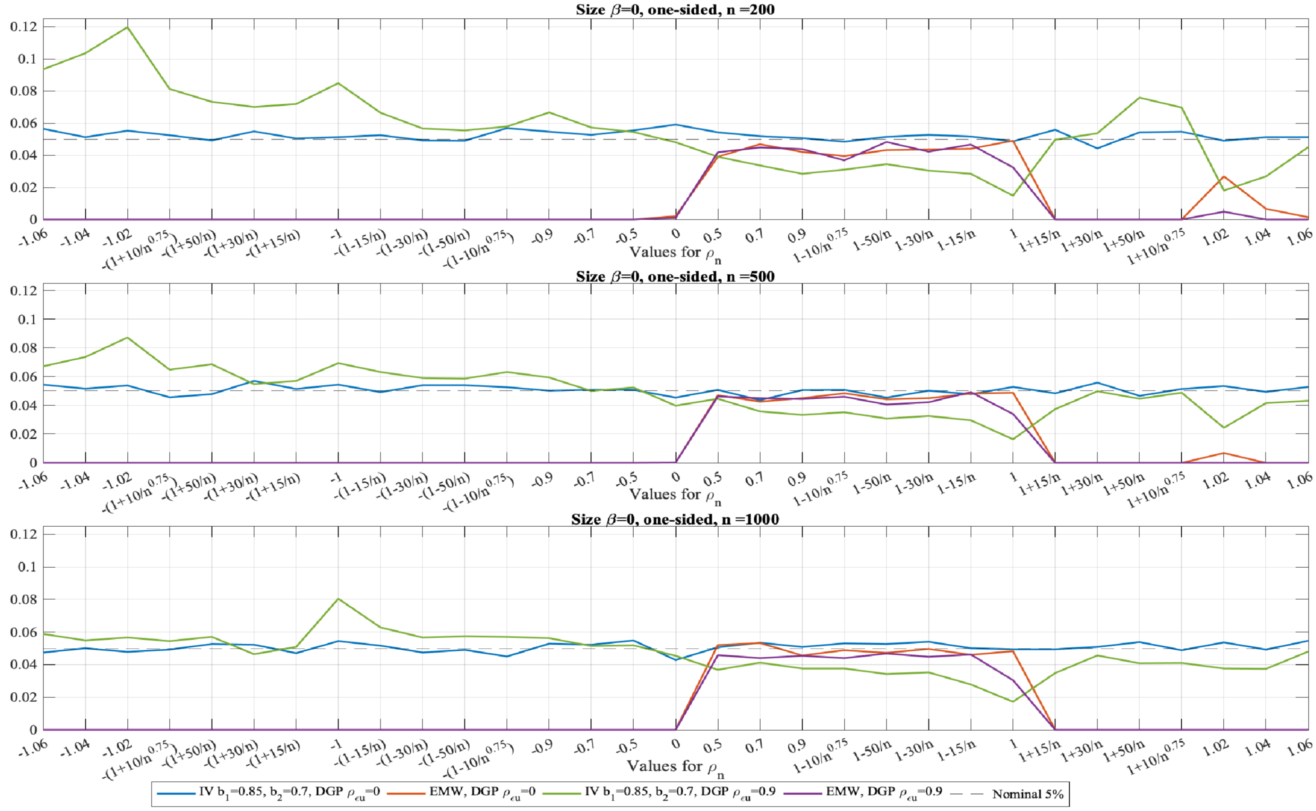



FIGURE 8

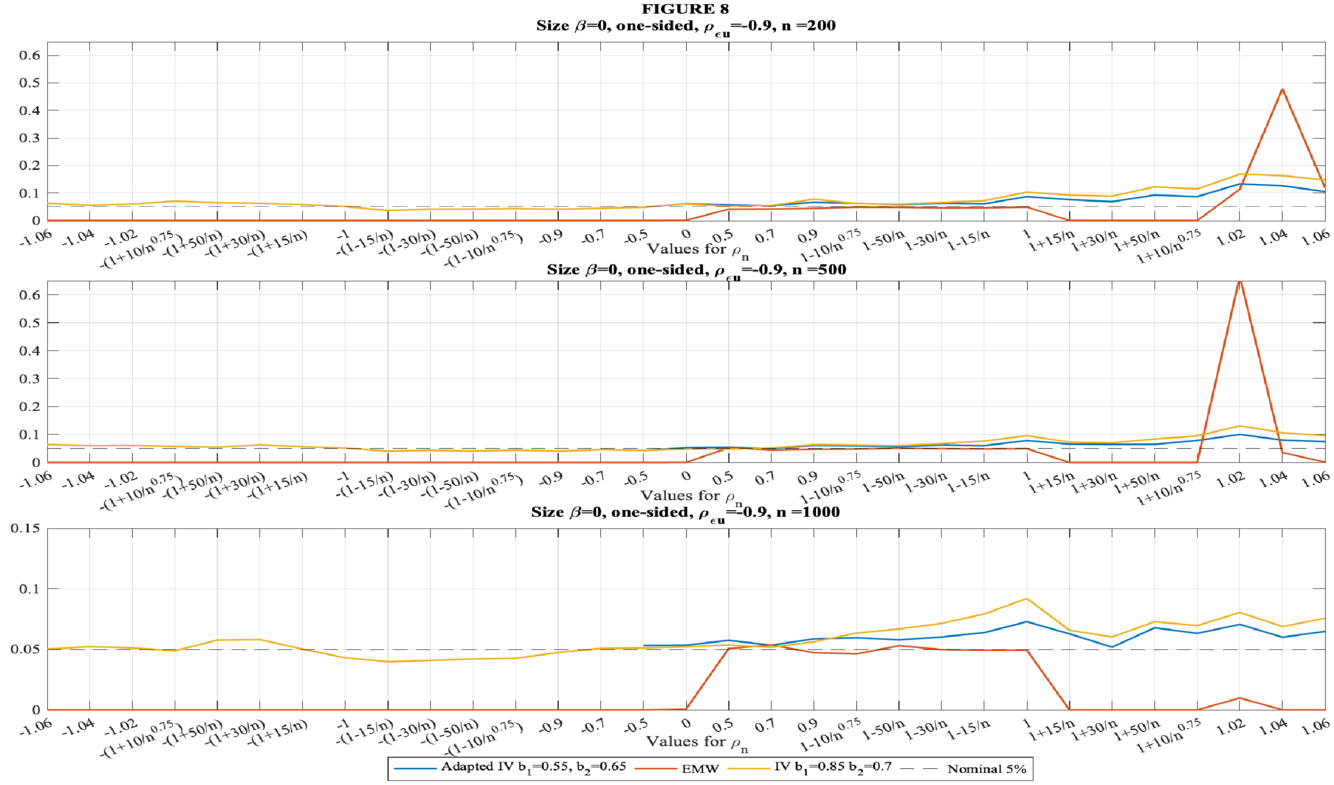


FIGURE 9

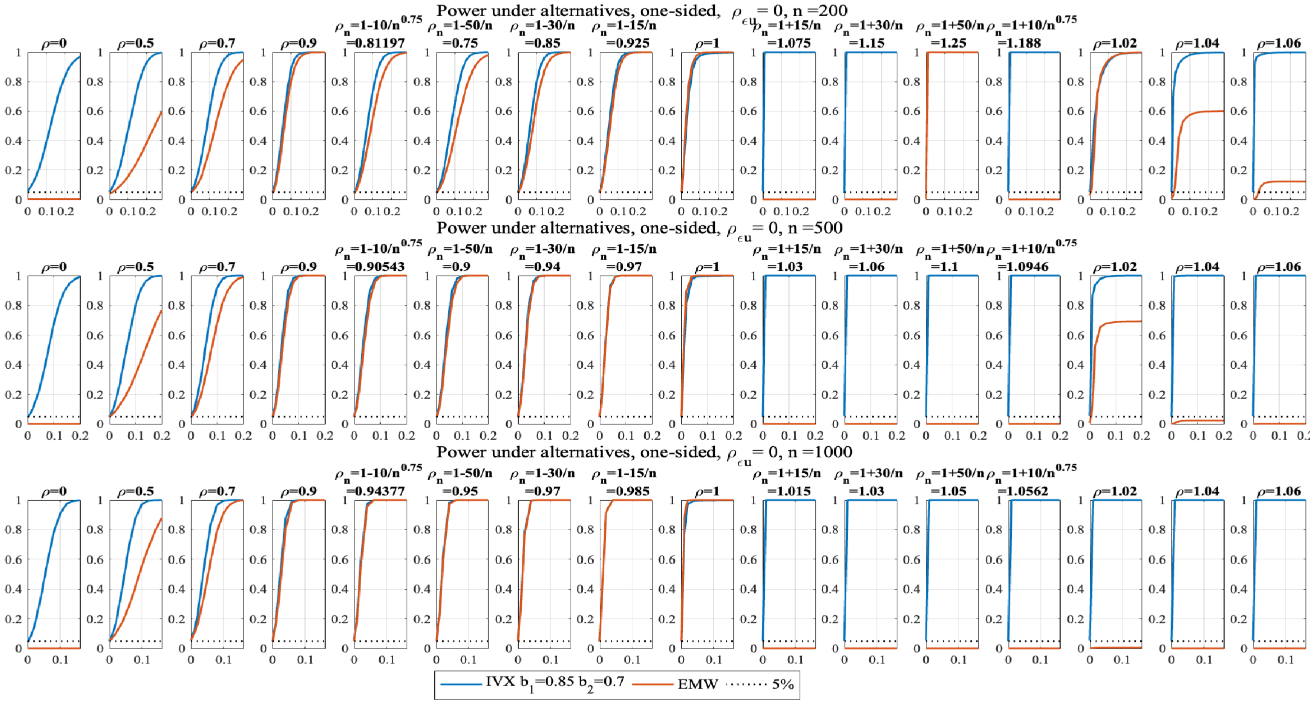
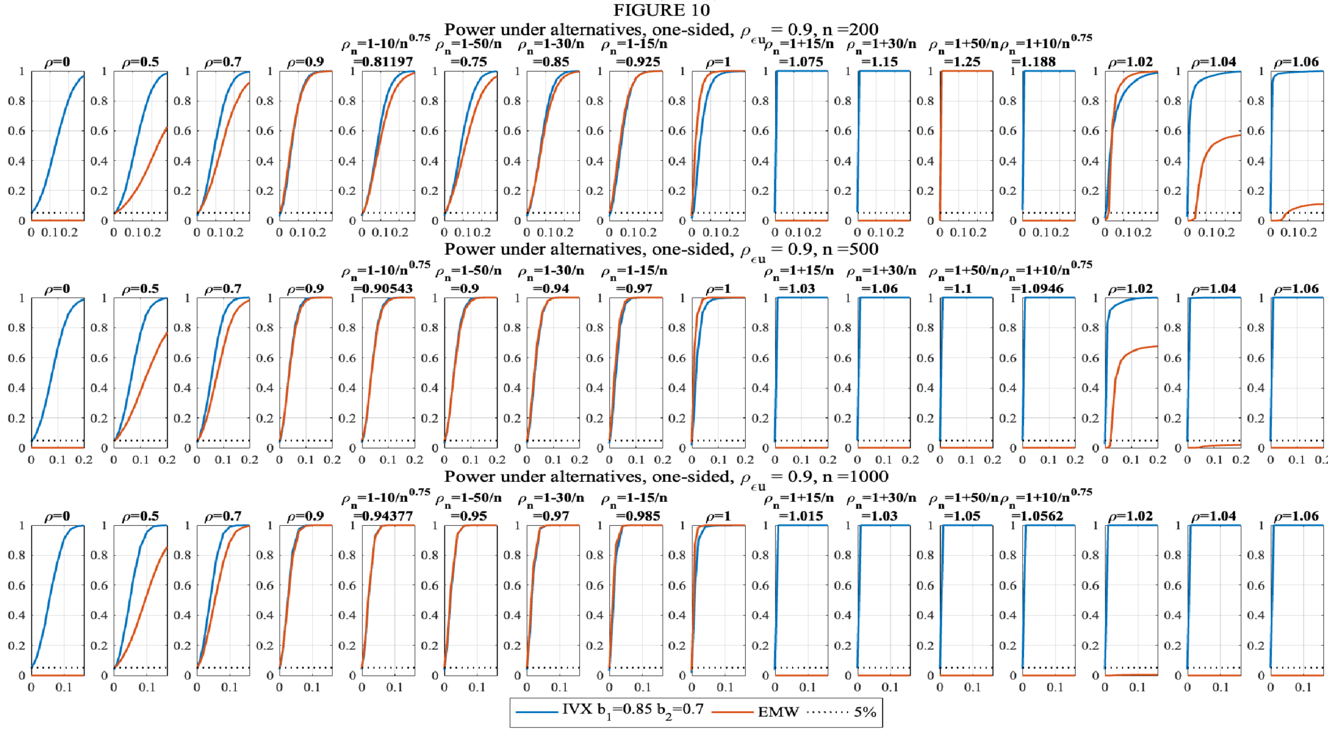



FIGURE 10

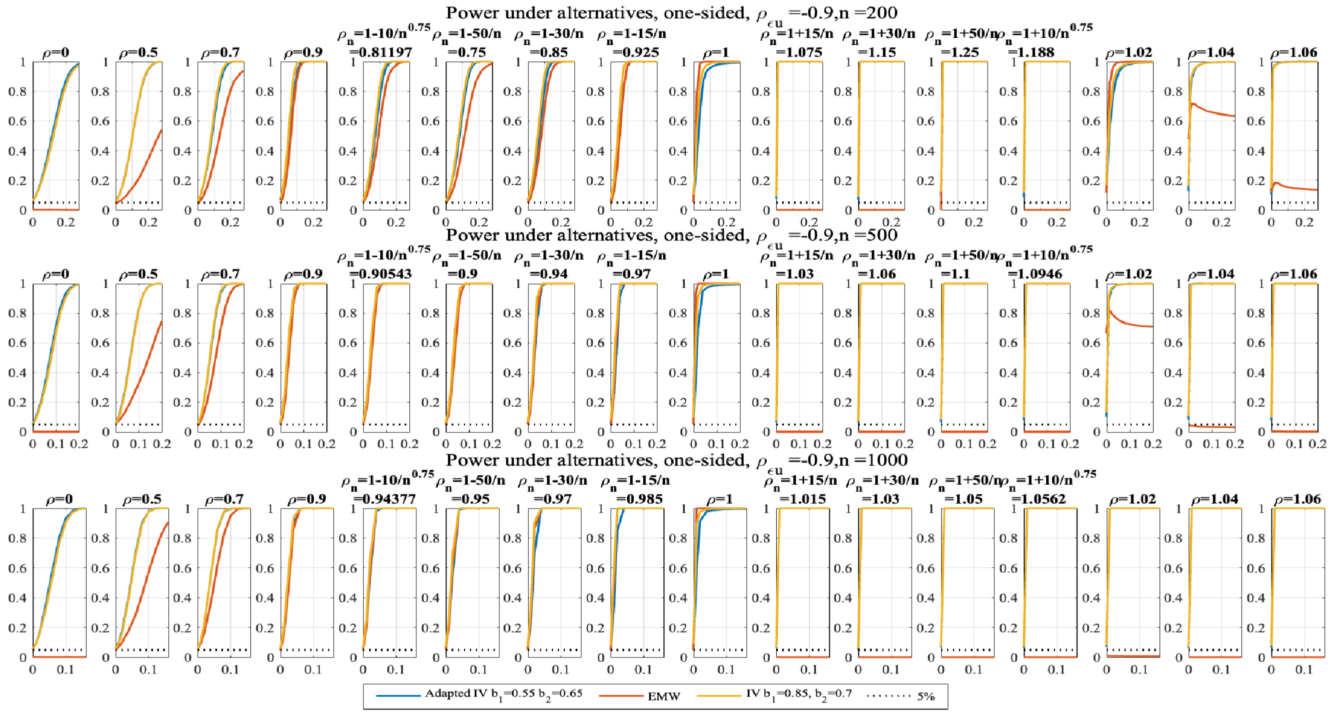



Instead, we propose using the following adaptive t-statistic:

$$T_n^A\left(\beta_n\right) = \mathbf{1}_{\{\hat{\rho}_{\varepsilon u} \leq L\} \cap \{\hat{\rho}_n \geq 0\}} T_n\left(\breve{z}_{1,t}\right) + \mathbf{1}_{\{\hat{\rho}_{\varepsilon u} > L\} \cup \{\hat{\rho}_n < 0\}} T_n\left(\breve{z}_{2,t}\right) \tag{31}$$

where $T_n\left(\breve{z}_{1,t}\right)$ and $T_n\left(\breve{z}_{2,t}\right)$ are the t-statistics in (23) based on two different choices for instruments $\breve{z}_{1,t}$ and $\breve{z}_{2,t}$, $\hat{\rho}_{\varepsilon u}$ is the sample correlation coefficient between the OLS residuals $\hat{u}_t$ and $\hat{\varepsilon}_t$, $\hat{\rho}_n$ is the OLS estimator for $\rho_n$ and $L$ is a threshold level below which a more conservative instrument selection is triggered. In this way, we can resolve the size distortion in the positive unit root case under strong negative correlation, without affecting power in all other cases.

FIGURE 11



We set $L = -0.7$ and use the values $b_1 = 0.55$ and $b_2 = 0.65$ in (28) for the construction of the conservative instrument $\breve{z}_{1,t}$. For $\breve{z}_{2,t}$ we continue to use the choice of instrument from Section 4.1 with $b_1 = 0.85$ and $b_2 = 0.7$. In the case of $\rho_{\varepsilon u} = -0.9$ in Figure 8, 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 (31) to illustrate the effect of using the adaptive procedure. For all other cases, in Figure 6-7, we display the rejection probability under the null based on the adaptive instrument which is identical to the original choice of instrument in Section 4.1 since the sample correlation coefficient $\hat{\rho}_{\varepsilon u}$ always exceeds the threshold -0.7. Figures 9-11 present the corresponding power curves [21]. We apply the procedure by Elliott et al. (2015) (EMW) in all regions for comparison, noting that their procedure is not designed to work outside $(-1, 1]$. There are several conclusions from the size and power comparisons in Figures 6-11. First, our adaptive procedure in (31) performs well in terms of empirical size in all correlation cases and in all persistence regions for the regressor and, as the sample size increases, any small sample distortions vanish. Second, we find that the EMW procedure never rejects the null to the right of unity (when the null is true and when it is not), except for a few cases with a small sample; for example in the -0.9 correlation case, its size reaches 40% in the case of $\rho_n = 1.02$ when $n = 100$, but the oversizing disappears as $n$ increases. Surprisingly, we find that the EMW procedure never rejects the null (even under the alternative) for stationary specifications with AR

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

roots in $(-1, 0]$, where it is expected to be valid, since, to our knowledge, it is supposed to switch to OLS. For this reason, in Figures 9-11, we only present power comparison for the cases $\rho_n > 0$, since EMW has zero power for any alternative for all cases $\rho_n \leq 0$. For the regions $\rho_n \leq 0$, our IV procedure has power curves that are a near mirror image of the corresponding non-oscillating cases $\rho_n > 0$. For the cases $\rho_n \in (0, 1]$, we reach a similar conclusion to the one in Section 4.3.1 in terms of power: namely, our procedure is always more powerful than EMW in all AR specifications 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 provides correct inference extending to all cases $\rho_n \leq 0$, including seasonal (near) unit root and oscillating (mildly) explosive processes as well as to the right of unity (the right-side of local-to-unity, mildly explosive and pure explosive regions) for which no alternative approaches are valid.

#### 4.3.3 Inference in the LP model

Next, we evaluate our proposed CIs for the impulse response parameter $r_h$ in the LP model in (3) and compare them to the lag-augmentation (henceforth LA) procedure by Olea and Plagborg-Moller (2021), which constructs the intervals by augmenting the LP model with an additional lag. In Figures 12 and 13, we report the coverage rates and lengths of the 95% CIs for horizons 1, 10 and 30 respectively for the IV estimator and LA procedure for $r_h$ for different AR regions and for different sample sizes. For the LA procedure, we use the Hall bootstrapped intervals[22]. Figure 12 presents evidence that our IV-based CIs have correct coverage and compare favourably to the CIs based on the LA procedure in $[-1, 1]$, while also providing correct coverage for $\rho_n$ in $(-\infty, -1] \cup [1, \infty)$ in the local-to-unity, mildly and purely explosive regions.

In terms of interval length, Figure 13 shows that: (i) for purely stationary DGPs, our intervals are comparable to the LA (any advantages of LA over our procedure, which is equivalent to the efficient OLS in this case, are due to small sample bootstrap improvements over asymptotic CIs); (ii) close to positive and negative unit roots, our procedure is considerably more powerful, which is expected since, for fixed horizon, the LA procedure rate of convergence at unit root is $\sqrt{n}$, instead of $n^{0.925}$ w.p. 2/3 and $n^{0.85}$ w.p. 1/3 implied by our choice of instrument; (iii) outside $[-1, 1]$, for DGPs which include not only explosive processes but also local-to-unity process, the LA procedure fails completely both in terms of coverage and length of the CIs. In summary, we find that our IV procedure is superior to the LA approach for any type of nonstationarity and horizon considered.

[22] We found that the t-percentile Hall bootstrapped CI, suggested by Olea and Plagborg-Moller (2021), give rise to negative variances and hence complex st. errors for oscillating DGPs.

FIGURE 12

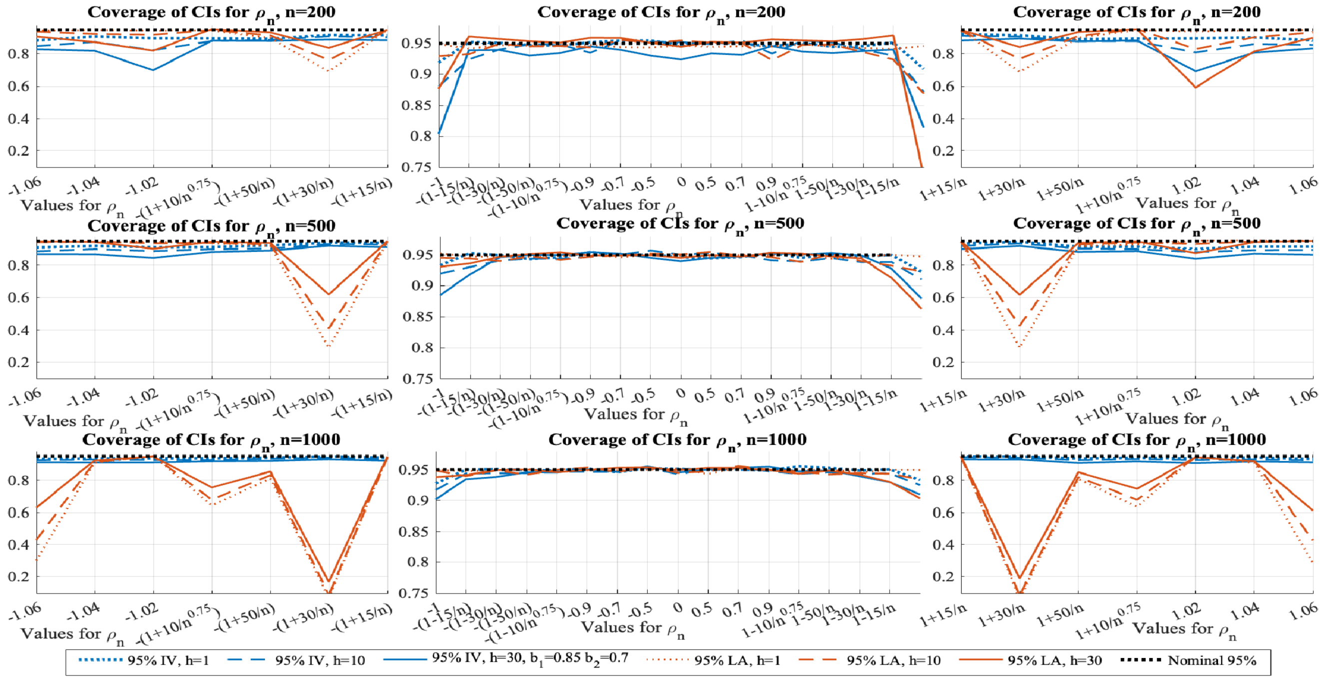
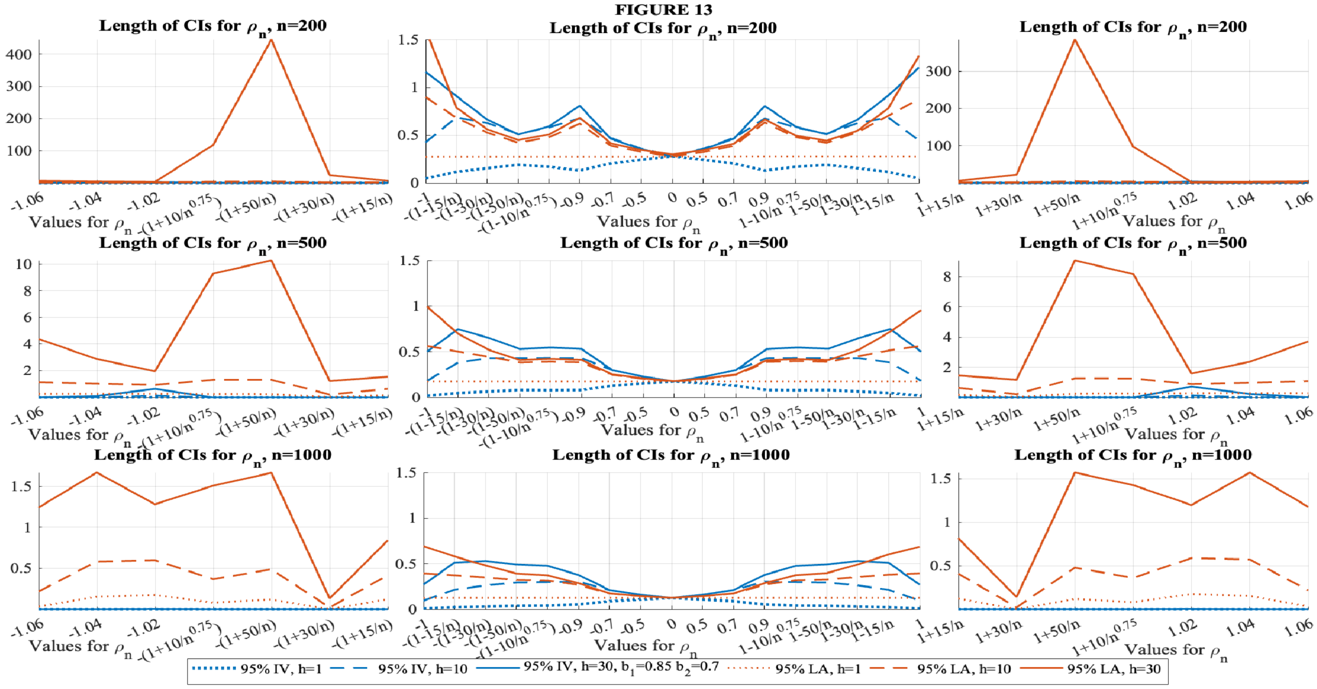



FIGURE 13




# 5 Empirical Applications

## 5.1 Inference in an epidemiological model of infection growth

In this section, we estimate a discrete-time susceptible-infected-removed (SIR) model on Covid-19 data; we briefly describe the model's equations below. The number of *infected*, *susceptible*, *recovered* and *deceased* at time $t$, denoted by $I_t$, $S_t$, $R_t$ and $D_t$ respectively, evolves according to a non-linear system of difference equations:

$$I_{t+1} = I_t\left(1 + \theta S_t/N - \gamma - \delta\right),\ S_{t+1} = S_t\left(1 - \theta I_t/N\right),\ R_{t+1} = R_t + \gamma I_t,\ D_{t+1} = D_t + \delta I_t \quad (32)$$

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. Since $S_t$ is a linear combination of the remaining states, it can be substituted out[23] in the equation for $I_{t+1}$. The model's parameters $\theta, \gamma, \delta \in (0, 1]$ are defined as follows: $\theta$ is the contact rate, the average number of individuals an infected person passes the infection in a period; $\gamma$ is the recovery rate and $\delta$ is the death rate. The model's dynamic behaviour is driven by the basic reproduction number (BRN) which in the model (32) is given by $r_0 = \theta/\left(\gamma + \delta\right)$ measuring the number of infections per infected individual, with $r_0 \geq 1$ implying the disease escalates into an epidemic and $r_0 < 1$ implying the infections' growth can be contained. The BRN $r_0$ is the key parameter for understanding the transmission mechanism of an epidemic. We use next generation matrix (NGM) approach and linearise the system in (32) around the decease-free equilibrium (DFE)[24] ($I = R = D = 0, S = N$). After adding a stochastic component in the form of a zero-mean measurement error $u_t = [u_{1t}, u_{2t}, u_{3t}]'$, the resulting linear system takes a triangular form

$$I_t = \rho I_{t-1} + u_{1t},\ \Delta R_t = \gamma I_t + u_{2t},\ \Delta D_t = \delta I_t + u_{3t}. \quad (33)$$

Linearising the model at the DFE reveals the inherently nonstationary dynamics of the series at the outbreak: $I_t$ follows an AR(1) model with root $\rho = 1 + \theta - \gamma - \delta$, satisfying: $\rho > 1$ whenever $r_0 > 1$, $\rho = 1$ whenever $r_0 = 1$, and $\rho < 1$ whenever $r_0 < 1$, i.e. at an outbreak, the number of infections displays exponential growth, a result that applies to a variety of epidemiological models (see e.g. Theorem 2.1 in Allen and Van den Driessche (2008)). The linearised equations for $\Delta R_t$ and $\Delta D_t$ both follow PR models with possibly explosive regressor.

Consequently, (i) inference based on standard procedures such as OLS/MLE in (33) is only valid when $\rho < 1$ corresponding to the case $r_0 < 1$ which is not empirically relevant at the outbreak but may become relevant after government intervention, (ii) when $\rho > 1$, the series for $I_t$ exhibit exponential growth and standard semi-parametric procedures such as OLS do not provide valid inference unless $u_t$ is i.i.d. Gaussian, and (iii) when $\rho$ is in vicinity of unity (i.e. when the contract rate $\theta$ is approximately equal to the removal rate $\gamma+\delta$), OLS/MLE procedures involve nonstandard unit root or local-to-unity asymptotics and $\mathcal{N}(0, 1)$ inference is invalid. Crucially, inference in the equations for $\Delta R_t$ and $\Delta D_t$ is affected by the level of persistence of $I_t$, and consequently, OLS/MLE inference on $\gamma$ and $\delta$ is only valid in the case $r_0 < 1$. Alternative robust procedures are also invalid

[23] The choice for removing $S_t$ facilitates estimation since data on $S_t$ are unavailable.

[24] This approximation is accurate at early stages of an epidemic, when $S_t$ is large relatively to $I_t, R_t$ and $D_t$. Even at the end of our sample, the proportion $S_t/N$ is 90%-96% for all countries, suggesting that the linearisation around DFE is a good approximation.

since their region of validity is restricted to $(0,1]$, which is not empirically relevant. The IV procedure of this paper remains valid for all parameter regions for $r_0$ and without distributional or homogeneity assumptions of the innovations. Epidemiologists consider $r_0$ the key parameter for determining whether an epidemic is controllable and for understanding its transmission mechanism and, therefore, being able to construct CIs with correct coverage regardless of the value of $r_0 \in (0,\infty)$ and without distributional assumptions is of practical importance for policy makers.

While the linearised SIR model in (33) is a very simple and stylised model and the Covid-19 data have been shown to suffer from serious measurement errors and omissions, we make use of the basic SIR model to illustrate the usefulness and empirical relevance of the uniform inference procedure proposed in this paper. Its main advantage is that it gives rise to CIs for the parameters of SIR-type models with correct coverage rates in both highly infectious and remissive periods, a property of crucial empirical relevance as this section demonstrates: $r_0$ may take values in $(0,1)$, $(1,\infty)$ as well as values in close vicinity to unity depending on the various stages of the epidemic. We are not aware of any alternative statistical procedure which can achieve this throughout the range $\rho \in (0,\infty)$ without restricting attention to a particular region of the parameter space and without imposing parametric assumptions on $u_t$ in the explosive region.

We apply the IV procedure of the paper to the linearised SIR model (33) on Covid-19 data and construct CIs for the parameters $\theta$, $\gamma$, $\delta$ and for the BRN $r_0$ across a panel of countries. The triangular system in (33) implies that the number of infections in the early stages of the Covid-19 outbreak, before any government intervention, follows an explosive AR model (since $r_0 > 1$ implies $\theta > (\gamma+\delta)$); the aim of containment policies was to reduce $r_0$ below unity. We use a dataset on daily number of confirmed, recovered and deceased individuals obtained from the John Hopkins University database (https://github.com/CSSEGISandData/COVID-19) for Italy, Germany, Denmark and Israel[25]. 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[26]. 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[27] (first reported death: 24/07/2020; 25/07/2020:26/11/2020, 27/11/2020:31/03/2021, 01/04/2020:04/08/2021). Our choice to conduct inference over subsamples is motivated by the unlikelihood that the model's parameters remain constant over time; aggressive government policies aimed at containing the early epidemic's dynamics aimed at either reducing the number of new infections through imposing lockdowns and

[25] The choice of countries is motivated by the availability and quality of series on the number of recovered.

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

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

social distancing measures (reducing $\theta$), through improved medical response to the outbreak: hospital bed availability, improved treatment (increasing $\gamma$, reducing $\delta$), or later on, through vaccination by reducing the proportion of susceptibles $S_0/N$. Since our procedure is valid uniformly over $r_0 \in (0, \infty)$, we are able to assess the effect of lockdowns on the value of $r_0$. We construct the CIs for $\theta$, $\gamma$, $\delta$ and $r_0$ for each country and subsample with the instrument choice of Section 4.1.

FIGURE 14

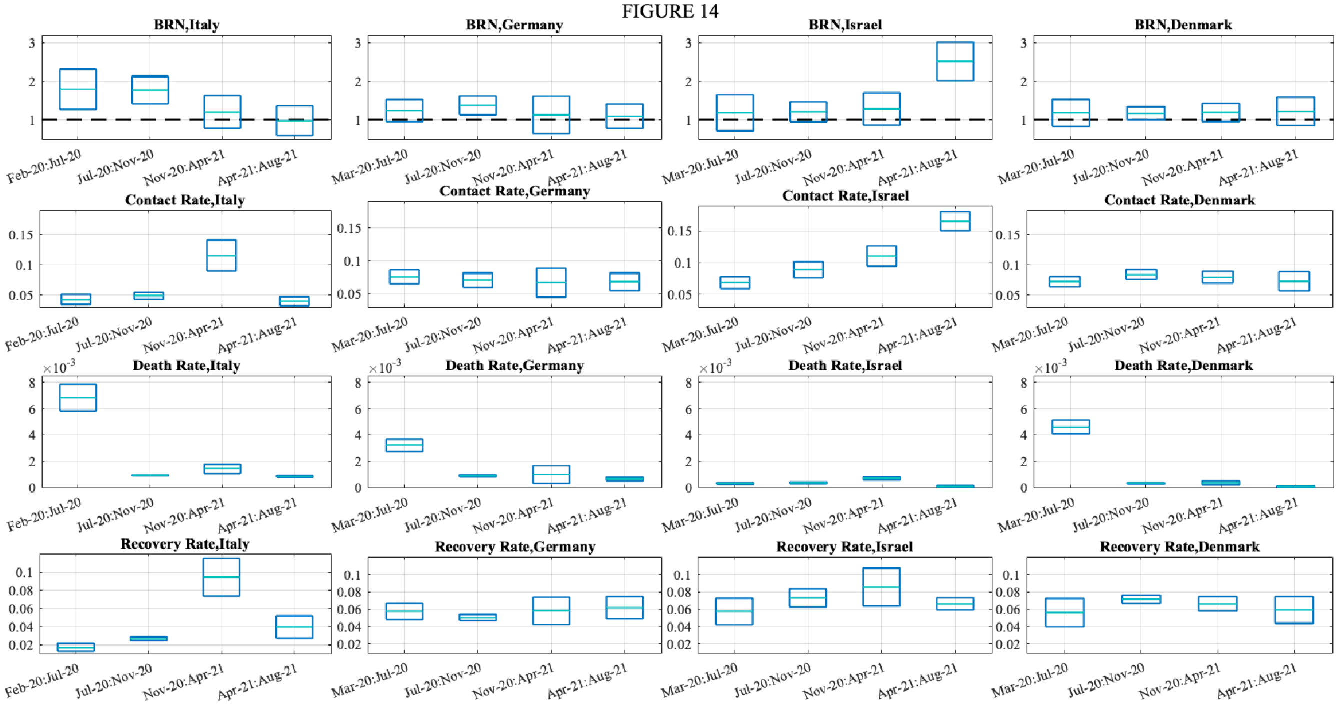



Figure 14 presents the IV estimates and 95% CIs for $r_0$, $\theta$, $\delta$ and $\gamma$ for each country and subsample. There are three main conclusions from our empirical analysis. First, the death rate has considerably fallen over time in all countries, and the recovery rate has increased over time for most countries. Second, the contact rate is constant over time for countries like Germany and Denmark, but increasing over time (especially during the winter of 2021) for Italy and Israel. Third, we find very different values for the BRN across countries: $r_0$ is relatively constant over time for countries like Denmark and Germany and while its value is usually above unity, $r_0 = 1$ is most of time included in the 95% CI. On the other hand, for Italy, we find that $r_0$ falls below unity in the period April-August 2021 while for Israel (whose experience was 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.

## 5.2 Bubbles in the Magnificent Seven's Stock Prices

In this Section, we apply our inference procedure to test for the presence of speculative bubbles in the stock price of the so-called Magnificent Seven tech stocks (Apple, Nvidia, Microsoft, Amazon, Google, Meta and Tesla). There has been an increased interest in whether on not the price rallies that these stocks have experienced are due to the underlying fundamentals changing or due to speculative bubbles, particularly following the release of ChatGPT and the subsequent bullishness in artificial intelligence. We use daily stock price data from Bloomberg for the period 04/01/2000-18/02/2025 (the samples for Google, Meta and Tesla are shorter due to availability). Pavlidis et al. (2017) argues that searching for bubbly episodes in the difference between realised asset prices and future price expectations can cancel possible explosive trends in the fundamentals of the asset, and hence is preferable for detection of purely speculative bubbles. For future expectations, we use series on 12-month mean price projections of analysts from the IBES database.

The existing tests of Phillips et al. (2011, 2015a,b) use recursive algorithms[28] with augmented Dickey-Fuller (ADF) unit root tests run over subsamples against one-sided alternatives $\rho > 1$, which require different simulated critical values depending on the recursive design of the test and window size used. A major advantage of our test statistic is that it is asymptotically $\mathcal{N}(0,1)$ regardless of: the presence or absence of bubbles (and even in moderately or purely stationary mean-reverting dynamics which could arise in the case of occasionally collapsing bubbles), linear trends, inclusion of intercepts or proportions of the sample used; hence we do not require different simulated critical values to cover all these cases. Moreover, our statistic is also robust to heterosckedasticity (of types covered by Assumption 2) both under the null $\rho = 1$ and under the alternative $\rho > 1$ and, crucially, our procedure remains valid in the presence of purely explosive bubbles. This is in contrast to existing procedures which only consider mildly explosive processes under the alternative. It is straightforward to design elaborate recursive inference procedures based on a rolling version of our IV-based t-statistic as well as more sophisticated datestamping schemes, along the lines of the procedures of Phillips et al. (2011, 2015a,b). For space considerations, we choose to perform a simple rolling window scheme here, but we stress that the simplicity and uniformity of our t-statistic make it straightforward to be extended to such screening algorithms, which are expected to further improve on the power of the simple rolling window test. We run the hypothesis test of $\rho = 1$ on the daily difference between spot and forward prices for each stock, against a one-sided alternative $\rho > 1$ over a 300-day one-sided rolling window[29] (one-sided windows are more suitable for real-time detection since they do not use future observations)[30].

In Figure 15, we compare our IV-test statistics (in dark blue) against the critical values at different significance levels, as well as against the IVX procedure of Phillips and Magdalinos (2009)

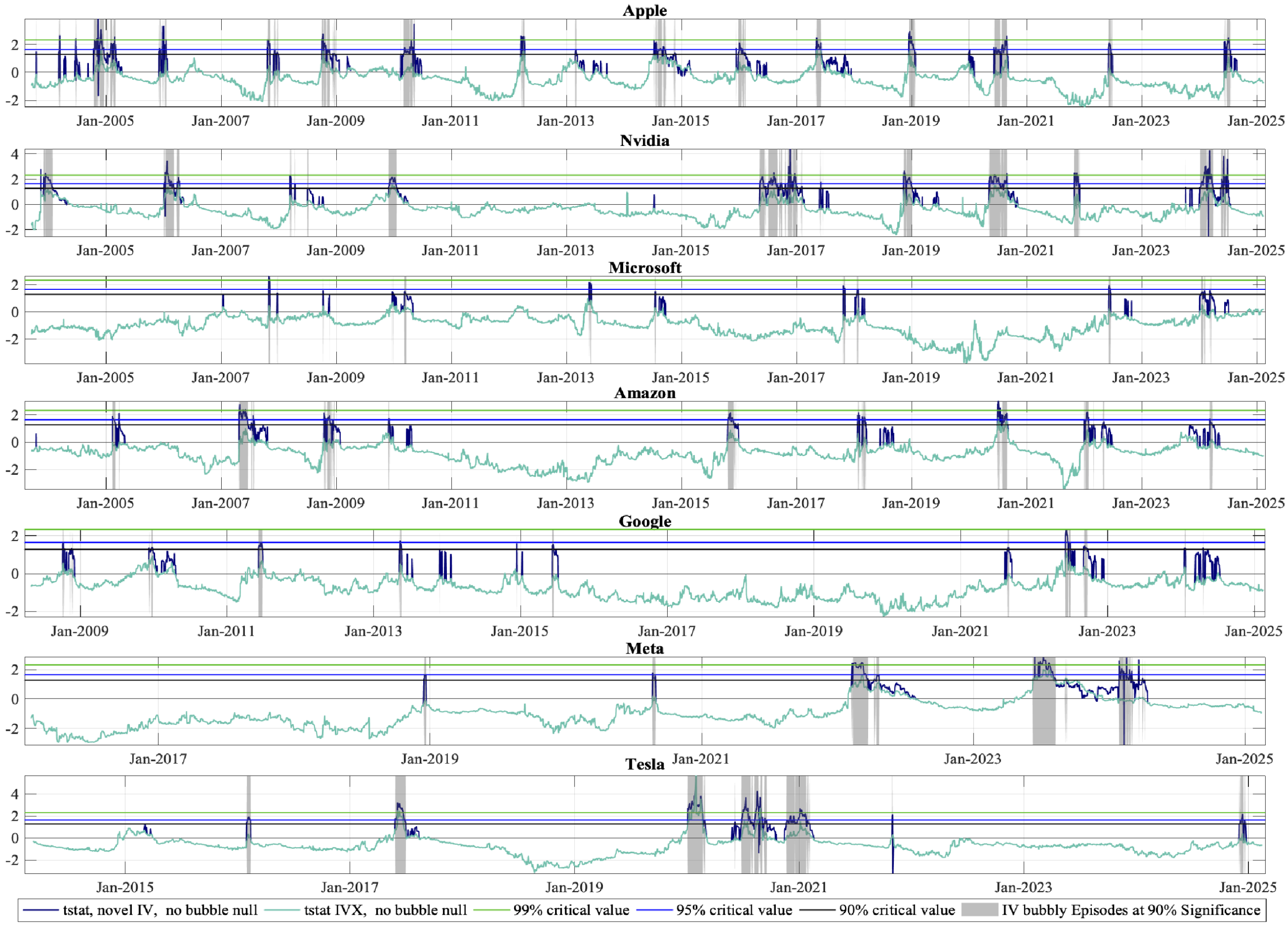
[28] There is a variety of recursive and rolling schemes in the literature designed to detect bubbles in real-time, including sup (S-) ADF, backward SADF, generalised (G) SADF, backward GSADF as well as CUMSUM tests.

[29] See e.g. Demeterscu et al. (2022) and Pavlidis et al. (2017) for applications of the IVX test methodology conducted on rolling windows.

[30] Our results are robust to different window sizes and the use of median instead of mean analysts' projections. We also find very similar bubble episodes if we run the test on the spot prices directly.

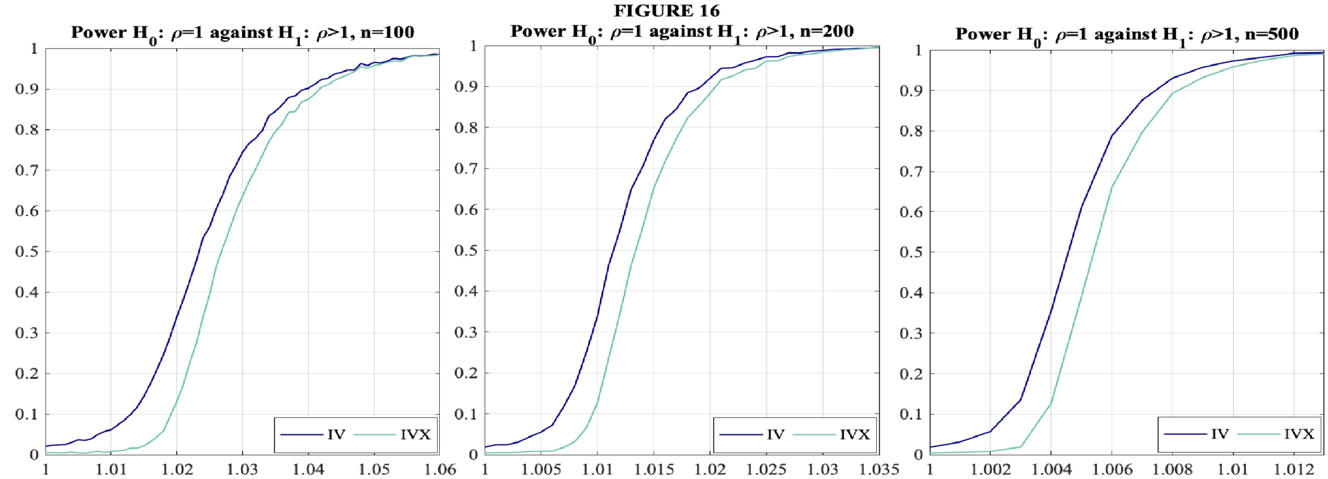
(in light blue). As it is clear from the figure, our procedure is considerably more powerful in detecting bubbles over the IVX procedure, which is only able to detect 12% of all the bubbly episodes our procedure identifies at 90% significance level. We run a simple simulation exercise based on 5,000 replications under i.i.d.$\mathcal{N}(0,1)$ innovations and different sample sizes to study these power gains in a controlled environment. In Figure 16, we display the rejection probabilities of the IV and IVX procedures for the one-sided hypothesis $\rho > 1$ at 90% significance under different alternatives, using the same choice of stationary instrument (and for the IV procedure the choice of mildly explosive instrument recommended in Section 4). It is clear from the figure that the new IV procedure has considerable power gains over the IVX procedure on the right side of unity in repeated simulation, which makes it significantly more effective in detecting bubbly episodes.

**FIGURE 15**




In terms of empirical evidence, our procedure detects speculative bubbles in all seven stocks in various periods: we find a long-lasting bubble in the price of Nvidia around the end of 2016 and beginning of 2017 and mid-2020 when Covid-19 pandemic increased demand for Nvidia's products. We also find a long period of multiple speculative bubbles in Tesla's price during 2020 and early 2021, when the Tesla's price was breaking records and surpassing analysts' projections. Finally, there are notable bubbles in Meta's price in the beginning of 2022 and the summer of 2023. More recently, we find bubbly periods: (i) in July 2024 for Apple and Nvidia, (ii) in January - March 2024 for Nvidia, Microsoft, Google, Amazon and Meta and (iii) after the presidential election in November and December 2024 for Tesla.

FIGURE 16




# 6 Conclusion

The paper proposes a unified, distribution-free framework for inference in autoregressive, predictive regression and local projection models, when the regressor's autoregressive root is in $(-\infty, \infty)$. This includes: (i) stable and near-stable processes, (ii) (seasonal) unit root and local-to-unity regressors, and (iii) regressors that exhibit stochastic exponential growth (e.g. explosive and mildly explosive). The unified inference is based on a novel estimation method that employs an instrumental variable approach with an artificially constructed instrument with a data-driven combination of a (possibly oscillating) moderately stationary and mildly explosive root. The resulting IV estimators for the AR parameter in the autoregression, the slope parameter in the predictive regression framework and the impulse response in the local projection model are all shown to have a mixed-Gaussian limit distribution under all persistence regimes, and independently of the distribution of the innovations and the initial condition. Consequently, the t-statistic based on the new estimators is asymptotically standard normal with uniform size over arbitrary closed subintervals of $(-\infty, \infty)$ and gives rise to asymptotically correctly-sized CIs. To our knowledge, this is the first method that delivers central limit theory and, consequently, general distribution-free asymptotic inference with a regressor with autoregressive root in $(-\infty, -1) \cup (1, \infty)$ and achieves uniform asymptotic inference over the entire autoregressive range $(-\infty, \infty)$.

We demonstrate that our inference procedure exhibits very good finite sample properties in an extensive Monte Carlo study and compares favourably to existing procedures for inference in autoregressions (Andrews and Guggenberger (2014)), predictive regressions (Elliott et al. (2015))

and local projection models (Olea and Plagborg-Moller (2021)) in their parametric validity range $(-1,1]$ while providing correct inference on $(-\infty,-1]\cup(1,\infty)$, where no existing alternative approach has general asymptotic validity.

Finally, we demonstrate how our inference procedure can be used to answer empirical questions for which alternative procedures are not well suited. In particular, we employ our procedure to construct CIs for the parameters of a standard SIR model without restricting the parameter space, i.e. without *a priori* knowledge of whether the epidemic is in a controllable or uncontrollable stage. We also demonstrate how our procedure can be employed for recursive search for the presence of speculative bubbles in the stock price of the Magnificent Seven tech stocks, delivering not only theoretical validity in the purely explosive region but also considerable power gains.

# References


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

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

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

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

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

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

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

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

Aue, A. and Horvath, L. (2007) A Limit Theorem For Mildly Explosive Autoregression With Stable Errors, *Econometric Theory* **23**(2): 201-220

Basawa, I. V.,Mallik, A. K., McCormick, W. P., Reeves, J. H. and Taylor, R. L. (1991) Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics* **19**: 1098–101

Blanchard, O.J. (1979) Speculative bubbles, crashes and rational expectations. *Economics Letters* **3**(4): 387-389.

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

Cavanagh, C. L., Elliott, G. and Stock, J. H. (1995) Inference in models with nearly integrated regressors. *Econometric Theory* **11**: 1131–47

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

Chambers, M., Ercolani, J. and Taylor, R. (2014) Testing for seasonal unit roots by frequency domain regression. *Journal of Econometrics* **178**(2): 243-258

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

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

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

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

Evans G. (1991) Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review* **81**(4): 922-930

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

Giraitis, L., Kokoszka, P., & Leipus, R. (2000). Stationary ARCH models: dependence structure and central limit theorem. Econometric Theory, 16(1), 3-22.

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

Goyal, A., Welch, I. and Zafirov, A. (2024) A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction, *The Review of Financial Studies*, **37**(11): 3490-3557

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., Sollis, R., Taylor, R. (2016) Tests for explosive financial bubbles in the presence of non-stationary volatility. *Journal of Empirical Finance* **38**: 548-574

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*

Hylleberg, S., Engle, R., Granger, C. and Yoo B. (1990) Seasonal integration and cointegration. Journal of Econometrics **44**(1): 215-238

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

Kendall, M. (1953) The analysis of economic time series, Part I. Prices. *Journal of the Royal Statistical Society* **96**:11-25

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 m.d. sequences satisfying the local Marcinkiewicz-Zygmund condition. *Bulletin of the Institute of Mathematics Academia Sinica* **11**: 1-13

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

Magdalinos, T. and Petrova, K. (2024) OLS limit theory for drifting sequences of parameters on the explosive side of unity, Working paper

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 AR 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

Olea, J.L.M. and Plagborg-Moller, M. (2021) Local projection inference is simpler and more robust than you think. *Econometrica* **89**(4): 1789-1823

Pavlidis, E., Paya, I, Peel, D. (2017) Testing for speculative bubbles using spot and forward prices. *International Economic Review* **58**(4): 1191-1226

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

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

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

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

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

Phillips P.C.B., Shi, S., Yu, J. (2015a) Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* **56**(4):1043-1078

Phillips P.C.B., Shi, S., Yu, J. (2015b) Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review* **56**(4): 1079-1134

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

Rossi, B. (2007) Expectation hypotheses tests at long horizons. *Econometrics Journal* **10**: 554-579

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

# Online Appendix

This Appendix contains auxiliary mathematical results, all proofs of the paper and additional simulation results.

## 1.1 Auxiliary Results

This section contains six auxiliary mathematical results that develop an asymptotic theory for sample moments of $x_{n,t}$ and $\tilde{z}_{n,t}$ and facilitate the proof of Theorems 1-3. Lemma 1 contains results on subsequential convergence that allows a passage from the AMG property of Theorem 1 to the uniform critical regions and confidence intervals of Theorems 2 and 3. Lemma 2 establishes the asymptotic separation property of the near-stationary and near-explosive classes and that of the nonstationary regular and oscillating classes, required for correct instrument selection. Lemma 3 establishes sample moment limit theory for $x_{n,t}$ in classes $C_+$(i)-$C_+$(ii). Lemma 4 establishes sample moment limit theory for $x_{n,t}$ in classes $C_+$(ii)-$C_+$(iii). Lemma 5 provides the functional central limit theorem required to establish the joint convergence and asymptotic independence of $n^{-1/2}\left(1-\rho_{nz}^2\right)^{1/2}\sum_{t=1}^n \tilde{z}_{n,t-1}\varepsilon_{n,t}$ and $n^{-1}\left(1-\rho_{nz}^2\right)\sum_{t=1}^n \underline{x}_{n,t-1}\tilde{z}_{n,t-1}$ in the local to unity case $C_+$(ii), where the denominator of the IV estimators (9) and (10) has a random limit. In Lemma 6, we show how the asymptotic behaviour of the oscillating classes $C_-$(i)-$C_-$(iii) may be derived from that of their regular counterparts $C_+$(i)-$C_+$(iii) via the transformation $x_{n,t} \mapsto (-1)^t x_{n,t}$. The proofs of Lemmata 1-6 are provided below.

**Lemma 1.**

(i) *Let $(\rho_n)_{n\in\mathbb{N}}$ satisfy $\rho_n \to \rho \in \mathbb{R}$. 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 7.*

(ii) *Let $\left(u_{n,t}^2\right)_{t\in\mathbb{Z}}$ be a uniformy integrable sequence for each $n \in \mathbb{N}$. Then, for any sequence $(\lambda_n)_{n\in\mathbb{N}}$ satisfying $\lambda_n \to \infty$, $\sup_{t\in\mathbb{Z}} \mathbb{E}\left(u_{n,t}^2 \mathbf{1}\left\{u_{n,t}^2 > \lambda_n\right\}\right) \to 0$.*

(iii) *Let $(u_{n,t})_{t\in\mathbb{N}}$ be an ARCH($\infty$) process satisfying Assumption 2(ii) for each $n \in \mathbb{N}$. For any subsequence $(m_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ there exists a further subsequence $(k_n)_{n\in\mathbb{N}} \subseteq (m_n)_{n\in\mathbb{N}}$ such that $\sigma_{k_n,t}^2 = \mathbb{E}_{\mathcal{F}_{k_n,t-1}} u_{k_n,t}^2$ satisfies (14).*

**Lemma 2.** *Let $(m_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of positive numbers such that $m_n \to \infty$. Under Assumptions 6 and 7:* **(i)** *if $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(i) then $m_n \mathbf{1}_{\bar{F}_n^+} \to_p 0$ and $m_n \mathbf{1}_{\bar{F}_n^-} \to_p 0$;* **(ii)** *if $(\rho_n)_{n\in\mathbb{N}}$ belongs to C(iii) then $m_n \mathbf{1}_{F_n^+} \to_p 0$ and $m_n \mathbf{1}_{F_n^-} \to_p 0$;* **(iii)** *if $\rho_n \to \rho \geq 1$ then $m_n \mathbf{1}_{F_n^-} \to_p 0$ and $m_n \mathbf{1}_{\bar{F}_n^-} \to_p 0$ ;* **(iv)** *if $\rho_n \to \rho \leq -1$ then $m_n \mathbf{1}_{F_n^+} \to_p 0$ and $m_n \mathbf{1}_{\bar{F}_n^+} \to_p 0$.*

For Lemmata 3-5, note that recursing the autoregression (24) up to $x_{n,0} = \mu + X_{n,0}$ implies

$$x_{n,t} = \mu + X_{n,0}\rho_n^t + x_{0t}^{(n)}, \quad x_{0t}^{(n)} \equiv x_{0t} := \textstyle\sum_{j=1}^t \rho_n^{t-j} u_{n,j} \tag{A.1}$$

where $x_{0t}^{(n)}$ denotes the autoregression (24) when $\mu = 0$ and $X_0(n) = 0$; we will drop the array superscript from $x_{0t}^{(n)}$ for notational economy. Equations (6), (7) and (8) imply that $\tilde{z}_{1t} := \tilde{z}_{n,t}\mathbf{1}_{F_n^+}$, $\tilde{z}_{1t}^- := \tilde{z}_{n,t}\mathbf{1}_{F_n^-}$, $\tilde{z}_{2t} := \tilde{z}_{n,t}\mathbf{1}_{\bar{F}_n^+}$ and $\tilde{z}_{2t}^- := \tilde{z}_{n,t}\mathbf{1}_{\bar{F}_n^-}$ satisfy the recursions

$$\tilde{z}_{1t} = \varphi_{1n}\tilde{z}_{1t-1} + \Delta x_{n,t} \text{ and } \tilde{z}_{1t}^- = \varphi_{1n}^-\tilde{z}_{1t-1}^- + \nabla \underline{x}_{n,t} \tag{A.2}$$

$$\tilde{z}_{2t} = \varphi_{2n}\tilde{z}_{2t-1} + \hat{u}_{n,t} \text{ and } \tilde{z}_{2t}^- = \varphi_{2n}^-\tilde{z}_{2t-1}^- + \hat{u}_{n,t}. \tag{A.3}$$

The (unobservable) near-stationary counterparts of (A.2) and near-explosive counterparts of (A.3) satisfy the recursions

$$z_{1t} = \varphi_{1n}z_{1t-1} + u_{n,t} \text{ and } z_{1t}^- = \varphi_{1n}^- z_{1t-1}^- + u_{n,t} \tag{A.4}$$

$$z_{2t} = \varphi_{2n}z_{2t-1} + u_{n,t} \text{ and } z_{2t}^- = \varphi_{2n}^- z_{2t-1}^- + u_{n,t}. \tag{A.5}$$

Recall the definitions of $\Gamma_n$ and $\Gamma$ in (15) and of $J_c(\cdot)$ in (27).

**Lemma 3.** *The following hold under Assumptions 6 and 7 when $x_{n,t}$ belongs to $C_+$(i)-$C_+$(ii):*

**(i)** $n^{-1}(1-\rho_n^2\varphi_{1n}^2)\sum_{t=1}^n \underline{x}_{n,t-1}\tilde{z}_{1t-1} = \tilde{\Psi}_n + o_p(1) \to_d \tilde{\Psi}(c)$ *where* $\tilde{\Psi}(c) = \sigma^2 + 2\rho\Gamma + \left(J_c(1)^2 - 2J_c(1)\int_0^1 J_c(r)\,dr\right)\mathbf{1}\{c \in \mathbb{R}\}$, $x_{0t}$ *is defined in (A.1) and*

$$\begin{aligned}\tilde{\Psi}_n &= (1+\rho_n)\left[\sigma^2 + 2\rho_n\Gamma_n + (2\rho_n - 1)\left(n^{-1}\sum_{t=1}^n x_{0t-1}u_{n,t} - \Gamma_n\right)\right] \\ &\quad -\rho_n\left(1-\rho_n^2\right)n^{-1}\sum_{t=1}^n x_{0t-1}^2 - 2\left(n^{-1/2}x_{0n}\right)n^{-3/2}\sum_{j=1}^n x_{0j-1}.\end{aligned} \tag{A.6}$$

**(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)** $n^{-1/2}(1-\rho_n^2\varphi_{1n}^2)^{1/2}\sum_{t=1}^n \tilde{z}_{1t-1}e_{n,t} \to_d \mathcal{N}(0, v(\rho))$ *where:* $v(\rho) = (\sigma^2 + 2\rho\Gamma)\sigma_e^2$ *when* $\mathbb{E}_{\mathcal{F}_{n,t-1}}(e_{n,t}^2) = \sigma_e^2$ *or when* $\rho_n \to 1$; *when* $\rho_n \to \rho \in (-1,1)$ *and Assumption 6(ii) holds with* $v_2(\rho)$ *replaced by* $v_*(\rho) = \lim_{n\to\infty}\mathbb{E}e_{n,1}^2\left(\sum_{j=0}^\infty \rho^j u_{n,-j}\right)^2$, $v(\rho) = (1-\rho^2)v_*(\rho)$.

For Lemma 4, consider the stochastic sequences

$$[Y_n, Y_n^\varepsilon, Z_n] := \left(\varphi_{2n}^2 - 1\right)^{1/2}\left[\sum_{t=1}^n \varphi_{2n}^{-(n-t)-1}u_{n,t}, \sum_{t=1}^n \varphi_{2n}^{-(n-t)-1}\varepsilon_{n,t}, \sum_{j=1}^n \varphi_{2n}^{-j}u_{n,j}\right] \tag{A.7}$$

and the convergence rates

$$\nu_n = \left(\rho_n^2 - 1\right)^{-1/2}|\rho_n|^n\mathbf{1}\{c = \infty\} + n^{1/2}\mathbf{1}\{c \in \mathbb{R}\}, \quad \nu_{n,z} = \left(\phi_{2n}^2 - 1\right)^{-1/2}|\phi_{2n}|^n, \tag{A.8}$$

$s_n = (\rho_n\phi_{2n} - 1)^{-1}\nu_{n,z}\nu_n$ and $\tau_n = (\varphi_{2n}^2 - 1)^{-1}\varphi_{2n}^n$, with $c$ and $\phi_{2n}$ defined in Ass. 7 and (26).

**Lemma 4.** *Let $Y_n$, $Y_n^\varepsilon$, $Z_n$ be the stochastic sequences in (A.7) and $X_n := x_{n,n}/\nu_n$. Let $Y, Z, X$ denote $\mathcal{N}(0,\omega^2)$ random variables and $Y^\varepsilon$* be $\mathcal{N}(0,\sigma_\varepsilon^2)$. *Under Assumptions 6 and 7, the following hold when $x_{n,t}$ belongs to $C_+$(ii)-$C_+$*(iii)*:*

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

$$\left[\tau_n^{-1}\sum_{t=1}^n \tilde{z}_{2t-1}u_{n,t}, \tau_n^{-2}\sum_{t=1}^n \tilde{z}_{2t-1}^2, s_n^{-1}\sum_{t=1}^n x_{n,t-1}\tilde{z}_{2t-1}\right] = \left[Y_nZ_n, Z_n^2, X_nZ_n\right] + o_p(1). \tag{A.9}$$

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

**(iii)** *Under $C_+$(iii) with $\rho_n \to \rho > 1$, $X_n \to_d X_\infty$ with $X_\infty \neq 0$ a.s.; $Y_n/X_n \to_d Y/X_\infty =_d \mathcal{MN}(0, \omega^2/X_\infty^2)$ and $Y_n^\varepsilon/X_n \to_d Y^\varepsilon/X_\infty =_d \mathcal{MN}(0, \sigma_\varepsilon^2/X_\infty^2)$.*

**Lemma 5.** *Consider* $B_n(s) = n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor} u_{n,t}$, $U_n(s) = (n(1-\varphi_{1n}^2)^{-1})^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor} z_{1t-1}e_{n,t}$, $Y_n(s) = (\varphi_{2n}^2-1)^{1/2}\sum_{t=1}^{\lfloor ns\rfloor}\varphi_{2n}^{-(\lfloor ns\rfloor-t)-1}u_{n,t}$ *and* $Z_n(s) = (\varphi_{2n}^2-1)^{1/2}\sum_{t=1}^{\lfloor ns\rfloor}\varphi_{2n}^{-t}u_{n,t}$ *as elements of* $D[0,1]$*. Under Assumption 6,* $[U_n(s), B_n(s), Y_n(s), Z_n(s)] \Rightarrow [U(s), B(s), Y, Z]$ *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* $\mathbb{E}B(s)^2 = s\omega^2$*,* $[Y,Z] =_d \mathcal{N}(0,\omega^2 I_2)$ *and* $(Y,Z)$ *is independent of* $[U(s), B(s)]$*.*

Lemma 6 shows how to obtain the asymptotic distribution of the IV estimators $(\tilde\rho_{1n}^-, \tilde\rho_{2n}^-)$ and $(\tilde\beta_{1n}^-, \tilde\beta_{2n}^-)$ generated by an oscillating autoregression $x_{n,t}$ in classes C$_-$(i)-C$_-$(iii) from that of their regular counterparts via the transformation $x \mapsto (-1)^{-t}x$. Defining $x_{n,t}$ and $\underline{x}_{n,t}$ as in (1) and (4) with $\rho_n < 0$, it is easy to see that $x_{n,t}^+ := (-1)^{-t}(x_{n,t}-\mu)$ and $\underline{x}_{n,t}^+ = (-1)^{-t}\underline{x}_{n,t}$ satisfy

$$x_{n,t}^+ = |\rho_n|\, x_{n,t-1}^+ + (-1)^{-t}u_{n,t}, \quad \underline{x}_{n,t}^+ = |\rho_n|\,\underline{x}_{n,t-1}^+ + (-1)^{-t}\underline{u}_{n,t} \tag{A.10}$$

so $x_{n,t}^+$ is a regular autoregression with root $|\rho_n|$. Define by $\hat\rho_n^+$ and $\hat u_{n,t}^+$ the OLS estimator and residuals in (4) with $\underline{x}_{n,t}$ replaced by $\underline{x}_{n,t}^+$ and define by $(F_n^{++}, \bar F_n^{++})$ the events $(F_n^+, \bar F_n^+)$ in (6) with $\hat\rho_n$ replaced by $\hat\rho_n^+$. Let $\underline{y}_{n,t}^+ = (-1)^{-t}\underline{y}_{n,t}$, $\tilde z_{1t}^+ = \sum_{j=1}^t |\varphi_{1n}^-|^{t-j}\Delta x_{n,j}^+$ and $\tilde z_{2t}^+ = \sum_{j=1}^t |\varphi_{2n}^-|^{t-j}\hat u_{n,j}^+$ and define by $\tilde\rho_{1n}^+$, $\tilde\beta_{1n}^+$ and $\tilde\rho_{2n}^+$, $\tilde\beta_{2n}^+$ the IV estimators in (9) and (10) by replacing $\underline{x}_{n,t}$ by $\underline{x}_{n,t}^+$, $\underline{y}_{n,t}$ by $\underline{y}_{n,t}^+$ and $\tilde z_{n,t}$ by $\tilde z_{1t}^+$ for $\tilde\rho_{1n}^+$, $\tilde\beta_{1n}^+$ and by $\tilde z_{2t}^+$ for $\tilde\rho_{2n}^+$, $\tilde\beta_{2n}^+$.

**Lemma 6.** *Consider an oscillating autoregression (1) with* $\rho_n < 0$*. Let* $\tilde\rho_{1n}^- = \tilde\rho_n \mathbf{1}_{F_n^-}$, $\tilde\beta_{1n}^- = \tilde\beta_n\mathbf{1}_{F_n^-}$, $\tilde\rho_{2n}^- = \tilde\rho_n\mathbf{1}_{\bar F_n^-}$ *and* $\tilde\beta_{2n}^- = \tilde\beta_n\mathbf{1}_{\bar F_n^-}$*. Then* $F_n^- = F_n^{++}$, $\bar F_n^- = \bar F_n^{++}$, $\tilde\rho_{2n}^- - \rho_n = -(\tilde\rho_{2n}^+ - |\rho_n|)$ *and* $\tilde\beta_{2n}^- - \beta = -(\tilde\beta_{2n}^+ - \beta)$*. Under Assumption 6 and C$_-$(i)-C$_-$(ii),* $\pi_n(\tilde\rho_{1n}^- - \rho_n) = -\pi_n(\tilde\rho_{1n}^+ - |\rho_n|) + o_p(1)$ *and* $\pi_n(\tilde\beta_{1n}^- - \beta) = -\pi_n(\tilde\beta_{1n}^+ - \beta) + o_p(1)$*.*

## 1.2 Mathematical Proofs

We use the abbreviations BW for the Bolzano-Weierstrass theorem, HH(1980) for Hall and Heyde (1980), MP(2020) for Magdalinos and Phillips (2020) and MP(2024) for Magdalinos and Petrova (2024). From the proof of Lemma 3 onwards, we define the following sequences for brevity: $\kappa_n := n \wedge |\rho_n^2 - 1|^{-1}$, $\lambda_{1n} := \kappa_n \wedge (1-\varphi_{1n}^2)^{-1}$, $\Lambda_{1n} := \kappa_n \vee (1-\varphi_{1n}^2)^{-1}$, $\lambda_{2n} := \kappa_n \wedge (\varphi_{2n}^2-1)^{-1}$ and $\Lambda_{2n} := \kappa_n \vee (\varphi_{2n}^2-1)^{-1}$.

**Proof of Lemma 1.** For part (i), the result 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 $(\rho_{m_n})_{n\in\mathbb{N}}$ 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: $(\rho_{s_n})_{n\in\mathbb{N}}$ belongs to C(i) if $c_\infty = -\infty$, or $(\rho_{s_n})_{n\in\mathbb{N}}$ belongs to C(ii) if $c_\infty \in \mathbb{R}$, or $(\rho_{s_n})_{n\in\mathbb{N}}$ belongs to C(iii) if $c_\infty = \infty$.

For part (ii), $(u_{m_n,t}^2)_{t\in\mathbb{Z}}$ is UI for any $(m_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$. Let $\upsilon_n(\lambda) := \sup_{t\in\mathbb{Z}}\mathbb{E}\left(u_{n,t}^2\mathbf{1}\left\{u_{n,t}^2 > \lambda\right\}\right)$.

By UI of $\left(u^2_{m_1,t}\right)_{t\in\mathbb{Z}}, \left(u^2_{m_2,t}\right)_{t\in\mathbb{Z}}, ...$ we obtain that for each $n \in \mathbb{N}$ there exists $b_n > 0$ such that

$$\lambda > b_n \Rightarrow \upsilon_{m_n}(\lambda) < 2^{-n}. \tag{A.11}$$

For the moment, let $\lambda_n \uparrow \infty$; for any $\beta > 0$ there exists $k_\beta \in \mathbb{N}$ such that $\lambda_{k_\beta} > \beta$. Running $\beta$ along $\{b_1, b_2, ...\}$ we obtain that for each $n \in \mathbb{N}$ there exists $k_{b_n} \in \mathbb{N}$ such that $\lambda_{k_{b_n}} > b_n$; letting $k_{\beta_n} := k_{b_n} \vee n$, the monotonicity of $(\lambda_n)$ implies that $\lambda_{k_{\beta_n}} > b_n$; (A.11) then implies that $\upsilon_{m_n}\left(\lambda_{k_{\beta_n}}\right) < 2^{-n}$. Since $k_{\beta_n} \geq n$, $\left(k_{\beta_n}\right)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ so choosing $m_n = k_{\beta_n}$ we obtain $\upsilon_{k_{\beta_n}}\left(\lambda_{k_{\beta_n}}\right) \to 0$. We conclude that for any subsequence $\{\upsilon_{r_n}(\lambda_{r_n}) : n \in \mathbb{N}\}$ of $\{\upsilon_n(\lambda_n) : n \in \mathbb{N}\}$ there exists a further subsequence $\{\upsilon_{s_n}(\lambda_{s_n}) : n \in \mathbb{N}\}$ (with $s_n = k_{b_{r_n}} \vee r_n$) such that $\upsilon_{s_n}(\lambda_{s_n}) \to 0$; this implies that $\upsilon_n(\lambda_n) \to 0$ as required when $\lambda_n \uparrow \infty$. If $\lambda_n \to \infty$ (not necessarily monotonically) the monotone subsequence lemma implies that $(\lambda_n)$ has an increasing subsequence; for any subsequence $\{\upsilon_{m_n}(\lambda_{m_n}) : n \in \mathbb{N}\}$ of $\{\upsilon_n(\lambda_n) : n \in \mathbb{N}\}$, there exists a further subsequence $\{\upsilon_{k_n}(\lambda_{k_n}) : n \in \mathbb{N}\}$ where $\lambda_{k_n} \uparrow \infty$, so $\upsilon_{k_n}(\lambda_{k_n}) \to 0$.

For part (iii), $\left(u_{n,t}, \sigma^2_{n,t}\right)$ satisfy Assumption 2(ii) for each $n$, so

$$u_{n,t} = \sigma_{n,t}\eta_{n,t}, \quad \sigma^2_{n,t} = \varpi_n + \textstyle\sum_{i=1}^{\infty} \alpha_{n,i} u^2_{n,t-i} \tag{A.12}$$

with $\mathcal{F}_{n,t} = \sigma\left(\eta_{n,t}, \eta_{n,t-1}, ...\right)$, $\alpha_{n,i} \geq 0$, $\inf_{n\geq1} \varpi_n \geq \delta$, $\sup_{n\geq1} \sum_{i=0}^{\infty} \alpha_{n,i} \leq 1 - \delta$ and

$$\sup_{n\geq1} \textstyle\sum_{i=M}^{\infty} \alpha_{n,i} \to 0 \ \text{ as } M \to \infty \tag{A.13}$$

the latter since $\sum_{i=M}^{\infty} \alpha_{n,i} \leq M^{-\delta} \sum_{i=M}^{\infty} i^\delta \alpha_{n,i}$ so $\sup_{n\geq1} \sum_{i=M}^{\infty} \alpha_{n,i} = O\left(M^{-\delta}\right)$ for some $\delta > 0$ by Assumption 2(ii). We first show that there exists a sequence $(\alpha_i)_{i\geq1}$ such that $\alpha_i \geq 0$ and $\sum_{i=1}^{\infty} \alpha_i < 1$ and a subsequence $(k_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ such that

$$A_{k_n} := \textstyle\sum_{i=1}^{\infty} |\alpha_{k_n,i} - \alpha_i| \to 0. \tag{A.14}$$

For each $i \in \mathbb{N}$, $\sup_{n\geq1} \alpha_{n,i} \leq \sup_{n\geq1} \sum_{i=1}^{\infty} \alpha_{n,i} \leq 1 - \delta$, so the BW theorem implies that there exists a subsequence $(k_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ such that $\alpha_{k_n,i} \to \alpha_i$ for each $i \in \mathbb{N}$. We define a sequence $(\alpha_i)_{i\geq1}$ from the above subsequential limits. Clearly, $\alpha_i \geq 0$ for all $i$ and for each $K \in \mathbb{N}$

$$\textstyle\sum_{i=1}^{K} \alpha_i = \sum_{i=1}^{K} \lim_{n\to\infty} \alpha_{k_n,i} = \lim_{n\to\infty} \sum_{i=1}^{K} \alpha_{k_n,i} \leq \sup_{n\geq1} \sum_{i=1}^{\infty} \alpha_{n,i} \leq 1 - \delta$$

i.e. $\sum_{i=1}^{\infty} \alpha_i = \sup_{K\geq1} \sum_{i=1}^{K} \alpha_i < 1$. By (A.13) and the summability of $(\alpha_i)_{i\geq1}$, for any $\varepsilon > 0$ there exist natural numbers $M_1(\varepsilon)$ and $M_2(\varepsilon)$ such that

$$\textstyle\sup_{n\geq1} \sum_{i=M}^{\infty} \alpha_{k_n,i} < \varepsilon/4 \ \ (\forall M \geq M_1(\varepsilon)) \ \text{ and } \ \sum_{i=M}^{\infty} \alpha_i < \varepsilon/4 \ \ (\forall M \geq M_2(\varepsilon)).$$

Letting $M_0(\varepsilon) := \max\{M_1(\varepsilon), M_2(\varepsilon)\}$ we obtain

$$\begin{aligned} A_{k_n} &\leq \textstyle\sum_{i=1}^{M_0(\varepsilon)-1} |\alpha_{k_n,i} - \alpha_i| + \sum_{i=M_0(\varepsilon)}^{\infty} \alpha_{k_n,i} + \sum_{i=M_0(\varepsilon)}^{\infty} \alpha_i \\ &< \textstyle\sum_{i=1}^{M_0(\varepsilon)-1} |\alpha_{k_n,i} - \alpha_i| + \varepsilon/2. \end{aligned} \tag{A.15}$$

Now the pointwise convergence $\alpha_{k_n,i} \to \alpha_i$ for each $i \in \mathbb{N}$ implies that for any $\varepsilon > 0$ and any $i \in \mathbb{N}$ there exist natural numbers $n_i(\varepsilon)$ such that

$$|\alpha_{k_n,i} - \alpha_i| < \varepsilon\left(2M_0(\varepsilon)\right)^{-1} \ \ (\forall n \geq n_i(\varepsilon)). \tag{A.16}$$

Taking $n_0(\varepsilon) := \max\{n_i(\varepsilon) : 1 \leq i < M_0(\varepsilon)\}$, (A.15) and (A.16) imply that for any $\varepsilon > 0$ there exists $n_0(\varepsilon) \in \mathbb{N}$ such that $A_{k_n} < \varepsilon$ for all $n \geq n_0(\varepsilon)$, showing (A.14).

Having established (A.14), we employ the approximating sequence $(\alpha_i)$ to define

$$\check{u}_{n,t} = \check{\sigma}_{n,t}\eta_{n,t}, \quad \check{\sigma}^2_{n,t} := \varpi_n + \textstyle\sum_{i=1}^{\infty} \alpha_i u^2_{n,t-i}. \tag{A.17}$$

By (A.14), $\sup_{t\geq 1} \left\|\sigma^2_{k_n,t} - \check{\sigma}^2_{k_n,t}\right\|_{L_2} \leq A_{k_n} \sup_{n\geq 1} \|u_{n,1}\|^2_{L_4} \to 0$ and, similarly, $\check{\sigma}^2_n := \mathbb{E}\check{\sigma}^2_{n,t}$ satisfies $\left|\sigma^2_{k_n} - \check{\sigma}^2_{k_n}\right| \to 0$. We may therefore write

$$\left\|\mathbb{E}_{\mathcal{F}_{k_n,t-1-m}}\left(\sigma^2_{k_n,t} - \sigma^2_{k_n}\right)\right\|_{L_2} \leq \left\|\mathbb{E}_{\mathcal{F}_{k_n,t-1-m}}\left(\check{\sigma}^2_{k_n,t} - \check{\sigma}^2_{k_n}\right)\right\|_{L_2} + \tilde{\psi}_n \tag{A.18}$$

where $\tilde{\psi}_n := \sup_{t\geq 1} \left\|\sigma^2_{k_n,t} - \check{\sigma}^2_{k_n,t}\right\|_{L_2} + \left|\check{\sigma}^2_{k_n} - \sigma^2_{k_n}\right| \to 0$. The remainder of the proof follows the lines of Example 1 in Arvanitis and Magdalinos (2019) using an $L_2$ (instead of $L_1$) norm. Theorem 4.1 and (4.11) of Giraitis et al. (2000) imply for each $n \in \mathbb{N}$

$$\check{u}^2_{n,t} - \check{\sigma}^2_n = \textstyle\sum_{i=0}^{\infty} \tilde{\alpha}_i \check{\sigma}^2_{n,t-i}\left(\eta^2_{n,t-i} - 1\right) \tag{A.19}$$

for a sequence $(\tilde{\alpha}_i)_{i\geq 0}$ defined by $\tilde{\alpha}(z) = \sum_{i=0}^{\infty} \tilde{\alpha}_i z^i = 1/\alpha(z)$ with $\alpha(z) = \sum_{i=0}^{\infty} \alpha_i z^i$ for $|z| \leq 1$, where $\sum_{i=1}^{\infty} \alpha_i < 1$ implies that $\sum_{i=0}^{\infty} |\tilde{\alpha}_i| < \infty$. (A.19) yields each $n, t, m \geq 1$,

$$\begin{aligned}\left\|\mathbb{E}_{\mathcal{F}_{k_n,t-1-m}}\left(\check{\sigma}^2_{k_n,t} - \check{\sigma}^2_{k_n}\right)\right\|_{L_2} &= \left\|\mathbb{E}_{\mathcal{F}_{k_n,t-1-m}}\left(\check{u}^2_{k_n,t} - \check{\sigma}^2_{k_n}\right)\right\|_{L_2} = \left\|\textstyle\sum_{i>m} \tilde{\alpha}_i \check{\sigma}^2_{k_n,t-i}\left(\eta^2_{k_n,t-i} - 1\right)\right\|_{L_2} \\ &\leq \left\{\mathbb{E}\left(\eta^2_{k_n,0} - 1\right)^2 \mathbb{E}\check{\sigma}^4_{k_n,0} \textstyle\sum_{i>m} |\tilde{\alpha}_i|\right\}^{1/2} \end{aligned} \tag{A.20}$$

since $\mathbb{E}\left(\check{\sigma}^2_{k_n,t-i}\right)^2\left(\eta^2_{k_n,t-i} - 1\right)^2 = \mathbb{E}\left(\eta^2_{k_n,0} - 1\right)^2 \mathbb{E}\check{\sigma}^4_{k_n,0}$ by independence of $\eta^2_{k_n,t-i}$ from $\mathcal{F}_{k_n,t-i-1}$. Combining (A.18) and (A.20) we conclude that

$$\sup_{t\geq 1} \left\|\mathbb{E}_{\mathcal{F}_{k_n,t-1-m}}\left(\sigma^2_{k_n,t} - \sigma^2_{k_n}\right)\right\|_{L_1} \leq b\psi_m + \tilde{\psi}_n$$

with $b = \sup_{n\geq 1}\left(\mathbb{E}\left(\eta^2_{k_n,0} - 1\right)^2 \mathbb{E}\check{\sigma}^4_{k_n,0}\right)^{1/2} < \infty$, $\psi_m = \left(\sum_{i>m} |\tilde{\alpha}_i|\right)^{1/2} \to 0$.

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

$$n(|\hat{\rho}_n| - 1) = n(|\rho_n| - 1)(1 - \epsilon_n), \quad \epsilon_n = \frac{|\hat{\rho}_n| - |\rho_n|}{1 - |\rho_n|} \tag{A.21}$$

for parts (i) and (ii) of the lemma. We first show that, under C(i) and C(iii),

$$\limsup_{n\to\infty} \mathbb{P}(\epsilon_n > 1 - \eta) = 0 \text{ for some } \eta \in (0,1). \tag{A.22}$$

The inequality $||x| - |y|| \leq |x - y|$ implies that $|\epsilon_n| \leq |\hat{\rho}_n - \rho_n| \,|1 - |\rho_n||^{-1} \to_p 0$ under C(iii) and Assumption 6 ($|\hat{\rho}_n - \rho_n|(\rho_n - 1)^{-1} = O_p\left(|\rho_n|^{-n}\right)$ by Theorem 1 of MP(2024)) as well as under C(i) and Assumption 5 ($|\hat{\rho}_n - \rho_n|(1 - |\rho_n|)^{-1} = O_p(n^{-1/2}(1 - |\rho_n|)^{-1/2})$), showing (A.22) for the above cases. It remains to prove (A.22) under C(i) and Assumption 6. In this case,

$$G_n := \left(1 - \rho^2_n\right)^{-1}(\hat{\rho}_n - \rho_n) \to_p \Gamma\left(\sigma^2 + 2\rho\Gamma\right)^{-1} \tag{A.23}$$

where $\Gamma$ is given in (15). To see this, the recursion for $x_{0t}$ in (A.1) yields

$$n^{-1}\left(1 - \rho^2_n\right)\textstyle\sum_{t=1}^{n} x^2_{0t-1} = n^{-1}\sum_{t=1}^{n} u^2_{n,t} + 2\rho_n n^{-1}\sum_{t=1}^{n} x_{0t-1}u_{n,t} - n^{-1}x^2_{0n} \to_p \sigma^2 + 2\rho\Gamma \tag{A.24}$$

since under Assumption 6: $n^{-1}\sum_{t=1}^{n} u^2_{n,t} \to_p \sigma^2$ by Lemma 1 of MP(2024); $n^{-1}\sum_{t=1}^{n} x_{0t-1}u_{n,t} \to_p \Gamma$ when $\rho \geq 0$ by Lemma 2.2(i) of MP(2020); when $\rho < 0$ $x^+_{n,t} := (-1)^t x_{0t}$ satisfies the regular autoregression (A.10) and $u^+_{n,t} := (-1)^t u_{n,t}$ satisfies Assumption 6 with ACV function

$\mathbb{E}u^+_{n,t}u^+_{n,t-k} = (-1)^{-k}\,\gamma_{u_n}(k)$, so

$n^{-1}\sum_{t=1}^n x_{0t-1}u_{n,t} = -n^{-1}\sum_{t=1}^n x^+_{n,t-1}u^+_{n,t} \to_p -\sum_{k=1}^\infty |\rho|^{k-1}(-1)^{-k}\gamma(k) = \sum_{k=1}^\infty \rho^{k-1}\gamma(k) = \Gamma$ by Lemma 2.2(i) of MP(2020). We have established that $n^{-1}\sum_{t=1}^n x_{0t-1}u_{n,t} \to_p \Gamma$ for $\rho \in [-1,1]$ under C(i) and Assumption 6; combining the last probability limit with (A.24) shows (A.23). Suppose first that $|\rho_n| \to 1$ under C(i) and Assumption 6; then

$$\begin{aligned}\epsilon_n &= \frac{|\rho_n|\left|1+\rho_n^{-1}\left(1-\rho_n^2\right)G_n\right| - |\rho_n|}{1-|\rho_n|} = \frac{|\rho_n|\left(1+\rho_n^{-1}\left(1-\rho_n^2\right)G_n\right) - |\rho_n|}{1-|\rho_n|} \\ &= \left(|\rho_n|/\rho_n\right)\left(1+|\rho_n|\right)G_n = 2\rho\Gamma/\left(\sigma^2+2\rho\Gamma\right) + o_p(1) \qquad \text{(A.25)}\end{aligned}$$

by (A.23) since $sign(\rho) = \rho \in \{-1,1\}$, where the second equality holds for all but finitely many $n$ since $\left(1-\rho_n^2\right)G_n = o_p(1)$. Since $\sigma^2 + 2\rho\Gamma > 0$, (A.22) follows immediately when $\rho = 1$ and $\Gamma \le 0$ and when $\rho = -1$ and $\Gamma \ge 0$ since the probability limit in (A.25) lies in $(-\infty, 0]$. When $\rho = 1$ and $\Gamma > 0$ or when $\rho = -1$ and $\Gamma < 0$, $\rho\Gamma > 0$, so the probability limit in (A.25) belongs to $(0,1)$ since $\sigma^2 > 0$. This completes the proof of (A.22) when $|\rho_n| \to 1$ under C(i). It remains to show (A.22) when $\rho_n \to \rho \in (-1,1)$. In this case, (A.23) implies that $\hat\rho_n \to_p b(\rho) := \rho + \left(1-\rho^2\right)\Gamma/\left(\sigma^2+2\rho\Gamma\right)$, so (A.21) gives $\epsilon_n \to_p \epsilon(\rho) := \left(|b(\rho)| - |\rho|\right)/\left(1-|\rho|\right)$ and (A.22) will follow by proving that $\epsilon(\rho) < 1$ which is equivalent to $|b(\rho)| < 1$. Since $\left(1-\rho^2\right)/\left(\sigma^2+2\rho\Gamma\right) > 0$ and $\rho \in (-1,1)$, $|b(\rho)| < 1$ will hold trivially if $\rho > 0$ and $\Gamma \le 0$ or if $\rho < 0$ and $\Gamma \ge 0$; it is therefore sufficient to prove $|b(\rho)| < 1$ when $\rho\Gamma > 0$. Now $|b(\rho)| < 1$ is equivalent to

$$-\sigma^2 < (1+\rho)\Gamma \ \text{ and } \ (1-\rho)\Gamma < \sigma^2. \qquad \text{(A.26)}$$

The left inequality in (A.26) holds trivially when $\Gamma \ge 0$ and the right when $\Gamma \le 0$; given the restriction $\rho\Gamma > 0$, it is enough to prove the left inequality in (A.26) for $\rho < 0$ and the right inequality in (A.26) for $\rho > 0$; for these values of $\rho$, (A.26) is equivalent to $\left(1-|\rho|\right)|\Gamma| < \sigma^2$. Since

$$\left(1-|\rho|\right)|\Gamma| \le \left(1-|\rho|\right)\sum_{k=1}^\infty |\rho|^{k-1}|\gamma(k)| < \sigma^2\left(1-|\rho|\right)\sum_{k=1}^\infty |\rho|^{k-1} = \sigma^2,$$

by (15) and $|\gamma(k)| < \sigma^2$, this shows (A.26), $|b(\rho)| < 1$ and completes the proof of (A.22).

For part (i), $\mathbf{1}_{\bar F_n^+} \le \mathbf{1}_{\bar F_n}$ and $\mathbf{1}_{\bar F_n^-} \le \mathbf{1}_{\bar F_n}$ so it is sufficient to show that $m_n\mathbf{1}_{\bar F_n} \to_p 0$. For some $\eta \in (0,1)$ that satisfies (A.22) and using (A.21), we obtain for any $\delta > 0$

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

for all $n \ge n_0(\delta)$ because $n\left(|\rho_n|-1\right)\eta \to -\infty$, so $\mathbf{1}\left\{n\left(|\rho_n|-1\right)\eta > 0\right\} = 0$ for all but finitely many $n$; since $\eta$ satisfies (A.22), part (i) follows. the proof of part (ii) is similar: for some $\eta$ satisfying (A.22) and any $\delta > 0$ we may write

$$\mathbb{P}\left(m_n\mathbf{1}_{F_n} > \delta\right) \le \mathbb{P}\left(m_n\mathbf{1}\left\{n\left(|\rho_n|-1\right)\eta \le 0\right\} > \delta\right) + \mathbb{P}\left(\epsilon_n > 1-\eta\right) = \mathbb{P}\left(\epsilon_n > 1-\eta\right)$$

for all $n \geq n_0(\delta)$ because $n(|\rho_n| - 1)\eta \to \infty$, so $\mathbf{1}\{n(|\rho_n| - 1)\eta \leq 0\} = 0$ for all but finitely many $n$; part (ii) follows since $\eta$ satisfies (A.22) and $\max\left(\mathbf{1}_{F_n^+}, \mathbf{1}_{F_n^-}\right) \leq \mathbf{1}_{F_n}$. For part (iii), $m_n \max\left(\mathbf{1}_{\bar{F}_n^-}, \mathbf{1}_{F_n^-}\right) \leq m_n \mathbf{1}\{\hat{\rho}_n < 0\}$; for arbitrary $\delta > 0$

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

for all $n \geq n_0(\delta)$, since $\rho_n \to \rho \geq 1$ so $\mathbf{1}\{\rho_n < 1/2\} = 0$ for all but finitely many $n$. Part (iii) follows since, by (A.23), $|\hat{\rho}_n - \rho_n| \to_p 0$ under Assumption 6 when $\rho_n \to \rho \geq 1$. For part (iv), $m_n \max\left(\mathbf{1}_{\bar{F}_n^+}, \mathbf{1}_{F_n^+}\right) \leq m_n \mathbf{1}\{\hat{\rho}_n \geq 0\}$ and the argument of part (iii) gives

$$\mathbb{P}(m_n \mathbf{1}\{\hat{\rho}_n \geq 0\} > \delta) \leq \mathbb{P}(m_n \mathbf{1}\{\rho_n > -1/2\} > \delta) + \mathbb{P}(|\hat{\rho}_n - \rho_n| \geq 1/2) \to 0$$

for all $\delta > 0$ when $\rho_n \to \rho \leq -1$.

**Proof of Lemma 3.** Recalling the definition of $x_{0t}$ in (A.1), denote by $\tilde{z}_{0t} = \sum_{j=1}^{t} \varphi_{1n}^{t-j} \Delta x_{0j}$ the restriction of the instrument $\tilde{z}_{1t}$ in (A.2) when $\mu = X_{n,0} = 0$. Applying (A.1), we obtain

$$\tilde{z}_{1t} - \tilde{z}_{0t} = \sum_{j=1}^{t} \varphi_{1n}^{t-j} (\Delta x_{n,j} - \Delta x_{0j}) = X_{n,0}(\rho_n - 1)\varphi_{1n}^{t-1} \sum_{j=0}^{t-1} (\rho_n/\varphi_{1n})^j \tag{A.27}$$

which yields the following decomposition for $\tilde{z}_{1t}$:

$$\tilde{z}_{1t} = \tilde{z}_{0t} + q_{nt}, \quad q_{nt} := X_{n,0}(\rho_n - 1)\left(\varphi_{1n}^t - \rho_n^t\right)/(\varphi_{1n} - \rho_n). \tag{A.28}$$

In the above form, (A.28) is useful when $\varphi_{1n} - \rho_n$ is not too small in that $n|\varphi_{1n} - \rho_n| \to \infty$; when $|\varphi_{1n} - \rho_n| = O(n^{-1})$ we apply the mean value theorem to the function $x \mapsto x^t$ between $\varphi_{1n}$ and $\rho_n$ to obtain $q_{nt} = (X_0(n) - \mu)(\rho_n - 1)t\phi_n^{t-1}$ where $|\phi_n - \varphi_{1n}| \leq |\varphi_{1n} - \rho_n| = O(n^{-1})$. The leading term $\tilde{z}_{0t}$ of (A.28) may be further decomposed as

$$\tilde{z}_{0t} = \begin{cases} z_{1t} - (1 - \rho_n)(\varphi_{1n} - \rho_n)^{-1}(z_{1t} - x_{0t}), & n|\varphi_{1n} - \rho_n| \to \infty \\ z_{1t} - (1 - \rho_n)\sum_{i=1}^{t-1} i\phi_n^{i-1} u_{n,t-i}, & n|\varphi_{1n} - \rho_n| = O(1) \end{cases} \tag{A.29}$$

where $z_{1t} := \sum_{j=1}^{t} \varphi_{1n}^{t-j} u_{n,j}$ for some $\phi_n$ satisfying $|\phi_n - \varphi_{1n}| \leq |\varphi_{1n} - \rho_n| = O(n^{-1})$. Writing $\Delta x_{0t} = u_{n,t} + (\rho_n - 1)x_{0t-1}$ from (A.1), we may obtain (A.29) as follows:

$$
\begin{aligned}
\tilde{z}_{0t} &= z_{1t} - (1 - \rho_n)\sum_{j=1}^{t} \varphi_{1n}^{t-j} \sum_{i=1}^{j-1} \rho_n^{j-1-i} u_{n,i} = z_{1t} - (1 - \rho_n)\sum_{i=1}^{t-1} \sum_{j=i+1}^{t} \varphi_{1n}^{t-j} \rho_n^{j-1-i} u_{n,i} \\
&= z_{1t} - (1 - \rho_n)\sum_{i=1}^{t-1} u_{n,t-i} \varphi_{1n}^{i-1} \sum_{j=0}^{i-1} (\rho_n/\varphi_{1n})^j. 
\end{aligned} \tag{A.30}
$$

When $n|\varphi_{1n} - \rho_n| \to \infty$, the sum in (A.30) is equal to $(\varphi_{1n} - \rho_n)^{-1}(\varphi_{1n} z_{1t-1} - \rho_n x_{0t-1})$ and (A.29) follows by adding and subtracting $u_{n,t}$ in $\varphi_{1n} z_{1t-1} - \rho_n x_{0t-1}$. When $n|\varphi_{1n} - \rho_n| = O(1)$ and $\rho_n = \varphi_{1n}$ (A.30) gives (A.29) with $\phi_n = 1$; when $\rho_n \neq \varphi_{1n}$, $(\varphi_{1n}^i - \rho_n^i)/(\varphi_{1n} - \rho_n) = i\phi_n^{i-1}$ by the MVT applied to the function $x \mapsto x^i$ between $\varphi_{1n}$ and $\rho_n$, and (A.29) follows from (A.30).

We first show that

$$r_{1n} := (1 - \rho_n \varphi_{1n}) n^{-1} \left(\sum_{t=1}^{n} x_{n,t-1} \tilde{z}_{n,t-1} - \sum_{t=1}^{n} x_{0t-1} \tilde{z}_{0t-1}\right) = o_p(1) \tag{A.31}$$

Since $|X_{n,0}| = O_p(1)$ and $1 - \rho_n \varphi_{1n} \asymp \lambda_{1n}^{-1}$, using the decompositions (A.1) and (A.28) we obtain

$$|r_{1n}| \le b\lambda_{1n}^{-1} n^{-1} \left\{ \left|\sum_{t=1}^{n-1} \rho_n^t \tilde{z}_{0t}\right| + \left|\sum_{t=1}^{n-1} \tilde{z}_{0t}\right| + \left|\sum_{t=1}^{n-1} x_{0t} q_{nt}\right| + 2\sum_{t=1}^{n-1} |q_{nt}| \right\} \tag{A.32}$$

for some $b > 0$. Since $\sum_{t=1}^{n-1} \tilde{z}_{0t} = O_p\left(n^{1/2}\lambda_{1n}\right)$ the second term of (A.32) is $O_p\left(n^{-1/2}\right)$; since $\sum_{t=1}^{n-1} |q_{nt}| \le O_p\left(\kappa_n^{-1}\right) \sum_{t=1}^{n-1} \left|\left(\varphi_{1n}^t - \rho_n^t\right) / \left(\varphi_{1n} - \rho_n\right)\right|$ the last term of (A.32) is $O_p\left(n^{-1}\kappa_n^{-1}\Lambda_{1n}\right)$ when $n|\varphi_{1n} - \rho_n| \to \infty$ and $O_p\left(\kappa_n^{-1}\right)\left(1 - \varphi_{1n}^2\right) n^{-1} \sum_{t=1}^{n-1} t\phi_n^{t-1} = O_p\left(\kappa_n^{-1} n^{-1} \left(1 - \phi_n\right)^{-1}\right)$ when $|\varphi_{1n} - \rho_n| = O\left(n^{-1}\right)$; since

$$\sum_{t=1}^{n-1} q_{nt}^2 = O_p(1)\, \kappa_n^{-2} \sum_{t=1}^{n-1} \left[\left(\varphi_{1n}^t - \rho_n^t\right) / \left(\varphi_{1n} - \rho_n\right)\right]^2 = O_p\left(\kappa_n^{-2} \lambda_{1n}^2 \Lambda_{1n}\right) \tag{A.33}$$

because the sum on the right is bounded by $2\left(\varphi_{1n} - \rho_n\right)^{-2} \sum_{t=1}^{n-1} \left(\varphi_{1n}^{2t} + \rho_n^{2t}\right) = O\left(\lambda_{1n}^2 \Lambda_{1n}\right)$ when $n|\varphi_{1n} - \rho_n| \to \infty$ and by $\sum_{t=1}^{n-1} t^2 \phi_n^{2(t-1)} = O\left(\kappa_n^3\right)$ when $|\varphi_{1n} - \rho_n| = O\left(n^{-1}\right)$ in which case $\kappa_n \asymp \Lambda_{1n} \asymp \lambda_{1n}$. Since $\sum_{t=1}^{n-1} x_{0t}^2 = O_p(n\kappa_n)$ by (A.24), the third term of (A.32) satisfies

$$\lambda_{1n}^{-1} n^{-1} \left|\sum_{t=1}^{n-1} x_{0t} q_{nt}\right| \le \lambda_{1n}^{-1} n^{-1} \left(\sum_{t=1}^{n-1} x_{0t}^2\right)^{1/2} \left(\sum_{t=1}^{n-1} q_{nt}^2\right)^{1/2} = O_p\left(n^{-1/2}\right). \tag{A.34}$$

Finally, for the first term of (A.32), $\sum_{t=1}^{n-1} \tilde{z}_{0t}^2 = O_p(n\lambda_n)$ by MP(2020), so

$$\lambda_{1n}^{-1} n^{-1} \left|\sum_{t=1}^{n-1} \rho_n^t \tilde{z}_{0t}\right| \le \lambda_{1n}^{-1} n^{-1} O\left(\kappa_n^{1/2}\right) \left(\sum_{t=1}^{n-1} \tilde{z}_{0t}^2\right)^{1/2} = O_p\left(n^{-1/2} \lambda_n^{-1/2} \kappa_n^{1/2}\right) = o_p(1)$$

completing the proof of (A.31). By (A.31) and equations (66)-(68) of MP(2020) we may write

$$\begin{aligned}(1 - \rho_n \varphi_{1n}) \frac{1}{n} \sum_{t=1}^n x_{n,t-1} \tilde{z}_{n,t-1} \;&=\; \sigma^2 + \frac{1}{n} \sum_{t=1}^n \tilde{z}_{0t-1} u_{n,t} + \frac{1}{n} (2\rho_n - 1) \sum_{t=1}^n x_{0t-1} u_{n,t} \\ &\quad + \rho_n (\rho_n - 1) \frac{1}{n} \sum_{t=1}^n x_{0t-1}^2 + o_p(1).\end{aligned} \tag{A.35}$$

Under Assumption 6 on $(u_{n,t})$, $n^{-1} \sum_{t=1}^n \tilde{z}_{0t-1} u_{n,t} = \Gamma_n + o_p(1)$ by Lemma 3.1(ii) of MP(2020).

Next, we analyse $\bar{z}_{1n-1} \bar{x}_{n-1}$. Letting $\bar{z}_{1n} = n^{-1} \sum_{t=1}^n \tilde{z}_{1t}$ and $\bar{\zeta}_n := n^{-1} \sum_{t=1}^n z_{1t}$, (A.29) gives

$$\bar{z}_{0n} = \bar{\zeta}_n - (1 - \rho_n)\left(\varphi_{1n} - \rho_n\right)^{-1} \left(\bar{\zeta}_n - \bar{x}_{0n}\right) = O_p\left(|\bar{x}_{0n}| \wedge \left|\bar{\zeta}_n\right|\right) = O_p\left(n^{-1/2} \lambda_{1n}\right) \tag{A.36}$$

since $\bar{\zeta}_n = O_p\left(n^{-1/2}\left(1 - \varphi_{1n}\right)^{-1}\right)$ and $\bar{x}_{0n} = O_p\left(n^{-1/2}\kappa_n\right)$. Denoting $\bar{q}_{nn} = n^{-1} \sum_{t=1}^n q_{nt}$, (A.1) and (A.28) imply that

$$\lambda_{1n}^{-1} \bar{x}_{n,n} \bar{z}_{1n} = \lambda_{1n}^{-1} \bar{z}_{0n} \bar{x}_{0n} + \lambda_{1n}^{-1} \bar{x}_{0n} \bar{q}_{nn} + O_p(1)\, \lambda_{1n}^{-1} \left(\bar{z}_{0n} + \bar{q}_{nn}\right) = \lambda_{1n}^{-1} \bar{z}_{0n} \bar{x}_{0n} + o_p(1) \tag{A.37}$$

under C$_+$(i)-C$_+$(ii), by (A.36), $\bar{x}_{0n} = O_p\left(n^{-1/2}\kappa_n\right)$ and $\bar{q}_{nn} = O_p\left(\kappa_n^{-1} n^{-1} \lambda_{1n} \Lambda_{1n}\right)$. Since $\lambda_{1n}^{-1} \bar{z}_{0n} \bar{x}_{0n} = O_p(\kappa_n / n)$ we conclude that $\lambda_{1n}^{-1} \bar{x}_{n,n} \bar{z}_{1n} = o_p(1)$ under C$_+$(i). Under C$_+$(ii), the recursion $\tilde{z}_{0t} = \varphi_{1n} \tilde{z}_{0t-1} + \Delta x_{0t}$ implies that $(1 - \varphi_{1n})\, \bar{z}_{0n-1} = x_{0n} - \tilde{z}_{0n}$, so (A.37) gives

$$(1 - \rho_n \varphi_{1n})\, \bar{z}_{1n-1} \bar{x}_{n-1} = (1 - \varphi_{1n})\, \bar{z}_{0n-1} \bar{x}_{0n-1} + o_p(1) = \frac{x_{0n}}{n^{1/2}} \frac{1}{n^{3/2}} \sum_{j=1}^n x_{0j-1} + o_p(1). \tag{A.38}$$

Combining (A.35)-(A.38) and using $\left(1 - \rho_n^2 \varphi_{1n}^2\right) / \left(1 - \rho_n \varphi_{1n}\right) \sim 1 + \rho_n$, we obtain that

$$\left(1 - \rho_n^2 \varphi_{1n}^2\right) \frac{1}{n} \sum_{t=1}^n \underline{x}_{n,t-1} \tilde{z}_{n,t-1} = \tilde{\Psi}_n + o_p(1) \tag{A.39}$$

with $\tilde{\Psi}_n$ defined in (A.6) under C$_+$(i)-C$_+$(ii), with the term in (A.38) being $o_p(1)$ under C$_+$(i). Under C$_+$(i), $n^{-1} \sum_{t=1}^n x_{0t-1} u_{n,t} \to_p \Gamma$ by Lemma 2.2(i) of MP(2020), so (A.24) implies that

$\tilde{\Psi}_n \to_p \sigma^2 + 2\rho\Gamma$ under $C_+$(i). Under $C_+$(ii), $\Gamma_n \to \lambda$ and

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

by standard local to unit asymptotics, e.g. Phillips (1987b), 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), we show that $r_{2n} := n^{-1}(1-\rho_n^2\varphi_{1n}^2)\left(\sum_{t=1}^n \tilde{z}_{1t}^2 - \sum_{t=1}^n \tilde{z}_{0t}^2\right) \to_p 0$. Using (A.28) we obtain $\sum_{t=1}^n \tilde{z}_{1t}^2 = \sum_{t=1}^n \tilde{z}_{0t}^2 + \sum_{t=1}^n q_{nt}^2 + 2\sum_{t=1}^n \tilde{z}_{0t}q_{nt}$ with (A.33) implying that $n^{-1}\lambda_{1n}^{-1}\sum_{t=1}^n q_{nt}^2 = O_p(n^{-1}\kappa_n^{-1}\Lambda_{1n}) = o_p(1)$ and $\lambda_{1n}^{-1}n^{-1}\left|\sum_{t=1}^{n-1}\tilde{z}_{0t}q_{nt}\right| = O_p\left(n^{-1/2}\right)$ by (A.34) since $\sum_{t=1}^{n-1}\tilde{z}_{0t}^2 = O_p\left(\sum_{t=1}^{n-1}x_{0t}^2\right)$, completing the proof of $r_{2n} = o_p(1)$. 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.24) 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, $(\rho_n)$ belongs to $C_+$(i) and equations (74) and (75) of MP(2020) imply that

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

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

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

since $\varphi_{1n} \to 1$ and $\rho_n \to 1$. For part (iii), $n^{-1/2}\lambda_{1n}^{-1/2}\sum_{t=1}^n q_{nt-1}e_{n,t} = O_p(n^{-1}\kappa_n^{-1}\Lambda_{1n})$ by (A.33), so (A.28) implies that it is sufficient to show the result for $\sum_{t=1}^n \xi_{nt}$ for the $\mathcal{F}_{n,t}$-martingale array $\xi_{nt} := n^{-1/2}(1-\rho_n^2\varphi_{1n}^2)^{1/2}\tilde{z}_{0t-1}e_{n,t}$. We will prove the following approximation under Assumption 6(i) or under Assumption 6(ii) with $\rho_n \to 1$:

$$\sum_{t=1}^n \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\xi_{nt}^2\right) = \sigma_e^2\left(1-\rho_n^2\varphi_{1n}^2\right)n^{-1}\sum_{t=1}^n \tilde{z}_{0t-1}^2 + o_p(1) \tag{A.40}$$

By part (ii) of the lemma, the right side of (A.40) satisfies $\sum_{t=1}^n \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\xi_{nt}^2\right) \to_p \sigma_e^2(\sigma^2 + 2\rho\Gamma)$. Under conditional homoskedasticity of $(e_{n,t})$, the equality in (A.40) holds exactly. Under Assumption 6(i), the Lindeberg condition (LC) follows by Lemma 3.2 of MP(2020) and $\sum_{t=1}^n \xi_{nt} \to_d \mathcal{N}(0, \sigma_e^2(\sigma^2 + 2\rho\Gamma))$ follows by a standard martingale CLT (e.g. Corollary 3.1 of HH(1980)). Under Assumption 6(ii) and $\rho_n \to 1$, consider a sequence $(l_n)$ satisfying $l_n \to \infty$ and $\lambda_{1n}^{-1}l_n^2 \to 0$ and denote the remainder in approximating $\tilde{z}_{0t}$ by $\tilde{z}_{0t-l_n}$ in (A.40) by

$$\begin{aligned} r_{nt} &= \tilde{z}_{0t} - \tilde{z}_{0t-l_n} = \sum_{j=t-l_n+1}^{t} \varphi_{1n}^{t-j} \Delta x_{0j} - \left(\varphi_{1n}^{l_n} - 1\right) \sum_{j=1}^{t-l_n} \varphi_{1n}^{t-l_n-j} \Delta x_{0j} \\ &= \sum_{i=0}^{l_n-1} \varphi_{1n}^{i} u_{n,t-i} + (\rho_n - 1) \sum_{i=0}^{l_n-1} \varphi_{1n}^{i} x_{0t-i-1} - \left(\varphi_{1n}^{l_n} - 1\right) \tilde{z}_{0t-l_n}. \end{aligned}$$

Since $\|u_{n,0}\|_{L_4}^2$, $\max_{t\le n} \left\|\kappa_n^{-1/2} x_{0t}\right\|_{L_4}^2$ and $\max_{t\le n} \left\|\lambda_{1n}^{-1/2} \tilde{z}_{0t}\right\|_{L_4}^2$ are all $O(1)$ we obtain

$$r_{nt}^4 \le 4^3 \left\{ l_n^2 \sum_{i=0}^{l_n-1} u_{n,t-i}^4 + (\rho_n - 1)^4 l_n^3 \sum_{i=0}^{l_n-1} x_{0t-i-1}^4 + \left(\varphi_{1n}^{l_n} - 1\right)^4 \tilde{z}_{0t-l_n}^4 \right\}$$

and $\lambda_{1n}^{-2} \max_{t\le n} \mathbb{E} r_{nt}^4 = O\left(\lambda_{1n}^{-2} l_n^3 + \lambda_{1n}^{-2} \kappa_n^{-2} l_n^4 + (\varphi_{1n}^{l_n} - 1)^4\right) = o(1)$ ($l_n(1-\varphi_{1n}^2) \to 0$ implies that $\varphi_{1n}^{l_n} \to 1$). Hence, $\tilde{z}_{0t}^2 = \tilde{z}_{0t-l_n}^2 + r_{nt}^2 + 2\tilde{z}_{0t-l_n} r_{nt}$ yields

$$\begin{aligned} n^{-1}\lambda_{1n}^{-1} \left\|\sum_{t=1}^{n} \left(\tilde{z}_{0t-1} - \tilde{z}_{0t-l_n-1}\right) \mathbb{E}_{\mathcal{F}_{n,t-1}} e_{n,t}^2 \right\|_{L_1} &= n^{-1}\lambda_{1n}^{-1} \left\|\sum_{t=1}^{n} \left(r_{nt-1}^2 + 2\tilde{z}_{0t-l_n-1} r_{nt-1}\right) \mathbb{E}_{\mathcal{F}_{n,t-1}} e_{n,t}^2 \right\|_{L_1} \\ &\le \left\|e_{n,0}^2\right\|_{L_4}^2 \lambda_{1n}^{-1} \max_{t\le n} \left(\|r_{nt}\|_{L_4}^2 + 2\|r_{nt}\|_{L_4} \|\tilde{z}_{0t}\|_{L_4}\right) = o(1) \end{aligned}$$

Hence, the left side of (A.40) becomes

$$\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{n,t-1}} \xi_{nt}^2 = \mathbb{E}\left(e_{n,t}^2\right) n^{-1} \left(1 - \rho_n^2 \varphi_{1n}^2\right) \sum_{t=1}^{n} \tilde{z}_{0t-l_n-1}^2 + n^{-1} \sum_{t=1}^{n} M_{n,t} + o_p(1)$$

where $M_{n,t} := (1 - \rho_n^2 \varphi_{1n}^2)\, \tilde{z}_{0t-l_n-1}^2 \mathbb{E}_{\mathcal{F}_{n,t-1}} \left[e_{n,t}^2 - \mathbb{E}\left(e_{n,t}^2\right)\right]$ is a uniformly integrable $L_1$-mixingale array with respect to $\mathcal{F}_{n,t-1}$:

$$\begin{aligned} \left\|\mathbb{E}_{\mathcal{F}_{n,t-l_n-1}}(M_{n,t})\right\|_{L_1} &= \left\|\left(1 - \rho_n^2 \varphi_{1n}^2\right) \tilde{z}_{0t-l_n-1}^2 \mathbb{E}_{\mathcal{F}_{n,t-l_n-1}} \left[e_{n,t}^2 - \mathbb{E}\left(e_{n,t}^2\right)\right]\right\|_{L_1} \\ &\le \left\|\mathbb{E}_{\mathcal{F}_{n,t-l_n-1}} \left[e_{n,t}^2 - \mathbb{E}\left(e_{n,t}^2\right)\right]\right\|_{L_2} \max_{t\le n} \left\|\lambda_{1n}^{-1/2} \tilde{z}_{0t}\right\|_{L_4}^2 \le b\left(\psi_{l_n} + \tilde{\psi}_n\right) \to 0 \end{aligned}$$

by (14). Proving UI of $(M_{n,t})$ is the same as proving UI of $\left(\lambda_n^{-1} \tilde{z}_{0t-1}^2 e_{n,t}^2\right)$: letting $\upsilon_n(x) = \mathbb{E} e_{n,0}^4 \mathbf{1}\left\{e_{n,0}^2 > x\right\}$,

$$\begin{aligned} \mathbb{E}\lambda_n^{-1} \tilde{z}_{0t-1}^2 e_{n,t}^2 \mathbf{1}\left\{\lambda_n^{-1} \tilde{z}_{0t-1}^2 e_{n,t}^2 > m_n\right\} &\le m_n^{1/4} \mathbb{E}\lambda_n^{-1} \tilde{z}_{0t-1}^2 \mathbf{1}\{\lambda_n^{-1} \tilde{z}_{0t-1}^2 > m_n^{3/4}\} + \mathbb{E}\lambda_n^{-1} \tilde{z}_{0t-1}^2 e_{n,t}^2 \mathbf{1}\{e_{n,t}^2 > m_n^{1/4}\} \\ &\le \max_{t\le n} \left\|\lambda_{1n}^{-1/2} \tilde{z}_{0t}\right\|_{L_4}^2 [m_n^{\frac{1}{4}} \max_{t\le n} \mathbb{P}(\lambda_n^{-1} \tilde{z}_{0t}^2 > m_n^{\frac{3}{4}}) + \upsilon_n^{1/2}(m_n^{\frac{1}{4}})]. \qquad \text{(A.41)} \end{aligned}$$

By the Markov inequality $m_n^{1/4} \max_{t\le n} \mathbb{P}\left(\lambda_n^{-1} \tilde{z}_{0t-1}^2 > m_n^{3/4}\right) \le m_n^{-1/2} \max_{t\le n} \left\|\lambda_{1n}^{-1/2} \tilde{z}_{0t}\right\|_{L_2}^2 \to 0$ for any $m_n \to \infty$; since $\mathbb{E} e_{n,0}^4 < \infty$, also $\upsilon_n\left(m_n^{1/4}\right) \to 0$ and the required UI follows since the right side of (A.41) is independent of $t$. A mixingale law of large numbers (e.g. Lemma 1 in MP(2024)) then implies that $\|n^{-1} \sum_{t=1}^{n} M_{n,t}\|_{L_1} \to 0$, completing the proof of (A.40) under Assumption 6(ii) when $\rho_n \to 1$. The LC $\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{n,t-1}} \left(\xi_{nt}^2 \mathbf{1}\left\{\xi_{nt}^2 > \delta\right\}\right) \to 0$ for any $\delta > 0$ under Assumption 6(ii) follows directly from the UI of $\left(\lambda_n^{-1} \tilde{z}_{0t-1}^2 e_{n,t}^2\right)$ proved in (A.41). This completes the proof of part (iii) under Assumption 6(ii) when $\rho_n \to 1$. When $\rho_n \to \rho \in (-1, 1)$, $\sum_{t=1}^{n} \xi_{nt} = n^{-1/2} (1 - \rho^2)^{1/2} \sum_{t=1}^{n} x_{0t-1} e_{n,t} + o_p(1)$ by MP (2020) and
$\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{n,t-1}} \xi_{nt}^2 = n^{-1} \left(1 - \rho^2\right) \sum_{t=1}^{n} x_{0t-1}^2 \mathbb{E}_{\mathcal{F}_{n,t-1}} e_{n,t}^2 + o_p(1) = \left(1 - \rho^2\right) \mathbb{E} x_{0t-1}^2 \mathbb{E}_{\mathcal{F}_{n,t-1}} e_{n,t}^2 + o_p(1)$,
by strict stationarity of $\left(u_{n,t}, \mathbb{E}_{\mathcal{F}_{n,t-1}} e_{n,t}^2\right)$ for each $n$ (implied by the strict stationarity of $\left(e_{n,t}, \mathbb{E}_{\mathcal{F}_{n,t-1}} e_{n,t}^2\right)$) where the last expectation does not depend on $t$ and is equal to $\upsilon_{n*}(\rho) = \mathbb{E}\left(\sum_{j=0}^{\infty} \rho^j u_{n,-j}\right)^2 e_{n,1}^2$ by $\mathcal{F}_{n,t-1}$-measurability of $x_{0t-1} = \sum_{j=0}^{\infty} \rho^j u_{n,t-1-j}$ and the law of iterated expectations. Since $\upsilon_{n*}(\rho) \to \upsilon_*(\rho)$ by assumption, $\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{n,t-1}} \xi_{nt}^2 \to_p (1 - \rho^2)\, \upsilon_*(\rho)$ as required. The LC follows

from the argument in (A.41) by setting $\lambda_{1n}^{-1} = 1$ and replacing $\tilde{z}_{0t-1}$ by $x_{0t-1} = \sum_{j=0}^{\infty} \rho^j u_{n,t-1-j}$. This concludes the proof of the lemma.

Finally, we prove an approximation used in the proof of Theorem 1: under Assumption 6

$$\sum_{t=1}^n \xi_{nt}^2 = \sum_{t=1}^n \mathbb{E}_{\mathcal{F}_{n,t-1}} \left(\xi_{nt}^2\right) + o_p(1). \tag{A.42}$$

Since the Lindeberg condition for $(\xi_{nt})$ has been verified above under Assumption 6 and the sequence $\left\{\sum_{t=1}^n \mathbb{E}_{\mathcal{F}_{n,t-1}} \xi_{nt}^2 : n \geq 1\right\}$ has been shown to converge in probability (and is tight as a consequence), (A.42) follows from Theorem 2.23 of HH.

**Proof of Lemma 4.** For $\delta > 0$, the Cauchy-Schwarz inequality implies that

$$\sum_{j=1}^{\infty} j^{\delta/4} |c_{n,j}| = \sum_{j=1}^{\infty} j^{-1/2-\delta/4} j^{1/2+\delta/2} |c_{n,j}| \leq \left(\sum_{i=1}^{\infty} i^{-1-\delta/2}\right)^{1/2} \left(\sum_{j=1}^{\infty} j^{1+\delta} c_{n,j}^2\right)^{1/2} \tag{A.43}$$

i.e. $\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} j^{\delta/4} |c_{n,j}| < \infty$ by Assumption 6. Hence, the fact that $(u_{n,t})$ satisfies Assumption 6 guarantees that $(u_{n,t})$ satisfies Assumption 2 and condition (13) of MP (2024). Since $z_{2t}$ in (A.5) is a mildly explosive process with root $\varphi_{2n}$ and innovation $u_{n,t}$, $[Y_n, Z_n] \to_d [Y, Z]$ and $[Y_n^{\varepsilon}, Z_n] \to_d [Y^{\varepsilon}, Z]$ follow from Lemma 2 of MP (2024) and the approximation

$$[\tau_n^{-1} \sum_{t=1}^n z_{2t-1} u_{n,t}, \tau_n^{-2} \sum_{t=1}^n z_{2t-1}^2] = [Y_n Z_n, Z_n^2] + o_p(1) \tag{A.44}$$

follows by Lemma 4 of MP(2024). We first prove that

$$s_n^{-1} \sum_{t=1}^n x_{n,t-1} z_{2t-1} = X_n Z_n + o_p(1). \tag{A.45}$$

The recursions for $x_{n,t}$ and $z_{2t}$ in (24) and (A.5) give

$$\begin{aligned}
(\rho_n \varphi_{2n} - 1) \sum_{t=1}^n x_{n,t-1} z_{2t-1} &= x_{n,n} z_{2n} - \varphi_{2n} \sum_{t=1}^n z_{2t-1} u_{n,t} - \rho_n \sum_{t=1}^n x_{n,t-1} u_{n,t} - \sum_{t=1}^n u_{n,t}^2 \\
&\quad + \varphi_{2n} \mu (\rho_n - 1) \sum_{t=1}^n z_{2t-1} + \mu (\rho_n - 1) \sum_{t=1}^n u_{n,t} \\
&= x_{n,n} z_{2n} + o_p(\nu_n \nu_{n,z})
\end{aligned} \tag{A.46}$$

because: $\nu_n^{-1} (\rho_n - 1) \sum_{t=1}^n u_{n,t}$ is of order $O_p(\rho_n^{-n} (n(\rho_n - 1))^{1/2}) = o_p(1)$ (by Lemma A1 of MP(2024)) under C$_+$(iii) and $O_p(n^{-1})$ under C$_+$(ii); by Lemma 4 and Lemma A1 of MP(2024): $\nu_n^{-1} \nu_{n,z}^{-1} \sum_{t=1}^n z_{2t-1} u_{n,t}$ is $O_p[\nu_n^{-1} (\varphi_{2n}^2 - 1)^{-1/2}] = o_p(\rho_n^{-n} [n(\rho_n - 1)]^{1/2}) = o_p(1)$ under C$_+$(iii) and $O_p[(n(\varphi_{2n} - 1))^{-1/2}]$ under C$_+$(ii) and $\nu_n^{-1} \nu_{n,z}^{-1} \sum_{t=1}^n x_{n,t-1} u_{n,t} = O_p(\varphi_{2n}^{-n} [n(\varphi_{2n} - 1)]^{1/2}) = o_p(1)$; $\sum_{t=1}^n z_{2t-1} = O_p\left((\varphi_{2n} - 1)^{-3/2} \varphi_{2n}^n\right)$, so $(\rho_n - 1) \nu_n^{-1} \nu_{n,z}^{-1} \sum_{t=1}^n z_{2t-1} = O_p(\rho_n^{-n} (\rho_n - 1)^{3/2} (\varphi_{2n} - 1)^{-1}) = o_p\left(n^{-1/2}\right)$ where we use that $\rho_n^{-n} n^{3/2} (\rho_n - 1)^{3/2} = O(1)$ ($o(1)$ under C$_+$(iii) by Lemma A1 of MP(2024)). This completes the proof of (A.46). By (A.46) we conclude that $s_n^{-1} \sum_{t=1}^n x_{n,t-1} z_{2t-1} = \nu_n^{-1} x_{n,n} \nu_{n,z}^{-1} z_{2n} + o_p(1)$ and (A.45) follows from the definitions of $Z_n$ in (A.7) and $X_n = x_{n,n}/\nu_n$.

Denoting $R_{1n} = \tau_n^{-1} \sum_{t=1}^n (\tilde{z}_{2t-1} - z_{2t-1}) u_{n,t}$, $R_{2n} = s_n^{-1} \sum_{t=1}^n x_{n,t-1} (\tilde{z}_{2t-1} - z_{2t-1})$ and $R_{3n} = \tau_n^{-2} \sum_{t=1}^n \left(\tilde{z}_{2t-1}^2 - z_{2t-1}^2\right)$, by (A.44) and (A.45), part (i) will follow from showing that $R_{1n} = o_p(1)$,

$R_{2n} = o_p(1)$ and $R_{3n} = o_p(1)$. Applying the identity $\hat{u}_t = \underline{u}_t - (\hat{\rho}_n - \rho_n)\underline{x}_{t-1}$ to $\tilde{z}_{2t} = \sum_{j=1}^{t} \varphi_{2n}^{t-j}\hat{u}_j$,

$$\begin{aligned}
\tilde{z}_{2t} &= \sum_{j=1}^{t} \varphi_{2n}^{t-j}\underline{u}_j - (\hat{\rho}_n - \rho_n)\sum_{j=1}^{t} \varphi_{2n}^{t-j}\underline{x}_{j-1} \\
&= z_{2t} - (\hat{\rho}_n - \rho_n)\sum_{j=1}^{t} \varphi_{2n}^{t-j}x_{0j-1} + (\hat{\rho}_n - \rho_n)(\bar{x}_{n-1} - \mu)\left(\varphi_{2n}^{t} - 1\right)/(\varphi_{2n} - 1) \\
&\quad - (\hat{\rho}_n - \rho_n)X_{n,0}\left(\rho_n^t - \varphi_{2n}^t\right)/(\rho_n - \varphi_{2n}) - \bar{u}_n\left(\varphi_{2n}^t - 1\right)/(\varphi_{2n} - 1) \qquad \text{(A.47)}
\end{aligned}$$

where the second equality follows by (A.1). The computation in (A.30) with $\varphi_{1n}$ replaced by $\varphi_{2n}$ on the second term of (A.47) gives the following decomposition:

$$\tilde{z}_{2t} = z_{2t} - (\hat{\rho}_n - \rho_n)q_{nt} + g_n\left(\varphi_{2n}^t - 1\right)/(\varphi_{2n} - 1), \quad \hat{\rho}_n - \rho_n = O_p\left(\kappa_n^{-1}\rho_n^{-n}\right) \qquad \text{(A.48)}$$

where $q_{nt} = (\rho_n - \varphi_{2n})^{-1}\left[\rho_n x_{0t-1} - \varphi_{2n}z_{2t-1} + X_{n,0}\left(\rho_n^t - \varphi_{2n}^t\right)\right]$, when $n|\varphi_{2n} - \rho_n| \to \infty$ and $q_{nt} = \sum_{i=1}^{t-1} i\phi_n^{i-1}u_{n,t-i} + X_{n,0}t\phi_n^{t-1}$ when $|\varphi_{2n} - \rho_n| = O(n^{-1})$ where $\phi_n$ is an intermediate point between $\varphi_{2n}$ and $\rho_n$ and $g_n = (\hat{\rho}_n - \rho_n)(\bar{x}_{n-1} - \mu) - \bar{u}_n$. The consistency rate for $\hat{\rho}_n$ in (A.48) under C$_+$(iii) under Assumption 6 follows from Theorem 1 in MP(2024). It is easy to see that $g_n = O_p\left(n^{-1/2}\right)$ under C$_+$(ii) and C$_+$(iii). We first show that

$$\sum_{t=1}^{n} q_{nt-1}^2 = O_p\left(\lambda_{2n}^2\kappa_n^2\rho_n^{2n}\right) + O_p\left(\lambda_{2n}^2\tau_n^2\right). \qquad \text{(A.49)}$$

When $n|\varphi_{2n} - \rho_n| \to \infty$

$$\sum_{t=1}^{n} q_{nt-1}^2 \leq 4(\rho_n - \varphi_{2n})^{-2}\left[\rho_n^2\sum_{t=1}^{n} x_{0t-1}^2 + \varphi_{2n}^2\sum_{t=1}^{n} z_{2t-1}^2 + X_{n,0}^2\left(\sum_{t=1}^{n}\rho_n^{2t} + \sum_{t=1}^{n}\varphi_{2n}^{2t}\right)\right]$$

with $(\rho_n - \varphi_{2n})^{-2} = O\left(\lambda_{2n}^2\right)$ and $\sum_{t=1}^{n} x_{0t-1}^2 = O_p\left(\kappa_n^2\rho_n^{2n}\right)$ and $\sum_{t=1}^{n} z_{2t-1}^2 = O_p\left(\tau_n^2\right)$ by Lemma 4 of MP(2024). When $|\varphi_{2n} - \rho_n| = O(n^{-1})$, $\lambda_{2n} \asymp \kappa_n$ and $\tau_n \asymp \kappa_n\rho_n^n$ so the right side of (A.49) is $O_p\left(\kappa_n^4\rho_n^{2n}\right)$ and $\sum_{t=1}^{n} q_{nt-1}^2 \leq 2\sum_{t=1}^{n}\left(\sum_{i=1}^{t-1} i\phi_n^{i-1}u_{n,t-i}\right)^2 + X_{n,0}\sum_{t=1}^{n} t^2\phi_n^{2(t-1)}$; since $\sum_{t=1}^{n} t^2\phi_n^{2(t-1)} = O(\phi_n^{2n}\left(\phi_n^2 - 1\right)^{-3}) = O\left(\rho_n^{2n}\kappa_n^3\right)$ and

$$\sum_{t=1}^{n}\mathbb{E}\left(\sum_{i=1}^{t-1} i\phi_n^{i-1}u_{n,t-i}\right)^2 = \sigma_n^2\sum_{t=1}^{n}\sum_{i=1}^{t-1} i^2\phi_n^{2(i-1)} \leq 2\sigma_n^2\left(\phi_n^2 - 1\right)^{-1}\sum_{t=1}^{n} t^2\phi_n^{2(t-1)} = O\left(\kappa_n^4\rho_n^{2n}\right),$$

(A.49) follows. We now show that $R_{1n} = o_p(1)$: by (A.48),

$$R_{1n} = -\tau_n^{-1}(\hat{\rho}_n - \rho_n)\sum_{t=1}^{n} q_{nt-1}u_{n,t} + \tau_n^{-1}g_n(\varphi_{2n} - 1)^{-1}\sum_{t=1}^{n}\left(\varphi_{2n}^{t-1} - 1\right)u_{n,t}.$$

By (A.49), $\sum_{t=1}^{n} q_{nt-1}u_{n,t} = O_p\left(\lambda_{2n}\kappa_n\rho_n^n\right) + O_p\left(\lambda_{2n}\tau_n\right)$ so, since $n\kappa_n^{-1}\rho_n^{-n} = O(1)$, the first term of $R_{1n}$ is $O_p\left(\varphi_{2n}^{-n}\right) + O_p\left(\lambda_{2n}/n\right)$; since $g_n = O_p\left(n^{-1/2}\right)$, the second term of $R_{1n}$ is $O_p(n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2}) = o_p(1)$, showing that $R_{1n} = o_p(1)$. To show that $R_{2n} = o_p(1)$, (A.48) gives

$$R_{2n} = -s_n^{-1}(\hat{\rho}_n - \rho_n)\sum_{t=1}^{n} x_{n,t-1}q_{nt-1} + s_n^{-1}g_n(\varphi_{2n} - 1)^{-1}\sum_{t=1}^{n} x_{n,t-1}\left(\varphi_{2n}^{t-1} - 1\right).$$

Since $\left|\sum_{t=1}^{n} x_{n,t-1}q_{nt-1}\right| \leq \left(\sum_{t=1}^{n} x_{n,t-1}^2\right)^{1/2}\left(\sum_{t=1}^{n} q_{nt-1}^2\right)^{1/2}$ and $|\hat{\rho}_n - \rho_n|\left(\sum_{t=1}^{n} x_{n,t-1}^2\right)^{1/2} = O_p(1)$, the first term of $R_{2n}$ is bounded by $O_p(1)s_n^{-1}\left(\sum_{t=1}^{n} q_{nt-1}^2\right)^{1/2} = O_p(\kappa_n^{1/2}\left(\varphi_{2n}^2 - 1\right)^{1/2}\varphi_{2n}^{-n}) + O_p(\kappa_n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2}\rho_n^{-n})$ by (A.49); by Lemma A1 of MP(2024) $\kappa_n^{1/2}\left(\varphi_{2n}^2 - 1\right)^{1/2}\varphi_{2n}^{-n} = o((\kappa_n/n)^{1/2})$ and $\kappa_n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2}\rho_n^{-n} = O(n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2})$, showing that first term of $R_{2n}$ is $o_p(1)$. For the second term of $R_{2n}$, a computation similar to (A.30) gives

$$s_n^{-1}\sum_{t=1}^{n} x_{0t-1}\varphi_{2n}^{t-1} = \nu_{n,z}^{-1}\nu_n^{-1}\left(\rho_n\varphi_{2n}^n x_{0n-1} - \sum_{i=1}^{n-1} u_{n,i}\varphi_{2n}^i\right) = O_p\left(\left(\varphi_{2n}^2 - 1\right)^{1/2}\right) + O_p\left(\rho_n^{-n}\kappa_n^{-1/2}\right).$$

Since $g_n = O_p\left(n^{-1/2}\right)$ and $\kappa_n^{-1/2}n^{1/2}\rho_n^{-n} = O(1)$, the second term of $R_{2n}$ is $O_p(n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2})$

completing the proof of $R_{2n} = o_p(1)$. To show that $R_{3n} = o_p(1)$, use (A.48) to write

$$R_{3n} = \tau_n^{-2} \sum_{t=1}^n h_{nt}^2 + 2\tau_n^{-2} \sum_{t=1}^n z_{2t-1} h_{nt} \leq \tau_n^{-2} \sum_{t=1}^n h_{nt}^2 + 2\left(\tau_n^{-2} \sum_{t=1}^{n-1} z_{2t}^2\right)^{1/2} \left(\tau_n^{-2} \sum_{t=1}^{n-1} h_{nt}^2\right)^{1/2}$$

where $h_{nt} := g_n \left(\varphi_{2n}^{t-1} - 1\right) / \left(\varphi_{2n} - 1\right) - \left(\hat{\rho}_n - \rho_n\right) q_{nt-1}$, satisfies

$$\tau_n^{-2} \sum_{t=1}^n h_{nt}^2 \leq 4\tau_n^{-2} g_n^2 \left(\varphi_{2n} - 1\right)^{-2} \left(\sum_{t=0}^{n-1} \varphi_{2n}^{2t} + n\right) + 2\tau_n^{-2} \left(\hat{\rho}_n - \rho_n\right)^2 \sum_{t=1}^n q_{nt-1}^2 = o_p(1)$$

since the first term is $O(n^{-1} (\varphi_{2n} - 1)^{-1})$ and the second term is $O_p\left(n^{-2} \lambda_{2n}^2\right)$ by (A.49) and (A.48). Since $\tau_n^{-2} \sum_{t=1}^n z_{2t}^2 = O_p(1)$ we conclude that $R_{3n} = o_p(1)$, completing the proof of part (i).

For part (ii), the martingale approximation of Lemma 2 of MP(2024) implies that

$$[Y_n, X_n]' = C(1) \sum_{j=1}^n b_{nj} e_{n,j} + o_p(1), \; b_{nj} = \left[\left(\varphi_{2n}^2 - 1\right)^{1/2} \varphi_{2n}^{-(n-j)-1}, \left(\rho_n^2 - 1\right)^{1/2} \rho_n^{-j}\right]' \quad \text{(A.50)}$$

We apply a standard martingale central limit theorem, e.g. Corollary 3.1 of HH(1980), to the martingale array in (A.50): the conditional variance matrix $V_n = \sum_{j=1}^n b_{nj} b_{nj}' \mathbb{E}_{\mathcal{F}_{n,j-1}}\left(e_{n,j}^2\right)$ has elements: $V_{11}^{(n)} = C(1)^2 \sigma_n^2 \left(\varphi_{2n}^2 - 1\right) \sum_{j=1}^n \varphi_{2n}^{-2j} \to \omega^2$; $V_{22}^{(n)} = C(1)^2 \sigma_n^2 \left(\rho_n^2 - 1\right) \sum_{j=1}^n \rho_n^{-2j} \to \omega^2$; $V_{12}^{(n)} = -C(1)^2 \sigma_n^2 \left(\rho_n^2 - 1\right)^{1/2} \left(\varphi_{2n}^2 - 1\right)^{1/2} \left(\varphi_{2n}^{-n} - \rho_n^{-n}\right) / \left(\varphi_{2n} - \rho_n\right)$. When $n\left|\rho_n - \varphi_{2n}\right| \to \infty$, $V_{12}^{(n)} = O\left(\varphi_{2n}^{-n}\right) + O\left(\rho_n^{-n}\right)$; when $\left|\rho_n - \varphi_{2n}\right| = O\left(n^{-1}\right)$, $\left(\varphi_{2n}^{-n} - \rho_n^{-n}\right) / \left(\varphi_{2n} - \rho_n\right) = -n\phi_n^{-n-1}$ for some $\phi_n$ in a $n^{-1}$-neighbourhood of $\rho_n$ and of $\varphi_{2n}$, so $V_{12}^{(n)} = O\left(n\left(\rho_n^2 - 1\right)\rho_n^{-n}\right) = o(1)$ by Lemma A1 of MP(2024). We conclude that $V_n \to \omega^2 I_2$ 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.50), the bound $\max_{j \leq n} \left\|b_{nj}\right\|^2 \leq 2\Lambda_{2n}$ (under $C_+$(iii), $\Lambda_{2n} = \left(\varphi_{2n}^2 - 1\right) \vee \left(\rho_n^2 - 1\right)$) yields

$$\sum_{j=1}^n \left\|b_{nj}\right\|^2 \mathbb{E}\left(e_{n,j}^2 \mathbf{1}\left\{\left\|b_{nj}\right\|^2 e_{n,j}^2 > \delta\right\}\right) \leq \max_{1 \leq j \leq n} \mathbb{E}\left(e_{n,j}^2 \mathbf{1}\left\{e_{n,j}^2 > \Lambda_{2n}^{-1} \delta / 2\right\}\right) \sum_{j=1}^n \left\|b_{nj}\right\|^2 \to 0$$

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

For part (iii), $\rho_n \to \rho > 1$ and (A.1) implies that $X_n = \left(\rho^2 - 1\right)\left(U_n + X_{n,0}\right) + o_p(1)$. Let $(k_n)_{n \in \mathbb{N}}$ be a sequence satisfying

$$k_n \to \infty \;\text{ and }\; k_n \left(\varphi_{2n}^2 - 1\right)^{1/4} \to 0 \quad \text{(A.51)}$$

(A.51) implies that $\left(n - k_n\right)\left(\varphi_{2n}^2 - 1\right) \to \infty$. Let $Y_n' = \left(\varphi_{2n}^2 - 1\right)^{1/2} C(1) \sum_{t=k_n+1}^n \varphi_{2n}^{-(n-t+1)} e_{n,t}$, $U_n' = \sum_{t=1}^{k_n-1} \rho^{-j} u_{n,j}$ and $X_n' = \left(\rho^2 - 1\right)\left(U_n' + X_{n,0}\right)$. It is easy to see that $\left\|Y_n - Y_n'\right\|_{L_2} = O\left(\varphi_{2n}^{-(n-k_n)}\right) = o(1)$, $\left\|U_n - U_n'\right\|_{L_2} = O\left(\rho^{-k_n}\right)$ and $\left|X_n - X_n'\right| = o_p(1)$. By Assumption 6, $\left(X_{n,0}, U_n\right)$ converges in distribution, so $\left(X_n\right)$ and $\left(X_n'\right)$ converge in distribution to the same limit denoted by $X_\infty$. By Lemma 3(ii) of MP(2024), a limit in distribution of any subsequence of $\left(X_n\right)$ is non-zero *a.s.*, so $\mathbb{P}\left(X_\infty = 0\right) = 0$. Since $\mathbb{P}\left(\left|X_\infty\right| \leq x\right)$ for all $x \leq 0$, 0 is a continuity point of $\mathbb{P}\left(\left|X_\infty\right| \leq \cdot\right)$); for any non-negative sequence $\left(\delta_n\right)_{n \in \mathbb{N}}$ such that $\delta_n \to 0$, $\mathbb{P}\left(Y_n / X_n \neq \mathbf{1}\left\{\left|X_n'\right| > \delta_n\right\} Y_n / X_n\right) \leq \mathbb{P}\left(\left|X_n'\right| \leq \delta_n\right) \to \mathbb{P}\left(\left|X_\infty\right| \leq 0\right) = 0$, so we may write

$$Y_n / X_n = Y_n' / X_n' \mathbf{1}\left\{\left|X_n'\right| > \delta_n\right\} + o_p(1) = \sum_{t=1}^{n-k_n} \xi_{n,t} + o_p(1) \quad \text{(A.52)}$$

where $\xi_{n,t} = C(1)\left(1 / X_n'\right) \mathbf{1}\left\{\left|X_n'\right| > \delta_n\right\} \alpha_{nt} e_{n,t+k_n}$, $\alpha_{nt} = \left(\varphi_{2n}^2 - 1\right)^{1/2} \varphi_{2n}^{-(n-k_n-t+1)}$. Since $X_n'$ is $\mathcal{F}_{n,k_n-1}$-measurable, $\left\{\left(\xi_{n,t}, \mathcal{F}_{n,t+k_n}\right) : 1 \leq t \leq n - k_n\right\}$ is a martingale difference array. We will show that quadratic variation of the martingale array in (A.52) satisfies

$$\sum_{t=1}^{n-k_n} \xi_{n,t}^2 = \omega^2 \left(X_n'\right)^{-2} \mathbf{1}\left\{|X_n'| > \delta_n\right\} + o_p(1) \to_d \omega^2 / X_\infty^2. \tag{A.53}$$

The proof of (A.53) follows by applying the law of large numbers of Lemma 1 of Arvanitis and Magdalinos (2019) (with a trivial adjustment to account for the extra $o(1)$ term $\tilde{\psi}_n$ in (14)) to $\sum_{t=1}^{n-k_n} \alpha_{nt}^2 \left(e_{n,t+k_n}^2 - \mathbb{E}e_{n,1}^2\right)$: since $\left(e_{n,t+k_n}^2 - \mathbb{E}e_{n,1}^2\right)_{t\geq 1}$ is a $L_1$-mixingale array, $\sup_{n\geq 1} \sum_{t=1}^{n-k_n} \alpha_{nt}^2 \leq 1$ and $\sum_{t=1}^{n-k_n} \alpha_{nt}^4 = O\left(\varphi_{2n}^2 - 1\right) = o(1)$, Lemma 1 of Arvanitis and Magdalinos (2019) implies that $\sum_{t=1}^{n-k_n} \alpha_{nt}^2 \left(e_{n,t+k_n}^2 - \mathbb{E}e_{n,1}^2\right) \to_{L_1} 0$, showing (A.53) since $\sum_{t=1}^{n-k_n} \alpha_{nt}^2 \to 1$, $\mathbb{E}e_{n,1}^2 \to \sigma_e^2$ and $\omega^2 = C(1)^2 \sigma_e^2$. Since convergence in (A.53) only applies in distribution, we employ the extended martingale CLT Theorem 3.4 of HH(1980). For the Lindeberg condition, for arbitrary $\eta > 0$ denote $\mathcal{L}_n(\eta) := \sum_{t=0}^{n-k_n} \mathbb{E}\left(\xi_{n,t}^2 \mathbf{1}\left\{\xi_{n,t}^2 > \eta\right\}\right)$, $\lambda_{n,\eta} := C(1)^{-2} \left(\varphi_{2n}^2 - 1\right)^{-1} \eta^2$ and $\upsilon_n(x) = \max_{1\leq t\leq n} \mathbb{E}\left(e_{n,t}^2 \mathbf{1}\left\{e_{n,t}^2 > x\right\}\right)$. Then

$$\mathcal{L}_n(\eta) \leq C(1)^2 \delta_n^{-2} \max_{1\leq t\leq n} \mathbb{E}\left(e_{n,t}^2 \mathbf{1}\left\{e_{n,t}^2 > \lambda_{n,\eta}\delta_n^2\right\}\right) \leq C(1)^2 \delta_n^{-2} \upsilon_n\left(\lambda_{n,\eta}^{1/2}\right)$$

by choosing $\delta_n^2 \geq \lambda_{n,\eta}^{-1/2}$; since $\lambda_{n,\eta} \to \infty$ for all $\eta > 0$, $\upsilon_n(\lambda_{n,\eta}^{1/2}) \to 0$ by UI of $\left(e_{n,t}^2\right)$, and we may choose $\delta_n^2 = \max\left\{\lambda_{n,\eta}^{-1/2}, \upsilon_n^{1/2}(\lambda_{n,\eta}^{1/2})\right\}$ which implies that $\mathcal{L}_n(\eta) \to 0$ for all $\eta > 0$, proving the LC for (A.52). We now verify the assumptions of Theorem 3.4 of HH(1980) for the martingale array in (A.52). The inequality $\mathbb{E}\left(\max_{1\leq t\leq n-k_n} \xi_{n,t}^2\right) \leq \eta^2 + \mathcal{L}_n(\eta)$ for all $\eta > 0$, shows that the unconditional LC $\mathcal{L}_n(\eta) \to 0$ for all $\eta > 0$ implies the negligibility conditions $\max_{1\leq t\leq n-k_n} \left|\xi_{n,t}\right| \to_p 0$ and $\sup_{n\geq 1} \mathbb{E}\left(\max_{1\leq t\leq n-k_n} \left|\xi_{n,t}\right|\right)^2 < \infty$. Letting $\mathcal{G}_n := \mathcal{F}_{n,k_n-1}$, $\mathcal{G}_n \subseteq \mathcal{F}_{n,t+k_n}$ for all $t \geq 1$; recalling that $\xi_{n,t}$ in (A.52) is $\mathcal{F}_{n,t+k_n}$-adapted, the first part of (A.53) and $\mathcal{G}_n$-measurability of $X_n'$ implies condition (3.28) of HH(1980); since $\mathcal{G}_{n,t+k_n} := \sigma\left(\mathcal{F}_{n,t+k_n} \cup \mathcal{G}_n\right) = \mathcal{F}_{n,t+k_n}$ for all $t \geq 1$, $\mathbb{E}_{\mathcal{G}_{n,t+k_n-1}}\left(\xi_{n,t}\right) = 0$ so condition (3.29) of HH(1980) holds trivially. Finally, the second part of (A.53) and Theorem 3.4 of HH(1980) allow us to conclude that $\sum_{t=0}^{n-k_n} \xi_{n,t} \to_d \mathcal{MN}\left(0, \omega^2/X_\infty\right)$, completing the proof of the statement for $Y_n/X_n$. The statement for $Y_n^\varepsilon/X_n$ follows by an identical argument by replacing $C(1)\, e_{n,t}$ by $\varepsilon_{n,t}$.

**Proof of Lemma 5.** Denote $\xi_{nt} = \left[\xi_{1,nt}, \xi_{2,nt}, \xi_{3,nt}, \xi_{4,nt}\right]'$ with $\xi_{1,nt} = (n\left(1 - \varphi_{1n}^2\right)^{-1})^{-1/2} z_{1t-1} e_{n,t}$, $\xi_{2,nt} = C(1)\, n^{-1/2} e_{n,t}$, $\xi_{3,nt} = C(1)\left(\varphi_{2n}^2 - 1\right)^{1/2} \varphi_{2n}^{-(\lfloor ns\rfloor - t)-1} e_{n,t}$ and $\xi_{4,nt} = C(1)\left(\varphi_{2n}^2 - 1\right)^{1/2} \varphi_{2n}^{-t} e_{n,t}$. The martingale approximation of Lemma 2 of MP(2024) for $Y_n(s)$ and $Z_n(s)$ and a standard approximation for $B_n(s)$ give

$$\left[U_n(s), B_n(s), Y_n(s), Z_n(s)\right]' = \sum_{t=1}^{\lfloor ns\rfloor} \xi_{nt} + o_p(1). \tag{A.54}$$

Since $z_{1t-1}$ is $\mathcal{F}_{n,t-1}$-measurable, $\xi_{nt}$ is a $\mathcal{F}_{n,t}$-martingale difference array and we may apply a Lindeberg-type functional CLT for vector-valued martingale difference arrays to (A.54): see Theorem 3.33 (pp. 478) of Jacod and Shiryaev (2003). The conditional Lindeberg condition (LC) on $\left\|\xi_{nt}\right\|^2$ (3.31 in Jacod and Shiryaev (2003)) is implied by the stronger unconditional LC on $\left\|\xi_{nt}\right\|^2$ which, in turn, is implied by establishing the LC on each of $\xi_{1,nt}^2, ..., \xi_{4,nt}^2$. The LC for $\xi_{1,nt}^2$ under

Assumption 6(i) is established by Proposition A1 and Lemma 3.3 of MP(2020); under Assumption 6(ii), $z_{1t}$ is the restriction of $\tilde{z}_{0t}$ when $\rho_n = 1$ so the LC follows from the LC in the proof of part (iii) of Lemma 3. The LC for $\xi_{j,nt}^2$ $j \in \{2,3,4\}$ follows from the bound

$$\sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}\left(\xi_{j,nt}^2 \mathbf{1}\left\{\xi_{j,nt}^2 > \delta\right\}\right) \leq C(1)^2 \max_{1\leq t\leq n} \mathbb{E}\left(e_{n,t}^2 \mathbf{1}\left\{e_{n,t}^2 > \lambda_n(\delta)^2\right\}\right), \quad \delta > 0 \tag{A.55}$$

where $\lambda_n(\delta) = n^{1/2}\delta^{1/2}C(1)^{-1}$ when $j = 2$ and $\lambda_n(\delta) = C(1)^{-1}(\varphi_{2n}^2 - 1)^{-1/2}\delta^{1/2}$ when $j \in \{3,4\}$. Since $\lambda_n(\delta) \to \infty$ for any $\delta > 0$ the LC for (A.54) follows by uniform integrability of $\left(e_{n,t}^2\right)_{t\in\mathbb{N}}$. Denote $\tilde{\sigma}_{n,t}^2 := \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(e_{n,t}^2\right)$. The conditional variance matrix of the array in (A.54) is given by $V^{(n)} := \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\xi_{nt}\xi_{nt}'\right)$ with typical elements denoted by $[V_{ij}^{(n)}]_{i,j=1}^4$: $V_{11}^{(n)} = n^{-1}\left(1-\varphi_{1n}^2\right)\sum_{t=1}^{\lfloor ns \rfloor} z_{1t-1}^2\tilde{\sigma}_{n,t}^2 \to_p \sigma_e^2\omega^2 s$, by the LLN (A.40) since $z_{1t}$ is the restriction of $\tilde{z}_{0t}$ when $\rho_n = 1$; $V_{22}^{(n)} = n^{-1}C(1)^2\sum_{t=1}^{\lfloor ns \rfloor}\tilde{\sigma}_{n,t}^2 \to \omega^2 s$ trivially under Ass. 6(i) and by the mixingale property (14) under Ass. 6(ii); $\left[V_{33}^{(n)}, V_{44}^{(n)}\right] = C(1)^2\left(\varphi_{2n}^2 - 1\right)\sum_{t=1}^{\lfloor ns \rfloor}\left[\varphi_{2n}^{-2(\lfloor ns \rfloor - t)-1}, \varphi_{2n}^{-2t}\right]\tilde{\sigma}_{n,t}^2 \to \omega^2[1,1]$ for all $s > 0$, trivially under Ass. 6(i) and using the mixingale LLN of Lemma 1 of Arvanitis and Magdalinos (2019) with $y_t = \tilde{\sigma}_{n,t}^2 - \mathbb{E}\left(e_{n,t}^2\right)$ under Ass. 6(ii); $\left[V_{23}^{(n)}, V_{24}^{(n)}\right] = C(1)^2 n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{1/2}\sum_{t=1}^{\lfloor ns \rfloor}\left[\varphi_{2n}^{-(\lfloor ns \rfloor - t)-1}, \varphi_{2n}^{-t}\right]\tilde{\sigma}_{n,t}^2$ satisfy $\left\|V_{2j}^{(n)}\right\|_{L_1} \leq O(n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2})$ for $j \in \{3,4\}$; by Lemma A1 of MP(2024), $V_{34}^{(n)} = C(1)^2\left(\varphi_{2n}^2 - 1\right)\varphi_{2n}^{-\lfloor ns \rfloor - 1}\sum_{t=1}^{\lfloor ns \rfloor}\tilde{\sigma}_{n,t}^2$ satisfies

$$\left\|V_{34}^{(n)}\right\|_{L_1} = C(1)^2\,\mathbb{E}\left(e_{n,0}^2\right)\left(\varphi_{2n}^2 - 1\right)\lfloor ns \rfloor\varphi_{2n}^{-\lfloor ns \rfloor} \to 0.$$

$\left[V_{13}^{(n)}, V_{14}^{(n)}\right] = C(1)^2\left(\varphi_{2n}^2 - 1\right)^{1/2}\left(1 - \varphi_{1n}^2\right)^{1/2} n^{-1/2}\sum_{t=1}^{\lfloor ns \rfloor}\left[\varphi_{2n}^{-(\lfloor ns \rfloor - t+1)}, \varphi_{2n}^{-t}\right] z_{1t-1}\tilde{\sigma}_{n,t}^2$ satisfies

$$\left\|V_{1j}^{(n)}\right\|_{L_1} \leq C(1)^2 \max_{t\leq n}\left\|\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(e_{n,t}^2\right)\right\|_{L_2}\left\|\left(1-\varphi_{1n}^2\right)^{1/2} z_{1t}\right\|_{L_2} n^{-1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2} = o(1)$$

for $j \in \{3,4\}$; finally, $V_{12}^{(n)} = C(1)^2 n^{-1}\left(1 - \varphi_{1n}^2\right)^{1/2}\sum_{t=1}^{\lfloor ns \rfloor} z_{1t-1}\tilde{\sigma}_{n,t}^2$. Under Ass 6(i), $V_{12}^{(n)} = o_p(1)$, since $\sum_{t=1}^{\lfloor ns \rfloor} z_{1t-1} = O_p(n^{1/2}\left(1 - \varphi_{1n}^2\right)^{-1})$. Under Ass 6(ii), write $r_{nt} = z_{1t} - z_{1t-l_n} = \sum_{j=t-l_n+1}^{t}\varphi_{1n}^{t-j}u_{n,j} - \left(\varphi_{1n}^{l_n} - 1\right)\sum_{j=1}^{t-l_n}\varphi_{1n}^{t-l_n-j}u_{n,j}$ for some sequence $(l_n)$ satisfying $l_n \to \infty$ and $l_n\left(1 - \varphi_{1n}^2\right)^{1/2} \to 0$, and note that $\left\|\left(1 - \varphi_{1n}^2\right)^{1/2} r_{nt}\right\|_{L_2} \leq B\left(\left(1 - \varphi_{1n}^2\right)^{1/2} l_n + \left|\varphi_{1n}^{l_n} - 1\right|\right) = o(1)$. Denoting $\tilde{V}_{12}^{(n)}$ as $V_{12}^{(n)}$ with $z_{1t-1}$ replaced by $z_{1t-l_n-1}$, the above bound for $r_{nt}$ and the CS inequality imply that $\left\|\tilde{V}_{12}^{(n)} - V_{12}^{(n)}\right\|_{L_1} \to 0$. Now

$$\tilde{V}_{12}^{(n)} = C(1)^2 n^{-1}\left(1 - \varphi_{1n}^2\right)^{1/2}\sum_{t=1}^{\lfloor ns \rfloor} z_{1t-l_n-1}\left(\mathbb{E}_{\mathcal{F}_{n,t-l_n-1}}\left(e_{n,t}^2\right) - \mathbb{E}e_{n,t}^2\right) + O_p\left(n^{-1/2}\left(1 - \varphi_{1n}^2\right)^{-1/2}\right)$$

which is $o_p(1)$ by applying the mixingale law of large numbers used in (A.40), showing $V_{12}^{(n)} = o_p(1)$. We conclude that $V^{(n)} \to_p diag\left(\sigma_e^2\omega^2 s, \omega^2 s, \omega^2, \omega^2\right)$ for $s \in [0,1]$, and applying Theorem 3.33 of Jacod and Shiryaev (2003) to (A.54), $\sum_{t=1}^{\lfloor ns \rfloor}\xi_{nt} \Rightarrow \xi(s)$ where $\xi(s)$ is a continuous Gaussian martingale with quadratic variation $\langle\xi\rangle_s = diag\left(\sigma_e^2\omega^2 s, \omega^2 s, \omega^2, \omega^2\right)$. By Levy's characterisation (e.g. Theorem 4.4 II of Jacod and Shiryaev (2003), $\xi(s)$ is characterised by its quadratic variation process, $\xi(s) =_d [U(s), B(s), Y, Z]'$ with the right side defined in the statement of the lemma and independence between the components of $\xi(s)$ implied by the diagonality of the quadratic variation matrix $\langle\xi\rangle_s$.

**Proof of Lemma 6.** Applying $\underline{x}_{n,t}^{+}=(-1)^{-t}\,\underline{x}_{n,t}$ to $\hat{\rho}_n$ yields the identity $\hat{\rho}_n=-\hat{\rho}_n^{+}$ and

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

and $\bar{F}_n^{-}=\left\{n\left(\hat{\rho}_n^{+}-1\right)>0\right\}\cap\left\{\hat{\rho}_n^{+}>0\right\}=\bar{F}_n^{++}$. For the OLS residuals, $\hat{\rho}_n=-\hat{\rho}_n^{+}$ yields

$$(-1)^{-t}\,\hat{u}_{n,t}=(-1)^{-t}\,\underline{x}_{n,t}+\hat{\rho}_n(-1)^{-(t-1)}\,\underline{x}_{n,t-1}=\underline{x}_{n,t}^{+}-\hat{\rho}_n^{+}\underline{x}_{n,t-1}^{+}=\hat{u}_{n,t}^{+}$$

which implies that $(-1)^t\,\tilde{z}_{2t}^{-}=\sum_{j=1}^{t}\left|\varphi_{2n}^{-}\right|^{t-j}\hat{u}_{n,j}^{+}=\tilde{z}_{2t}^{+}$. Hence, the denominator and numerator of $\tilde{\rho}_{2n}^{-}-\rho_n$ can be written as: $\sum_{t=1}^{n}\underline{x}_{n,t-1}\tilde{z}_{2t-1}^{-}=\sum_{t=1}^{n}\underline{x}_{n,t-1}^{+}\tilde{z}_{2t-1}^{+}$ and $\sum_{t=1}^{n}\tilde{z}_{2t-1}^{-}\underline{u}_{n,t}=-\sum_{t=1}^{n}\tilde{z}_{2t-1}^{+}(-1)^{-t}\,\underline{u}_{n,t}$; $\tilde{\rho}_{2n}^{-}-\rho_n=-\left(\tilde{\rho}_{2n}^{+}-|\rho_n|\right)$ now follows from (A.10). The same argument shows that $\tilde{\beta}_{2n}^{-}-\beta=-\left(\tilde{\beta}_{2n}^{+}-\beta\right)$. By (A.10), the instrument $\tilde{z}_{1t}^{-}=\sum_{j=1}^{t}\left(\varphi_{1n}^{-}\right)^{t-j}\nabla\underline{x}_{n,j}$ in (A.2) can be written as

$$(-1)^{-t}\,\tilde{z}_{1t}^{-}=\sum_{j=1}^{t}\left|\varphi_{1n}^{-}\right|^{t-j}\left(x_{n,j}^{+}-x_{n,j-1}^{+}\right)-2\left(1-\varphi_{1n}^{-}\right)^{-1}(\bar{x}_{n,n}-\mu)(-1)^{-t}\left(1-\left(\varphi_{1n}^{-}\right)^{t}\right)=:\tilde{z}_{1t}^{+}-2g_n$$

where, since $\left(1-\varphi_{1n}^{-}\right)^{-1}\to 1/2$, $|g_n|\leq|\bar{x}_{n,n}-\mu|=O_p\left(n^{-1/2}\right)$ because (1) implies that $\bar{x}_{n,n}-\mu=\rho_n(\bar{x}_{n,n-1}-\mu)+\bar{u}_{n,n}$ or

$$\bar{x}_{n,n}-\mu=(1-\rho_n)^{-1}(\bar{u}_{n,n}-\rho_n x_{n,n}/n)=O_p\left(n^{-1/2}\right)$$

under C$_-$(i)-C$_-$(ii) since $(1-\rho_n)^{-1}\in[1/2,1]$. Since $g_n\sum_{t=1}^{n}\underline{u}_{n,t}=O_p(1)$ and $g_n\sum_{t=1}^{n}\underline{x}_{n,t-1}^{+}=O_p(\kappa_n)$, $\pi_n^{-1}\sum_{t=1}^{n}\tilde{z}_{1t-1}^{-}\underline{u}_{n,t}=-\pi_n^{-1}\sum_{t=1}^{n}\tilde{z}_{1t-1}^{+}(-1)^{-t}\,\underline{u}_{n,t}+O_p(\pi_n^{-1})$ and $\pi_n^{-2}\sum_{t=1}^{n}\underline{x}_{n,t-1}\tilde{z}_{1t-1}^{-}=\pi_n^{-2}\sum_{t=1}^{n}\underline{x}_{n,t-1}^{+}\tilde{z}_{1t-1}^{+}+O_p(\pi_n^{-2}\kappa_n)$, which implies that $\pi_n\left(\tilde{\rho}_{1n}^{-}-\rho_n\right)=-\pi_n\left(\tilde{\rho}_{1n}^{+}-|\rho_n|\right)+o_p(1)$ under C$_-$(i)-C$_-$(ii). The proof of $\pi_n\left(\tilde{\beta}_{1n}^{-}-\beta\right)=-\pi_n\left(\tilde{\beta}_{1n}^{+}-\beta\right)+o_p(1)$ follows similarly.

**Proof of Theorem 1.** It will be convenient to define

$$[N_n,N_{n,\varepsilon}]=\left(\sum_{t=1}^{n}\tilde{z}_{n,t-1}^{2}\right)^{-1/2}\left[\sigma^{-1}\sum_{t=1}^{n}\tilde{z}_{n,t-1}\underline{u}_{n,t},\sigma_{\varepsilon}^{-1}\sum_{t=1}^{n}\tilde{z}_{n,t-1}\underline{\varepsilon}_{n,t}\right]\tag{A.56}$$

and to denote by $[N_{jn},N_{jn,\varepsilon}]$ the expression in (A.56) with $\tilde{z}_{n,t-1}$ replaced by $\tilde{z}_{jt-1}$ and by $\left[N_{jn}^{-},N_{jn,\varepsilon}^{-}\right]$ the expression in (A.56) with $\tilde{z}_{n,t-1}$ replaced by $\tilde{z}_{jt-1}^{-}$ for $j\in\{1,2\}$.

Under C$_+$(i)-C$_+$(ii), (A.28) and (A.36) give $\bar{z}_{1n}=O_p\left[n^{-1}\lambda_{1n}\left(n^{1/2}+\kappa_n^{-1}\Lambda_{1n}\right)\right]$. Since

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

with $1-\rho_n^2\varphi_{1n}^2\asymp\lambda_{1n}^{-1}$, $\bar{u}_n=O_p\left(n^{-1/2}\right)$ and the above order for $\bar{z}_{1n}$ imply that $n^{-1/2}\lambda_{1n}^{-1/2}n\bar{z}_{1n-1}\bar{u}_n$ is of order $O_p(n^{-1/2}\lambda_{1n}^{1/2})+O_p(n^{-1}\Lambda_{1n}\kappa_n^{-1/2})=o_p(1)$; similarly for $\tilde{\beta}_{1n}$, $n^{-1/2}\lambda_{1n}^{-1/2}n\bar{z}_{1n-1}\bar{\varepsilon}_n=o_p(1)$. By Lemma 3(ii), the common denominator of $\pi_n(\tilde{\rho}_{1n}-\rho_n)$ and $\pi_n(\tilde{\beta}_{1n}-\beta)$ is asymptotically equivalent to $\tilde{\Psi}_n$ in (A.6) we obtain, under C$_+$(i)-C$_+$(ii),

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left[\tilde{\rho}_{1n}-\rho_n,\tilde{\beta}_{1n}-\beta\right]=[1+o_p(1)]\,\tilde{\Psi}_n^{-1}\left[\tilde{U}_n(1),\tilde{U}_n^{\varepsilon}(1)\right]\tag{A.57}$$

where $\tilde{U}_n(\cdot)$ is defined as $U_n(\cdot)$ in Lemma 5 with $z_{1t-1}$ replaced by $\tilde{z}_{1t-1}$ (and $e_{n,t}=u_{n,t}$ under Assumption 5) and $\tilde{U}_n^{\varepsilon}(\cdot)$ as $\tilde{U}_n(\cdot)$ with $e_{n,t}$ replaced by $\varepsilon_{n,t}$.

Under C$_+$(i) and Assumption 5, $u_{n,t}=e_{n,t}$ and $\Gamma=0$ so $\tilde{\Psi}(c)=\sigma^2$ and $\tilde{U}_n(1)\to_d\mathcal{N}(0,v(\rho))$, where, by Lemma 3(iii), $v(\rho)=\sigma^4$ when $\rho_n\to 1$ or when $\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(u_{n,t}^2\right)=\sigma^2$; under Assumption

5(ii), $v(\rho) = (1-\rho^2) v_1(\rho)$ when $\rho_n \to \rho \in (-1,1)$. Substituting into (A.57) yields

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}(\tilde{\rho}_{1n}-\rho_n) \to_d \mathcal{N}\left(0, v(\rho)/\sigma^4\right) \tag{A.58}$$

with the asymptotic variance in (A.58) being equal to 1 when $\rho_n \to 1$ or when $\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(u_{n,t}^2\right) = \sigma^2$, in which case also $N_{1n} \to_d \mathcal{N}(0,1)$ by Lemma 2(i) and parts (ii) and (iii) of Lemma 3. For $\tilde{\beta}_n$ under $\mathrm{C}_+$(i) and Assumption 6, $\tilde{\Psi}(c) = \sigma^2 + 2\rho\Gamma$ by Lemma 3(i) and $\tilde{U}_n^{\varepsilon}(1) \to_d \mathcal{N}(0, v_\varepsilon(\rho))$ where, by Lemma 3(iii) with $e_{n,t}$ replaced by $\varepsilon_{n,t}$, $v_\varepsilon(\rho) = (\sigma^2+2\rho\Gamma)\sigma_\varepsilon^2$ when $\rho_n \to 1$ or when $\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\varepsilon_{n,t}^2\right) = \sigma_\varepsilon^2$, in which case also $N_{1n,\varepsilon} \to_d \mathcal{N}(0,1)$ for the same reason as $N_{1n}$; under Assumption 6(ii), $v_\varepsilon(\rho) = (1-\rho^2) v_2(\rho)$ when $\rho_n \to \rho \in (-1,1)$. We conclude that

$$n^{1/2}\left(1-\rho_n^2\varphi_{1n}^2\right)^{-1/2}\left(\tilde{\beta}_{1n}-\beta\right) \to_d \mathcal{N}\left(0, v_\varepsilon(\rho)/\left(\sigma^2+2\rho\Gamma\right)\right). \tag{A.59}$$

We now prove part (a) of the theorem for $\tilde{\rho}_n$ under C(i): by Lemma 2(i),

$$[\pi_n(\tilde{\rho}_n-\rho_n), N_n] = \left[\pi_n(\tilde{\rho}_{1n}-\rho_n)\mathbf{1}_{F_n^+} + \pi_n\left(\tilde{\rho}_{1n}^- - \rho_n\right)\mathbf{1}_{F_n^-}, N_{1n}\mathbf{1}_{F_n^+} + N_{1n}^-\mathbf{1}_{F_n^-}\right] + o_p(1) \tag{A.60}$$

under Assumption 6. When $\rho_n \to 1$, Lemma 2(iii) implies that $\pi_n(\tilde{\rho}_n-\rho_n) = \pi_n(\tilde{\rho}_{1n}-\rho_n) + o_p(1) \to_d \mathcal{N}(0, v(\rho)/\sigma^4)$ by (A.58) and $N_n = N_{1n} + o_p(1) \to_d \mathcal{N}(0,1)$. When $\rho_n \to -1$, Lemma 2(iv) and Lemma 6 imply that

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

since $\hat{\rho}_n^+ \to_p |\rho_n| = 1$ implies that $\mathbf{1}_{F_n^{++}} \to_p 1$. From its definition in Lemma 6, $\tilde{\rho}_{1n}^+$ is an IV estimator generated by the (regular) $\mathrm{C}_+$(i) autoregression (A.10) and the regular near-stationary instrument $\tilde{z}_{1t}^+$ of Lemma 6; hence, (A.58) implies that $\pi_n\left(\tilde{\rho}_{1n}^+ - \rho_n\right) \to_d \mathcal{N}(0,1)$, showing that $\pi_n(\tilde{\rho}_n-\rho_n) \to_d \mathcal{N}(0,1)$ when $\rho_n \to -1$. Similarly, $N_n = N_{1n}^-\mathbf{1}_{F_n^-} + o_p(1) = -N_{1n}^+\mathbf{1}_{F_n^{++}} + o_p(1)$

$$N_{1n}^+ = -\mathbf{1}_{F_n^{++}}\left(\sum_{t=1}^n \tilde{z}_{1,t-1}^{+2}\right)^{-1/2}\sigma^{-1}\sum_{t=1}^n \tilde{z}_{1t-1}^+(-1)^{-t}u_{n,t} + o_p(1) \to_d \mathcal{N}(0,1), \tag{A.61}$$

showing that $N_n \to_d \mathcal{N}(0,1)$ when $\rho_n \to -1$. To complete the proof of part (a) for $\tilde{\rho}_n$, it remains to deal with the stationary case $\rho_n \to \rho \in (-1,1)$, where under Assumption 5, $\pi_n \sim n^{1/2}(1-\rho_n^2)^{-1/2}$, $\mathbf{1}_{F_n^+} \to_p \mathbf{1}_{\{\rho\geq 0\}}$ and $\mathbf{1}_{F_n^-} = \mathbf{1}_{F_n^{++}} \to_p \mathbf{1}_{\{\rho<0\}}$ by Lemma 2 and the consistency of $\hat{\rho}_n$; (A.60) and Lemma 6 then yield

$$\pi_n(\tilde{\rho}_n-\rho_n) = \pi_n(\tilde{\rho}_{1n}-\rho_n)\mathbf{1}_{\{\rho\geq 0\}} - \pi_n\left(\tilde{\rho}_{1n}^+ - |\rho_n|\right)\mathbf{1}_{\{\rho<0\}} + o_p(1) \to_d \mathcal{N}\left(0, v(\rho)/\sigma^4\right)$$

by (A.58) and $N_n = N_{1n}\mathbf{1}_{\{\rho\geq 0\}} - N_{1n}^+\mathbf{1}_{\{\rho<0\}} \to_d \mathcal{N}(0,1)$ by (A.61) when $\sigma_{n,t}^2 = \sigma_n^2$. This completes the proof of part (a) for $\tilde{\rho}_n$. The proof of part (a) for $\tilde{\beta}_n$ follows a similar argument by replacing $\left(\tilde{\rho}_n, \tilde{\rho}_{1n}, \tilde{\rho}_{1n}^-, \tilde{\rho}_{1n}^+\right)$ by $(\tilde{\beta}_n, \tilde{\beta}_{1n}, \tilde{\beta}_{1n}^-, \tilde{\beta}_{1n}^+)$ and $\left(N_n, N_{1n}, N_{1n}^-, N_{1n}^+\right)$ with $\left(N_{n,\varepsilon}, N_{1n,\varepsilon}, N_{1n,\varepsilon}^-, N_{1n,\varepsilon}^+\right)$: Lemma 2 shows that when $\rho_n \to \rho \in (-1,1)$, $\hat{\rho}_n \to_p b(\rho) := \rho + (1-\rho^2)\Gamma/(\sigma^2+2\rho\Gamma)$ under Assumption 6, which implies that $\mathbf{1}_{F_n^+} \to_p \mathbf{1}_{\{b(\rho)\geq 0\}}$ and $\mathbf{1}_{F_n^-} = \mathbf{1}_{F_n^{++}} \to_p \mathbf{1}_{\{b(\rho)<0\}}$. Hence,

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

since both $\pi_n(\tilde{\beta}_{1n}-\beta)$ and $\pi_n(\tilde{\beta}_{1n}^+-\beta)$ have the same limit distribution given in (A.59) and $N_{n,\varepsilon} = N_{1n,\varepsilon}\mathbf{1}_{\{b(\rho)\geq 0\}} - N_{1n,\varepsilon}^+\mathbf{1}_{\{b(\rho)<0\}} \to_d \mathcal{N}(0,1)$ when $\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\varepsilon_{n,t}^2\right) = \mathbb{E}(\varepsilon_n^2)$. This completes the proof of part (a) of the theorem.

Under $C_+$(ii)-$C_+$(iii), by MP(2024), $\sum_{t=1}^n z_{2t} = O_p\left(\left(\varphi_{2n}-1\right)^{-3/2}\varphi_{2n}^n\right)$ and $\hat{\rho}_n - \rho_n = O_p\left(\kappa_n^{-1}\rho_n^{-n}\right)$, so summing (A.48), we obtain

$$\sum_{t=1}^n \tilde{z}_{2t} = \sum_{t=1}^n z_{2t} + O_p\left(n^{-1/2}\left(\varphi_{2n}-1\right)^{-2}\varphi_{2n}^n\right) = O_p\left(\left(\varphi_{2n}-1\right)^{-3/2}\varphi_{2n}^n\right). \tag{A.62}$$

We conclude that $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left(n\bar{z}_{2n-1}\bar{u}_n\right) = O_p\left(n^{-1/2}\left(\varphi_{2n}-1\right)^{-1/2}\right) = o_p(1)$. For $\tilde{\beta}_{2n}$, the same order applies to $\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left(n\bar{z}_{2n-1}\bar{\varepsilon}_n\right) = o_p(1)$. The above and (A.9) 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

$$\left(\varphi_{2n}^2-1\right)\varphi_{2n}^{-n}\left[\sum_{t=1}^n \tilde{z}_{2t-1}u_{n,t}, \sum_{t=1}^n \tilde{z}_{2t-1}\varepsilon_{n,t}\right] = \left[Y_n Z_n, Y_n^{\varepsilon} Z_n\right] + o_p(1). \tag{A.63}$$

Since $\bar{x}_{n-1} = O_p\left(n^{-1}\kappa_n^{3/2}\rho_n^n\right)$ by MP(2024), (A.62) implies that

$$ns_n^{-1}\bar{x}_{n-1}\bar{z}_{2n-1} = \left[1+o_p(1)\right]\frac{\kappa_n}{n}\nu_{n,z}^{-1}\left(\rho_n\varphi_{2n}-1\right)\sum_{t=1}^n z_{2t-1}\frac{\nu_n^{-1}}{\kappa_n}\sum_{j=1}^n x_{n,j-1} \tag{A.64}$$

which is $o_p(1)$ under $C_+$(iii): $O_p\left(\kappa_n/n\right)$ if $\left(\rho_n-1\right)/\left(\varphi_{2n}-1\right) \to 0$ and $O_p(\left(\left(\varphi_{2n}-1\right)^{-1}/n\right)$ if $\left(\varphi_{2n}-1\right)/\left(\rho_n-1\right) = O(1)$. Under $C_+$(ii), (A.64) becomes $Z_n n^{-3/2}\sum_{j=1}^n x_{n,j-1} + o_p(1)$ by (A.62), showing that (A.64) contributes asymptotically under $C_+$(ii). Combining the above with the approximation of $s_n^{-1}\sum_{t=1}^n x_{n,t-1}z_{2t-1}$ 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

$$s_n^{-1}\sum_{t=1}^n \underline{x}_{n,t-1}z_{2t-1} = Z_n\underline{X}_n + o_p(1), \quad \underline{X}_n := X_n - n^{-3/2}\sum_{j=1}^n x_{n,j-1} \tag{A.65}$$

under $C_+$(ii)-$C_+$(iii) and Assumption 6, where $Z_n$ is defined in (A.7) and $X_n = x_{n,n}/\nu_n$. Recalling the definition of $s_n$ and noting that $\rho_n\varphi_{2n}-1 \sim \varphi_{2n}-1$ under $C_+$(ii), the normalisation under $C_+$(ii)-$C_+$(iii) becomes $s_n/(\left(\varphi_{2n}^2-1\right)^{-1}\varphi_{2n}^n) = \pi_n$ in (26). Combining (A.65), (A.63) and (A.9),

$$\pi_n\left[\tilde{\rho}_{2n}-\rho_n, \tilde{\beta}_{2n}-\beta\right] = \frac{1}{\underline{X}_n}\left[Y_n, Y_n^{\varepsilon}\right] + o_p(1), \quad \left[N_{2n}, N_{2n,\varepsilon}\right] = \frac{Z_n}{|Z_n|}\left[\frac{Y_n}{\sigma}, \frac{Y_n^{\varepsilon}}{\sigma_{\varepsilon}}\right] + o_p(1) \tag{A.66}$$

under $C_+$(ii)-$C_+$(iii) and Assumption 6. We now prove part (c): under $C_+$(iii), $\underline{X}_n = X_n + o_p(1)$ and applying parts (ii) and (iii) of Lemma 4 and the continuous mapping theorem to (A.66),

$$\pi_n\left[\tilde{\rho}_{2n}-\rho_n, \tilde{\beta}_{2n}-\beta\right] \to_d \left[Y/X, Y^{\varepsilon}/X\right] \text{ and } N_{2n}, N_{2n,\varepsilon} \to_d \mathcal{N}(0,1) \tag{A.67}$$

where $X \neq 0$ *a.s.* (by Lemma 3(ii) of MP(2024) when $\rho > 1$), $X =_d \mathcal{N}(0,\omega^2)$ when $\rho_n \to 1$, $X$ is independent of $(Y, Y^{\varepsilon})$ and $Y =_d \mathcal{N}(0,\sigma^2)$, $Y^{\varepsilon} =_d \mathcal{N}(0,\sigma_{\varepsilon}^2)$. To justify the second part of (A.67), Lemma 4(i) implies that $\left(Z_n/|Z_n|\right)\sigma^{-1}Y_n \to_d sign(Z)\,\sigma^{-1}Y =_d \mathcal{N}(0,1)$ because

$$\begin{aligned}\mathbb{P}\left(sign(Z)\,\sigma^{-1}Y \leq x\right) &= \mathbb{P}\left(\sigma^{-1}Y \leq x\right)\mathbb{P}(Z>0) + \mathbb{P}\left(-\sigma^{-1}Y \leq x\right)\mathbb{P}(Z<0) \\ &= \Phi(x)\left[\mathbb{P}(Z>0)+\mathbb{P}(Z<0)\right] = \Phi(x)\end{aligned} \tag{A.68}$$

since $Z$ is independent of $Y$ and $\sigma^{-1}Y =_d \mathcal{N}(0,1)$ by Lemma 4 and $\mathbb{P}(Z=0) = 0$ by Gaussianity. Under Assumption 5, $\omega^2 = \sigma^2$, so $Y/X =_d \mathcal{MN}(0,\sigma^2/X^2)$; under Assumption 6, $Y^{\varepsilon}/X =_d \mathcal{MN}(0,\sigma_{\varepsilon}^2/X^2)$ with $X \neq 0$ *a.s.* by Gaussianity when $\rho_n \to 1$ and by Lemma 3(ii) of MP(2024) when $\rho_n \to \rho > 1$. Thus, (A.67) gives the correct limit distributions for $C_+$(iii), the theorem under $C_+$(iii) follows from the asymptotic equivalences $\pi_n\left(\tilde{\rho}_n-\tilde{\rho}_{2n}\right) = o_p(1)$, $\pi_n(\tilde{\beta}_n-\tilde{\beta}_{2n}) = o_p(1)$ and $\left[N_n, N_{n,\varepsilon}\right] = \left[N_{2n}, N_{2n,\varepsilon}\right] + o_p(1)$ by applying parts (ii) and (iii) of Lemma 2 to (9) and (10). Under

C$_-$(iii), parts (ii) and (iv) of Lemma 2 imply that $\pi_n(\tilde{\rho}_n - \rho_n) = \pi_n\left(\tilde{\rho}_{2n}^- - \rho_n\right)\mathbf{1}_{\bar{F}_n^-} + o_p(1)$, $\pi_n(\tilde{\beta}_n - \beta) = \pi_n(\tilde{\beta}_{2n}^- - \beta)\mathbf{1}_{\bar{F}_n^-} + o_p(1)$ and $[N_n, N_{n,\varepsilon}] = \left[N_{2n}^-, N_{2n,\varepsilon}^-\right]\mathbf{1}_{\bar{F}_n^-} + o_p(1)$; by Lemma 6, $\pi_n\left[\tilde{\rho}_{2n}^- - \rho_n, \tilde{\beta}_{2n}^- - \beta\right]\mathbf{1}_{\bar{F}_n^-} = -\pi_n\left[\tilde{\rho}_{2n}^+ - |\rho_n|, \tilde{\beta}_{2n}^+ - \beta\right]\mathbf{1}_{\bar{F}_n^{++}} = -\pi_n\left[\tilde{\rho}_{2n}^+ - |\rho_n|, \tilde{\beta}_{2n}^+ - \beta\right] + o_p(1)$ since $\mathbf{1}_{\bar{F}_n^{++}} \to_p 1$ by Lemma 2 and $\left[N_{2n}^-, N_{2n,\varepsilon}^-\right]\mathbf{1}_{\bar{F}_n^-} = -sign\left(Z_n^+\right)\left[Y_n^+/\sigma, Y_n^{\varepsilon+}/\sigma_\varepsilon\right]\mathbf{1}_{\bar{F}_n^{++}} + o_p(1)$, where $[Y_n^+, Y_n^{\varepsilon+}, Z_n^+]$ are defined as $[Y_n, Y_n^\varepsilon, Z_n]$ in (A.7) with $\varphi_{2n}$, $u_{n,t}$ and $\varepsilon_{n,t}$ replaced by $\left|\varphi_{2n}^-\right|$, $(-1)^{-t}u_{n,t}$ and $(-1)^{-t}\varepsilon_{n,t}$. Since $\tilde{\rho}_{2n}^+$ is generated by the regular C$_+$(iii) autoregression (A.10) and the regular mildly explosive instrument $\tilde{z}_{2t}^+$, (A.67) implies that $\pi_n(\tilde{\rho}_n - \rho_n) \to_d -Y/X =_d Y/X$ and $\pi_n\left(\tilde{\beta}_n - \beta\right) \to_d -Y^\varepsilon/X =_d Y^\varepsilon/X$ by the symmetry of $\mathcal{MN}(0, \sigma^2/X^2)$ and $\mathcal{MN}(0, \sigma_\varepsilon^2/X^2)$ around 0. For $N_n$, Lemma 2(ii) implies that

$$N_n = N_{2n}\mathbf{1}_{\bar{F}_n^+} + N_{2n}^-\mathbf{1}_{\bar{F}_n^-} + o_p(1) \to_d sign(Z)\,Y/\sigma\mathbf{1}_{\{\rho\geq 0\}} - sign\left(Z^+\right)Y^+/\sigma\mathbf{1}_{\{\rho<0\}} =_d \mathcal{N}(0,1)$$

since $(Z^+, Y^+) =_d (Z, Y)$; similarly $N_{n,\varepsilon} \to_d \mathcal{N}(0,1)$. This proves part (c) of the theorem.

We prove part (b) of the theorem under Assumption C$_+$(ii). In the notation of (A.57) and Lemma 5, $\left|\tilde{U}_n(1) - U_n(1)\right| = o_p(1)$, because $n^{-1/2}\left(1-\varphi_{1n}^2\right)^{1/2}\sum_{t=1}^n q_{nt-1}e_{n,t} = O_p\left(n^{-1}\right)$ by the approximation leading to (A.40) and $n^{-1/2}\left(1-\varphi_{1n}^2\right)^{1/2}\sum_{t=1}^n\left(\tilde{z}_{0t-1} - z_{1t-1}\right)e_{n,t} = o_p(1)$ under C$_+$(ii) by Lemma 3.2(i) of MP(2020). Using Lemma 2(iii) and combining (9), (A.57) and (A.66) and recalling (26) and the above approximation for $\tilde{U}_n(1)$, we obtain

$$\begin{aligned}\pi_n(\tilde{\rho}_n - \rho_n) &= n^{1/2}\left(1-\varphi_{1n}^2\right)^{-1/2}(\tilde{\rho}_{1n} - \rho_n)\mathbf{1}_{F_n^+} + 2n^{1/2}\left(\varphi_{2n}^2 - 1\right)^{-1/2}(\tilde{\rho}_{2n} - \rho_n)\mathbf{1}_{\bar{F}_n^+} + o_p(1) \\ &= \mathbf{1}_{F_n}U_n(1)/\tilde{\Psi}_n + \mathbf{1}_{\bar{F}_n}Y_n(1)/\underline{X}_n + o_p(1), \qquad\qquad (A.69)\end{aligned}$$

where $U_n(\cdot)$ and $Y_n(\cdot)$ are defined in Lemma 5 (with $u_{n,t} = e_{n,t}$ under Assumption 5) and the last line follows since $\left|\mathbf{1}_{F_n^+} - \mathbf{1}_{F_n}\right| \to_p 0$ and $\left|\mathbf{1}_{\bar{F}_n^+} - \mathbf{1}_{\bar{F}_n}\right| \to_p 0$ by Lemma 2(iii). $\tilde{\Psi}_n$ in (A.6), $n^{-1/2}x_n$, $\mathbf{1}_{F_n}$ and $\mathbf{1}_{\bar{F}_n}$ are functionals of $B_n(s) = n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor}u_{n,t}$, on $D[0,1]$, so the functional CLT of Lemma 5 on $[U_n(s), B_n(s), Y_n(s)]$ and the continuous mapping theorem imply that

$$\mathbf{1}_{F_n}U_n(1)/\tilde{\Psi}_n + \mathbf{1}_{\bar{F}_n}Y_n(1)/\underline{X}_n \to_d \mathbf{1}_{F_c}U(1)/\left(\omega^2\Psi_1(c)\right) + \mathbf{1}_{\bar{F}_c}Y/(\omega\Psi_2(c)) \qquad (A.70)$$

since, by Lemma 3(i), $\tilde{\Psi}_n \to_d \tilde{\Psi}(c)$ with $\sigma^2 + 2\rho\Gamma = \omega^2$ under C$_+$(ii), $\tilde{\Psi}(c) = \omega^2\Psi_1(c)$ on the event $F_c$ and $2\left(J_c(1) - \int_0^1 J_c(r)\,dr\right) = 2\omega\Psi_2(c)$ on the event $\bar{F}_c$. The continuous mapping theorem is applicable to (A.70) because $x = 0$ is the only discontinuity point of the function $x \mapsto \mathbf{1}_{(-\infty,0]}(x)$ and $\mathbb{P}(K_c + c = 0) = 0$ since $K_c$ in (27) is a continuously distributed random variable for all $c \in \mathbb{R}$. Denoting $\zeta := \left[\sigma^{-2}U(1), \sigma^{-1}Y\right]'$, Lemma 5 implies that $\zeta$ is independent of $\mathcal{F}_B = \sigma(B(s) : s \in [0,1])$ and $\zeta =_d \mathcal{N}(0, I_2)$. Since the random variables $J_c(1)$, $\Psi(c)$, $\mathbf{1}_{F_c}$ and $\mathbf{1}_{\bar{F}_c}$ are $\mathcal{F}_B$-measurable (as non-stochastic functionals of $B(r)$ on $D[0,1]$) the independence of $\zeta$ and $\mathcal{F}_B$ implies the independence of the random vectors $\zeta$ and $\left[J_c(1), \tilde{\Psi}(c), \mathbf{1}_{F_c}, \mathbf{1}_{\bar{F}_c}\right]'$. Under Assumption 5, $\omega^2 = \sigma^2$ and we conclude that the limit in (A.70) is given by $\left[\Psi_1^{-1}(c)\mathbf{1}_{F_c}, \Psi_2^{-1}(c)\mathbf{1}_{\bar{F}_c}\right]\zeta$ has a $\mathcal{MN}(0, V_1)$ distribution with $V_1 = \Psi_1^{-2}(c)\mathbf{1}_{F_c} + \Psi_2^{-2}(c)\mathbf{1}_{\bar{F}_c} = \Psi^{-2}(c)$ as required by the theorem for $\pi_n(\tilde{\rho}_n - \rho_n)$. For $\pi_n(\tilde{\beta}_n - \beta)$, defining $U_n^\varepsilon(s)$ and $Y_n^\varepsilon(s)$ as $U_n(s)$ and $Y_n(s)$ with $e_{n,t}$ replaced by $\varepsilon_{n,t}$, the same argument applies with $U_n(s)$ and $Y_n$ replaced by $U_n^{\varepsilon}(s)$ and $Y_n^{\varepsilon}$ in (A.69):

$$\pi_n\left(\tilde{\beta}_n-\beta\right)\to_d\left(\omega^2\Psi_1(c)\right)^{-1}U^{\varepsilon}(1)\mathbf{1}_{F_c}+\left(\omega\Psi_2(c)\right)^{-1}Y^{\varepsilon}\mathbf{1}_{\bar{F}_c}=\left(\sigma_{\varepsilon}/\omega\right)V_1\tilde{\zeta}\qquad\text{(A.71)}$$

by Lemma 5, where $\tilde{\zeta}=_d\mathcal{N}(0,I_2)$, establishing the limit distribution under $\mathrm{C}_+$(ii). Under $\mathrm{C}_-$(ii), Lemma 2(iv) and Lemma 6 imply that

$$\begin{aligned}\pi_n\left(\tilde{\rho}_n-\rho_n\right) &= \pi_n\left(\tilde{\rho}_{1n}^{-}-\rho_n\right)\mathbf{1}_{F_n^-}+\pi_n\left(\tilde{\rho}_{2n}^{-}-\rho_n\right)\mathbf{1}_{\bar{F}_n^-}+o_p(1)\\ &= -\left[\pi_n\left(\tilde{\rho}_{1n}^{+}-|\rho_n|\right)\mathbf{1}_{F_n^{++}}+\pi_n\left(\tilde{\rho}_{2n}^{+}-|\rho_n|\right)\mathbf{1}_{\bar{F}_n^{++}}\right]+o_p(1)\\ &= -\pi_n\left(\tilde{\rho}_n^{+}-|\rho_n|\right)+o_p(1)\end{aligned}\qquad\text{(A.72)}$$

where $\tilde{\rho}_{1n}^{+}$ and $\tilde{\rho}_{2n}^{+}$ are defined in Lemma 6, $\tilde{\rho}_n^{+}:=\tilde{\rho}_{1n}^{+}\mathbf{1}_{F_n^{++}}+\tilde{\rho}_{2n}^{+}\mathbf{1}_{\bar{F}_n^{++}}$ is a IV estimator generated by the (regular) $\mathrm{C}_+$(ii) autoregression (A.10) and the regular instrument $\tilde{z}_t^{+}:=\tilde{z}_{1t}^{+}\mathbf{1}_{F_n^{++}}+\tilde{z}_{2t}^{+}\mathbf{1}_{\bar{F}_n^{++}}$ and the last equality in (A.72) holds exactly since $\mathbf{1}_{F_n^{++}}+\mathbf{1}_{\bar{F}_n^{++}}=1$. Under Assumption 6, an FCLT on the innovation sequence of (A.10) continues to hold: $B_n^{+}(s):=n^{-1/2}\sum_{j=1}^{\lfloor ns\rfloor}(-1)^j u_{n,j}\Rightarrow B^{+}(s)$ on $D[0,1]$ where $B_+(\cdot)$ is a BM with variance $\omega^2$ (e.g. Theorem 3.33 of Jacod and Shiryaev (2003)) and $n(1-|\rho_n|)=n(1+\rho_n)\to -c$ imply $n^{-1/2}x_{\lfloor nt\rfloor}^{+}\Rightarrow\int_0^t e^{c(t-s)}dB^{+}(s)$. Therefore, Lemma 5 continues to apply to $B_n^{+}(t)$, $U_n^{+}(s)$ and $Y_n^{+}(s)$, defined as their counterparts in Lemma 5 with $(\varphi_{1n},\varphi_{2n},z_{1t},e_{n,t},u_{n,t})$ replaced by $\left(\left|\varphi_{1n}^{-}\right|,\left|\varphi_{2n}^{-}\right|,z_{1t}^{+},(-1)^{-t}e_{n,t},(-1)^{-t}u_{n,t}\right)$; hence $\pi_n\left(\tilde{\rho}_n^{+}-|\rho_n|\right)\to_d\mathcal{MN}(0,V_1)$ by the $\mathrm{C}_+$(ii) case established above; hence, (A.72) and the symmetry of a centred mixed Gaussian distribution around 0 imply that $\pi_n(\tilde{\rho}_n-\rho_n)\to_d\mathcal{MN}(0,V_1)$ under C_(ii). The proof for $\tilde{\beta}_n$ under $\mathrm{C}_-$(ii) Assumption 6 follows from a similar argument, completing the proof of part (i) of the theorem. We proceed by completing the proof of

$$N_n\to_d\mathcal{N}(0,1)\quad\text{and}\quad N_{n,\varepsilon}\to_d\mathcal{N}(0,1)\qquad\text{(A.73)}$$

under Assumption 5 when $|\rho|\geq 1$ or $\sigma_{n,t}^2=\sigma_n^2$ $a.s.$ for $N_n$ and under Assumption 6 when $|\rho|\geq 1$ or $\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(\varepsilon_{n,t}^2\right)=\mathbb{E}\left(\varepsilon_{n,t}^2\right)$ for $N_{n,\varepsilon}$. Having established (A.73) under C(i) and C(iii), we need to provide a proof under C(ii). Writing

$$\begin{aligned}N_n &= N_{1n}\mathbf{1}_{F_n^+}+N_{1n}^{-}\mathbf{1}_{F_n^-}+N_{2n}\mathbf{1}_{\bar{F}_n^+}+N_{2n}^{-}\mathbf{1}_{\bar{F}_n^-}\\ &= \sigma^{-2}U_n(1)\mathbf{1}_{F_n^+}-\sigma^{-2}U_n^{+}(1)\mathbf{1}_{F_n^-}+sign(Z_n)(Y_n/\sigma)\mathbf{1}_{\bar{F}_n^+}-sign\left(Z_n^{+}\right)\left(Y_n^{+}/\sigma\right)\mathbf{1}_{\bar{F}_n^-}+o_p(1)\end{aligned}$$

where $U_n^{+}(\cdot)$ is defined in the same way as $U_n(\cdot)$ in Lemma 5 with $(\varphi_{1n},z_{1t},u_{n,t})$ replaced by $\left(\left|\varphi_{1n}^{-}\right|,(-1)^{-t}z_{1t},(-1)^{-t}u_{n,t}\right)$, Lemma 6 and Lemma 3(ii) yield

$$N_n\to_d N_1\mathbf{1}_{F_c}+N_2\mathbf{1}_{\bar{F}_c},$$

where $N_1=\sigma^{-2}U(1)\mathbf{1}_{\{\rho\geq 0\}}-\sigma^{-2}U^{+}(1)\mathbf{1}_{\{\rho<0\}}$ and $N_2=sign(Z)(Y/\sigma)\mathbf{1}_{\{\rho\geq 0\}}-sign(Z^{+})(Y^{+}/\sigma)\mathbf{1}_{\{\rho<0\}}$, $N_1$ and $N_2$ are $\mathcal{N}(0,1)$ under Assumption 5 by (A.68) and joint convergence in distribution follows by Lemma 5. By Lemma 5, $(U(1),U^{+}(1))$ and $(Z,Y,Z^{+},Y^{+})$ are independent of the BM $\{B(s):s\in[0,1]\}$, which implies that $(N_1,N_2)$ is independent of $\left(F_c,\bar{F}_c\right)$; hence

$$\begin{aligned}\mathbb{P}\left(N_1\mathbf{1}_{F_c}+N_2\mathbf{1}_{\bar{F}_c}\leq x\right) &= \mathbb{P}(N_1\leq x,F_c)+\mathbb{P}\left(N_2\leq x,\bar{F}_c\right)=\mathbb{P}(N_1\leq x)\mathbb{P}(F_c)+\mathbb{P}(N_2\leq x)\mathbb{P}\left(\bar{F}_c\right)\\ &= \Phi(x)\left[\mathbb{P}(F_c)+\mathbb{P}\left(\bar{F}_c\right)\right]=\Phi(x)\end{aligned}$$

completing the proof of (A.73) for $N_n$ under Assumption 5; the proof for $N_{n,\varepsilon}$ in (A.73) under Assumption 6 follows the same argument.

For part (ii), we first derive the limit distribution of

$$\left[\tilde{T}_n, \bar{T}_n\right] = \left(|\Psi_n| / \Psi_n\right) \left[\left(\sigma/\hat{\sigma}_n\right) N_n, \left(\sigma_\varepsilon/\hat{\sigma}_{n,\varepsilon}\right) N_{n,\varepsilon}\right], \ \Psi_n = \textstyle\sum_{t=1}^n \underline{x}_{n,t-1}\tilde{z}_{n,t-1}. \tag{A.74}$$

The consistency of $\hat{\sigma}_n$, $\hat{\sigma}_{n,\varepsilon}$ and (A.73) imply: $\left(\sigma/\hat{\sigma}_n\right) N_n \to_d \zeta =_d \mathcal{N}(0,1)$ and $\left(\sigma_\varepsilon/\hat{\sigma}_{n,\varepsilon}\right) N_{n,\varepsilon} \to_d \zeta_\varepsilon =_d \mathcal{N}(0,1)$. By part (i) of the theorem, we know that $\pi_n^{-2}\Psi_n \to_d \Psi \neq 0$ $a.s.$ and Lemma 5 implies that $\left(\Psi_n, N_n, N_{n,\varepsilon}\right) \to_d \left(\Psi, \zeta, \zeta_\varepsilon\right)$ where $\Psi$ is independent of $\left(\zeta, \zeta_\varepsilon\right)$; hence

$$\begin{aligned}
\lim\nolimits_{n\to\infty} \mathbb{P}_{\theta_n'}\left(T_n \le x\right) &= \mathbb{P}_{\theta_0}\left(sign\left(\Psi\right)\zeta \le x\right) = \mathbb{P}_{\theta_0}\left(\zeta \le x, \Psi > 0\right) + \mathbb{P}_{\theta_0}\left(-\zeta \le x, \Psi < 0\right) \\
&= \mathbb{P}_{\theta_0}\left(\zeta \le x\right)\mathbb{P}_{\theta_0}\left(\Psi > 0\right) + \mathbb{P}_{\theta_0}\left(-\zeta \le x\right)\mathbb{P}_{\theta_0}\left(\Psi < 0\right) \\
&= \Phi\left(x\right)\left[\mathbb{P}_{\theta_0}\left(\Psi > 0\right) + \mathbb{P}_{\theta_0}\left(\Psi < 0\right)\right] = \Phi\left(x\right).
\end{aligned}$$

We have shown that $T_n \to_d \mathcal{N}(0,1)$ under Assumption 5 when $\sigma_{n,t}^2 = \sigma^2$ or $|\rho_n| \to |\rho| \ge 1$ and, the same argument shows that $\bar{T}_n \to_d \mathcal{N}(0,1)$ under Assumption 6 when $\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(e_{n,t}^2\right) = \sigma_{e,n}^2$ or $|\rho_n| \to |\rho| \ge 1$, completing the proof of part (ii) for $T_n$ under $\mathbb{P}_{\theta_n'}$ and part (iii) for $\bar{T}_n$ under $\mathbb{P}_{\bar{\theta}_n'}$.

Next, we establish the approximations

$$\pi_n^{-2}\textstyle\sum_{t=1}^n \tilde{z}_{1t-1}^2\left(\hat{u}_{n,t}^2 - u_{n,t}^2\right) = o_p(1) \ \text{ and } \ \pi_n^{-2}\sum_{t=1}^n \tilde{z}_{1t-1}^2\left(\hat{\varepsilon}_{n,t}^2 - \varepsilon_{n,t}^2\right) = o_p(1) \tag{A.75}$$

under Assumptions 5 and 6 respectively. We only need to consider cases C(i)-C(ii) as Lemma 2 implies that (A.75) holds trivially under C(iii). The identity $\hat{u}_{n,t} = u_{n,t} - \bar{u}_n - \left(\hat{\rho}_n - \rho_n\right)\underline{x}_{n,t-1}$ and the inequality $\left(x+y\right)^2 \le 2\left(x^2+y^2\right)$ give

$$\begin{aligned}
\pi_n^{-2}\textstyle\sum_{t=1}^n \tilde{z}_{1t-1}^2\left(\hat{u}_{n,t}^2 - u_{n,t}^2\right) &\le 2\pi_n^{-2}\left(\left(\hat{\rho}_n - \rho_n\right)^2\sum_{t=1}^n \tilde{z}_{1t-1}^2\underline{x}_{n,t-1}^2 + |\bar{u}_n|\left|\sum_{t=1}^n \tilde{z}_{1t-1}^2 u_{n,t}\right|\right) \\
&\quad + 2\pi_n^{-2}\left(|\hat{\rho}_n - \rho_n|\left|\textstyle\sum_{t=1}^n \tilde{z}_{1t-1}^2 u_{n,t}\underline{x}_{n,t-1}\right| + \bar{u}_n^2\sum_{t=1}^n \tilde{z}_{1t-1}^2\right)
\end{aligned} \tag{A.76}$$

Since $|\bar{u}_n| = O_p\left(n^{-1/2}\right)$, the last term of (A.76) is $O_p\left(n^{-1}\right)$. The same is true for the second term, since $\pi_n^{-2}\left|\sum_{t=1}^n \tilde{z}_{1t-1}^2 u_{n,t}\right| \le \left(\pi_n^{-2}\sum_{j=1}^n \tilde{z}_{1j-1}^2\right)^{1/2}\left(\pi_n^{-2}\sum_{t=1}^n \tilde{z}_{1t-1}^2 u_{n,t}^2\right)^{1/2} = O_p(1)$ by (A.40) and (A.42). The inequality $\left|\sum_{t=1}^n \tilde{z}_{1t-1}^2 u_{n,t}\underline{x}_{n,t-1}\right| \le \left(\sum_{t=1}^n \tilde{z}_{1t-1}^2 u_{n,t}^2\right)^{1/2}\left(\sum_{t=1}^n \tilde{z}_{1t-1}^2\underline{x}_{n,t-1}^2\right)^{1/2}$ on the third term of (A.76) implies that (A.76) is $o_p(1)$ if $\pi_n^{-2}\left(\hat{\rho}_n - \rho_n\right)^2\sum_{t=1}^n \tilde{z}_{1t-1}^2\underline{x}_{n,t-1}^2 = o_p(1)$ i.e. if $n^{-2}\lambda_{1n}^{-1}\kappa_n^{-1}\sum_{t=1}^n \tilde{z}_{1t-1}^2\underline{x}_{n,t-1}^2 = o_p(1)$. By (A.1), it is easy to see that proving $r_n := n^{-2}\lambda_{1n}^{-1}\kappa_n^{-1}\sum_{t=1}^n \tilde{z}_{1t-1}^2 x_{0t-1}^2 = o_p(1)$ is sufficient. Under Assumption 5(i) and C(i), Proposition A1(ii) of MP(2020) implies that the LC $\mathcal{L}_n(\delta) := \sum_{t=1}^n \mathbb{E}\left(\xi_{n,t}^2\mathbf{1}\left\{\xi_{n,t}^2 > \delta\right\}\right) \to 0$ for any $\delta > 0$ is satisfied by $\xi_{n,t} := n^{-1/2}\kappa_n^{-1/2}x_{0t}$; since $\mathcal{L}_n(\delta) \to 0$ for any $\delta > 0$ implies that $\max_{1\le t\le n}\left|\xi_{n,t}\right| = o_p(1)$ (see (3.4) and (3.5) in HH(1980)), we conclude that $n^{-1/2}\kappa_n^{-1/2}\max_{1\le t\le n}|x_{0t}| = o_p(1)$, giving $r_n \le \left(n^{-1/2}\kappa_n^{-1/2}\max_{1\le t\le n}|x_{0t}|\right)^2 n^{-1}\lambda_{1n}^{-1}\sum_{t=1}^n \tilde{z}_{1t-1}^2 = o_p(1)$. Under Assumption 5(i) and C(ii), $n^{-1/2}\max_{1\le t\le n}|x_{0t}| \to_d \sup_{r\in[0,1]}|J_c(r)|$ by the FCLT and the continuous mapping theorem, which implies that $r_n \le \max_{1\le t\le n} x_{0t}^2 n^{-3}\lambda_{1n}^{-1}\sum_{t=1}^n \tilde{z}_{1t-1}^2 = O_p\left(n^{-1}\right)$. Under Assumption 5(ii), $\|r_n\|_{L_1} \le n^{-1}\max_{1\le t\le n}\left\|\lambda_{1n}^{-1/2}\tilde{z}_{1t}\right\|_{L_4}^2\left\|\kappa_n^{-1/2}x_{0t}\right\|_{L_4}^2 = O\left(n^{-1}\right)$. This proves the first part of (A.75) under Assumption 5; the second part under Assumption 6 follows similarly.

Next, we show $\left|T_n - T_n^{EW}\right| = o_{\mathbb{P}_{\theta_n}}(1)$ under Assumption 5 in all cases apart from $\rho_n \to \rho \in (-1,1)$ and $\sigma_{n,t}^2 \neq \sigma^2$. (A.75) and consistency of $\hat{\sigma}_n^2$ imply that under C$_+$(i)-C$_+$(ii),

$$\pi_n^{-2} \sum_{t=1}^n \tilde{z}_{n,t-1}^2 \left(\hat{u}_{n,t}^2 \mathbf{1}_{F_n} + \hat{\sigma}_n^2 \mathbf{1}_{\bar{F}_n}\right) = \mathbf{1}_{F_n} \pi_n^{-2} \sum_{t=1}^n \tilde{z}_{1t-1}^2 u_{n,t}^2 + \mathbf{1}_{\bar{F}_n} \sigma^2 \pi_n^{-2} \sum_{t=1}^n \tilde{z}_{2t-1}^2 + o_p(1). \quad \text{(A.77)}$$

By (A.28), $\pi_n^{-2} \sum_{t=1}^n \left(\tilde{z}_{1t-1}^2 - \tilde{z}_{0t-1}^2\right) u_{n,t}^2 \leq A_n + 2B_n^{1/2} A_n^{1/2}$ where $A_n = \pi_n^{-2} \sum_{t=1}^n q_{nt-1}^2 u_{n,t}^2 = o_p(1)$ and $B_n = \pi_n^{-2} \sum_{t=1}^n \tilde{z}_{0t-1}^2 u_{n,t}^2 = O_p(1)$ by (A.42) and $\|A_n\|_{L_1} = O\left(n^{-1} \kappa_n^{-2} \lambda_{1n} \Lambda_{1n}\right) = o(1)$ by (A.33). By (A.42) and (A.40) ($e_{n,t} = u_{n,t}$ under Assumption 5), $\pi_n^{-2} \sum_{t=1}^n \tilde{z}_{0t-1}^2 \left(u_{n,t}^2 - \sigma_{n,t}^2\right) = o_p(1)$ and $\pi_n^{-2} \sum_{t=1}^n \tilde{z}_{0t-1}^2 \sigma_{n,t}^2 = \sigma^2 \pi_n^{-2} \sum_{t=1}^n \tilde{z}_{0t-1}^2 + o_p(1)$; since $\pi_n^{-2} \sum_{t=1}^n \tilde{z}_{1t-1}^2 = \pi_n^{-2} \sum_{t=1}^n \tilde{z}_{0t-1}^2 + o_p(1)$ by Lemma 3(ii), the right side of (A.77) becomes

$$\sigma^2 \pi_n^{-2} \sum_{t=1}^n \left(\tilde{z}_{1t-1}^2 \mathbf{1}_{F_n} + \tilde{z}_{2t-1}^2 \mathbf{1}_{\bar{F}_n}\right) + o_p(1) = \sigma^2 \pi_n^{-2} \sum_{t=1}^n \tilde{z}_{n,t-1}^2 + o_p(1) \quad \text{(A.78)}$$

proving $T_n^{EW} = T_n + o_p(1)$ under C$_+$(i)-C$_+$(ii). Under C$_-$(i)-C$_-$(ii), (A.77) holds with $\tilde{z}_{1t-1}$ replaced by $\tilde{z}_{1t-1}^-$ and $\tilde{z}_{2t-1}$ replaced by $\tilde{z}_{2t-1}^-$; since $\left(\tilde{z}_{1t-1}^-\right)^2 = \left(\tilde{z}_{1t-1}^+\right)^2$ and $\left(\tilde{z}_{2t-1}^-\right)^2 = \left(\tilde{z}_{2t-1}^+\right)^2$, $T_n^{EW} = T_n + o_p(1)$ follows by (A.78). This completes the proof of $\left|T_n - T_n^{EW}\right| = o_{\mathbb{P}_{\theta_n}}(1)$ under Assumption 5(i) or Assumption 5(ii) with $|\rho| \geq 1$. The above argument with $\left(\hat{u}_{n,t}^2, u_{n,t}^2\right)$ replaced by $\left(\hat{\varepsilon}_{n,t}^2, \varepsilon_{n,t}^2\right)$ shows that $\left|\bar{T}_n - \bar{T}_n^{EW}\right| = o_{\mathbb{P}_{\bar{\theta}_n}}(1)$ under Assumption 6(i) or Assumption 6(ii) with $|\rho| \geq 1$. To complete the proof of part (ii) for $T_n^{EW}$ under $\mathbb{P}_{\theta_n}$ and of part (iii) for $\bar{T}_n^{EW}$ under $\mathbb{P}_{\bar{\theta}_n}$, it is enough to show that $T_n^{EW} \to_d \mathcal{N}(0,1)$ and $\bar{T}_n^{EW} \to_d \mathcal{N}(0,1)$ under Assumptions 5(ii) and 6(ii) respectively when $\rho \in (-1,1)$. By Lemma 2, $\bar{T}_n^{EW}$ is asymptotically equivalent to its restriction when $\tilde{z}_{n,t-1} = \tilde{z}_{1t-1}$, so (A.75) and Lemma 3 imply that

$$\bar{T}_n^{EW} = sign\left(\sigma^2 + 2\rho\Gamma\right) \left(\sum_{t=1}^n \tilde{z}_{1,t-1}^2 \varepsilon_{n,t}^2\right)^{-1/2} \left(\sum_{t=1}^n \tilde{z}_{1,t-1}^2 \mathbb{E}_{\mathcal{F}_{n,t-1}} \varepsilon_{n,t}^2\right)^{1/2} \zeta_n + o_p(1) \quad \text{(A.79)}$$

where $\zeta_n = \left(\sum_{t=1}^n \tilde{z}_{1,t-1}^2 \mathbb{E}_{\mathcal{F}_{n,t-1}} \varepsilon_{n,t}^2\right)^{-1/2} \sum_{t=1}^n \tilde{z}_{1t-1} \varepsilon_{n,t} \to_d \mathcal{N}(0,1)$ under $\mathbb{P}_{\bar{\theta}_n}$ from the proof of Lemma 3(iii); since $\left(n^{-1} \sum_{t=1}^n \tilde{z}_{1,t-1}^2 \varepsilon_{n,t}^2\right)^{-1/2} \left(n^{-1} \sum_{t=1}^n \tilde{z}_{1,t-1}^2 \mathbb{E}_{\mathcal{F}_{n,t-1}} \varepsilon_{n,t}^2\right)^{1/2} \to_p 1$ by (A.42), (A.79) implies that $\bar{T}_n^{EW} \to_d \mathcal{N}(0,1)$ under $\mathbb{P}_{\bar{\theta}_n}$ when $\rho \in (-1,1)$. The same argument works for $T_n^{EW}$ by replacing $\varepsilon_{n,t}^2$ by $u_{n,t}^2$ and $sign\left(\sigma^2 + 2\rho\Gamma\right)$ by $sign\left(\sigma^2\right) = 1$. This completes the proof of part (ii) of the theorem and of part (iii) of the theorem for $\bar{T}_n$ and $\bar{T}_n^{EW}$.

It remains to prove the remainder of part (iii) of the theorem for $\beta_n^*$, $T_n^*$ and $T_n^{*EW}$. We prove that $\pi_n\left(\tilde{\beta}_n - \beta_n^*\right) \to_p 0$; since $\beta_n^* = \tilde{\beta}_n$ on the event $\bar{F}_n \cup \{\hat{\rho}_n < 0\}$, it is enough to provide a proof under C$_+$(i)-C$_+$(ii): $\pi_n\left(\tilde{\beta}_{1n} - \beta_{1n}^*\right) \to_p 0$ with $\pi_n \asymp n^{1/2} \lambda_{1n}^{1/2}$. From (10) and (22),

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

since $\bar{z}_{1n} x_{n,n} = O_p[n^{-1} \lambda_{1n} (\Lambda_{1n}^{1/2} \kappa_n^{1/2} + \kappa_n^{-1/2} \Lambda_{1n})] = O_p(\lambda_{1n}) = o_p(\pi_n)$. For $T_n^*$, by (23) and (17),

$$T_n^* - \bar{T}_n = \pi_n^{-1} [\Sigma_n^{-1/2} - \hat{\sigma}_\varepsilon^{-1} \left(\underline{X}' P_{\tilde{Z}} \underline{X}\right)^{1/2}] \pi_n (\tilde{\beta}_n - \beta_n) + \pi_n^{-1} \Sigma_n^{-1/2} \pi_n (\beta_n^* - \tilde{\beta}_n).$$

The first term on the right is $o_p(1)$ because $\pi_n^{-2} \Delta_n = \pi_n^{-2} n \bar{z}_{1,n-1}^2 = O_p(n^{-1} \lambda_{1n})$ implies that $(\pi_n^2 \Sigma_n)^{-1/2} = \left|\pi_n^{-2} \underline{X}' \tilde{Z}\right| (\pi_n^{-2} \tilde{Z}' \tilde{Z} - \pi_n^{-2} \Delta_n)^{-1/2} \hat{\sigma}_\varepsilon^{-1} = \hat{\sigma}_\varepsilon^{-1} (\pi_n^{-2} \underline{X}' P_{\tilde{Z}} \underline{X})^{1/2} + o_p(1)$; the second term is $o_p(1)$ because $\pi_n^{-1} \Sigma_n^{-1/2} = O_p(1)$ and $\pi_n(\beta_n^* - \tilde{\beta}_n) = o_p(1)$. We conclude that $T_n^* = \bar{T}_n + o_p(1)$.

Finally, since $\pi_n^{-2}\Delta_n \to_p 0$, $\left(\pi_n^2 Q_{n,\varepsilon}^*\right)^{-1/2} = \left(\pi_n^2 Q_{n,\varepsilon}\right)^{-1/2} + o_p(1)$, giving

$$T_n^{*EW} = \left(\pi_n^2 Q_{n,\varepsilon}\right)^{-1/2} \pi_n\left(\beta_n^* - \beta_n\right) + o_p(1) = \bar{T}_n^{EW} + \left(\pi_n^2 Q_{n,\varepsilon}\right)^{-1/2} \pi_n(\tilde{\beta}_n - \beta_n^*) + o_p(1)$$

and $\pi_n(\tilde{\beta}_n - \beta_n^*) = o_p(1)$ implies that $T_n^{*EW} = \bar{T}_n^{EW} + o_p(1)$.

**Proof of Theorems 2 and 3.** By the definition of supremum, we know that there exists a sequence $(\theta_n^*)_{n\in\mathbb{N}}$ in $\Theta^{\text{hom}}$ such that $\sup_{\theta\in\Theta^{\text{hom}}} \mathbb{P}_\theta(\mathcal{R}_{n,\alpha}) = \lim_{n\to\infty} \mathbb{P}_{\theta_n^*}(\mathcal{R}_{n,\alpha})$; $\mathbb{P}_{\theta_n^*}(\mathcal{R}_{n,\alpha}) = \mathbb{P}_{\theta_n^*}(|T_n| > \Phi^{-1}(1-\alpha/2))$ so, if we could show that $\lim_{n\to\infty} \mathbb{P}_{\theta_n^*}(|T_n| \le x) = \Phi(x)$ along the sequence $(\theta_n^*)$ for which the supremum is attained, $\sup_{\theta\in\Theta^{\text{hom}}} \mathbb{P}_\theta(\mathcal{R}_{n,\alpha}) = \alpha$ would follow. Proving the stronger result $\lim_{n\to\infty} \mathbb{P}_{\theta_n}(|T_n| \le x) = \Phi(x)$ *for any* sequence $(\theta_n)_{n\in\mathbb{N}}$ in $\Theta^{\text{hom}}$ is sufficient.

We employ the above proof strategy: showing that

$$\lim_{n\to\infty} \mathbb{P}_{\theta_n'}(|T_n| \le x) = \Phi(x) \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}_{\theta_n}\left(\left|T_n^{EW}\right| \le x\right) = \Phi(x) \text{ for all } x \in \mathbb{R} \tag{A.80}$$

for arbitrary sequences $(\theta_n)_{n\in\mathbb{N}} \subseteq \Theta$ and $(\theta_n')_{n\in\mathbb{N}} \subseteq \Theta^{\text{hom}}$ is sufficient to prove Theorem 2. For brevity, we denote $p_n(\theta_n', x) := \mathbb{P}_{\theta_n'}(|T_n| \le x)$, $q_n(\theta_n, x) := \mathbb{P}_{\theta_n}\left(\left|T_n^{EW}\right| \le x\right)$, $\bar{p}_n\left(\bar{\theta}_n', x\right) := \mathbb{P}_{\bar{\theta}_n'}\left(\left|\bar{T}_n\right| \le x\right)$, $\bar{q}_n\left(\bar{\theta}_n, x\right) := \mathbb{P}_{\bar{\theta}_n}\left(\left|\bar{T}_n^{EW}\right| \le x\right)$, $p_n^*\left(\bar{\theta}_n', x\right) := \mathbb{P}_{\bar{\theta}_n'}(|T_n^*| \le x)$, $q_n^*\left(\bar{\theta}_n, x\right) := \mathbb{P}_{\bar{\theta}_n}(|T_n^*| \le x)$. Theorem 3 will follow by showing that

$$\lim_{n\to\infty} \bar{p}_n\left(\bar{\theta}_n', x\right) = \lim_{n\to\infty} p_n^*\left(\bar{\theta}_n', x\right) = \lim_{n\to\infty} \bar{q}_n\left(\bar{\theta}_n, x\right) = \lim_{n\to\infty} q_n^*\left(\bar{\theta}_n, x\right) = \Phi(x) \tag{A.81}$$

for all $x \in \mathbb{R}$ for arbitrary sequences $\left(\bar{\theta}_n\right)_{n\in\mathbb{N}} \subseteq \bar{\Theta}$ and $\left(\bar{\theta}_n'\right)_{n\in\mathbb{N}} \subseteq \bar{\Theta}^{\text{hom}}$. We show (A.80) and (A.81) by verifying the following criterion for each of the sequences $p_n$, $q_n$, $\bar{p}_n$, $p_n^*$, $\bar{q}_n$, $q_n^*$.

(s) For any subsequence $\{q_{m_n}(\theta_{m_n}, x) : n \in \mathbb{N}\}$ of $\{q_n(\theta_n, x) : n \in \mathbb{N}\}$ there exists a subsequence $\{q_{k_n}(\theta_{k_n}, x) : n \in \mathbb{N}\}$ of $\{q_{m_n}(\theta_{m_n}, x) : n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} q_{k_n}(\theta_{k_n}, x) = \Phi(x)$, $x \in \mathbb{R}$.

Let $(\theta_n)_{n\in\mathbb{N}} \subseteq \Theta = \Theta_\rho \times \Theta_u \times \Theta_{X_0}$ with $\theta_n = \left(\rho_n, (F_{n,t})_{t\in\mathbb{Z}}, F_{n,X_0}\right)$. If $\{\mathbb{P}_{\theta_n}(u_{n,t} \le x) : t \in \mathbb{Z}\}$ is an element of $\Theta_u$, $\mathbb{E}u_{n,t}^2 = \sigma_n^2$ for all $t$ and $\sup_{n\ge1} \sigma_n^2 < \infty$. We start by extracting a subsequence along $\Theta_\rho$: since $\rho_n \in \Theta_\rho = [-M, M]$ for all $n$, BW implies the existence of a subsequence $\left(\rho_{m_n}\right)_{n\in\mathbb{N}}$ of $(\rho_n)_{n\in\mathbb{N}}$ such that $\rho_{m_n} \to \rho \in [-M, M]$; by Lemma 1(i), there exists a 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 7. We continue by extracting a subsequence along $\Theta_{X_0}$: $\sup_{n\ge1} \mathbb{P}_{\theta_n'}(|X_{n,0}| > \lambda) \le \lambda^{-\eta} \sup_{n\ge1} \mathbb{E}|X_{n,0}|^\eta \to 0$ as $\lambda \to \infty$ by Assumption 3 and $\sup_{n\ge1} \mathbb{E}|U_n| \le (|\rho| - 1)^{-1} \sup_{n\ge1} \sigma_n$ when $|\rho| > 1$; hence $(X_{n,0})_{n\in\mathbb{N}}$ and $(U_n)_{n\in\mathbb{N}}$ are tight sequences. Given the subsequence $(s_n)_{n\in\mathbb{N}}$ of $\mathbb{N}$ along which $\left(\rho_{s_n}\right)_{n\in\mathbb{N}}$ satisfies Assumption 7, tightness of $(X_{s_n,0})_{n\in\mathbb{N}}$ implies that there exists a subsequence $(X_{l_n,0})_{n\in\mathbb{N}}$ of $(X_{s_n,0})_{n\in\mathbb{N}}$ that converges in distribution: $X_{l_n,0} \to_d X_0$ where $X_0$ is $\sigma(\cup_{n\ge1}\mathcal{F}_{l_n,0})$ measurable, since $X_{l_n,0}$ is $\mathcal{F}_{l_n,0}$-adapted; tightness of $(U_{l_n,0})_{n\in\mathbb{N}}$ implies that there exists a subsequence $(U_{r_n,0})_{n\in\mathbb{N}}$ of $(U_{l_n,0})_{n\in\mathbb{N}}$ that converges in distribution; tightness $(\alpha X_{l_n,0} + \beta U_{l_n})_{n\in\mathbb{N}}$ for any constants $\alpha, \beta$ implies the existence of a subsequence $(\nu_n)_{n\in\mathbb{N}}$ of $(l_n)_{n\in\mathbb{N}}$ such that $(\alpha X_{\nu_n,0} + \beta U_{\nu_n})_{n\in\mathbb{N}}$ converges in distribution for all constants $\alpha, \beta$; the Cramér-Wold device then implies that $(X_{\nu_n,0}, U_{\nu_n})_{n\in\mathbb{N}}$ converges in distribution. We have shown that:

(i) there exists a subsequence $(\theta_{\nu_n})_{n\in\mathbb{N}}$ of $(\theta_n)_{n\in\mathbb{N}}$ in $\Theta$ such that $(\theta_{\nu_n})_{n\in\mathbb{N}}$ satisfies Assumption 7, $X_{\nu_n,0} \to_d X_0$ under $\mathbb{P}_{\theta_{\nu_n}}$ for $\sigma(\cup_{n\geq 1}\mathcal{F}_{\nu_n,0})$ measurable $X_0$ and $(U_{\nu_n})_{n\in\mathbb{N}}$ converges in distribution when $|\rho| > 1$ jointly with $X_{\nu_n,0}$ under $\mathbb{P}_{\theta_{\nu_n}}$.

(ii) there exists a subsequence $\left(\bar{\theta}_{\nu_n}\right)_{n\in\mathbb{N}}$ of $\left(\bar{\theta}_n\right)_{n\in\mathbb{N}}$ in $\bar{\Theta}$ such that $\left(\bar{\theta}_{\nu_n}\right)_{n\in\mathbb{N}}$ satisfies the conclusion in (i) under $\mathbb{P}_{\bar{\theta}_{\nu_n}}$.

If $\theta' \in \Theta^{\text{hom}}$ with $F_t(\cdot) = \mathbb{P}_{\theta'}(u_t \leq \cdot)$, the Lyapounov inequality $\mathbb{E}_{\mathcal{F}_{t-1}}|u_t| \leq \sigma$ and (12) imply that $\sigma^2 \geq \delta^2$, so that $\sigma^2 \in \left[\delta^2, B\right]$ for some $B > \delta > 0$, the upper bound obtained by UI of $(u_t^2)_{t\in\mathbb{Z}}$. Now consider a sequence $(\theta'_n)_{n\in\mathbb{N}} \subseteq \Theta^{\text{hom}} = \Theta_\rho \times \Theta_u^{\text{hom}} \times \Theta_{X_0}$ with $\theta'_n = \left(\rho_n, (F_{n,t})_{t\in\mathbb{Z}}, F_{n,X_0}\right)$. If $F_{n,t}(\cdot) = \mathbb{P}_{\theta'_n}(u_{n,t} \leq \cdot)$, $(u_{n,t})_{t\in\mathbb{Z}}$ satisfies Assumption 2(i) for each $n$, so $\sigma_n^2 = \mathbb{E}u_{n,1}^2 = \left[\delta^2, B\right]$ for all $n$; by BW, there exists a subsequence $(k_n)_{n\in\mathbb{N}}$ of $(\nu_n)_{n\in\mathbb{N}}$ in (i) such that $\sigma_{k_n}^2 \to \sigma^2 > 0$. Moreover, $(u_{n,t}, \mathcal{F}_{n,t})_{t\in\mathbb{Z}}$ is a martingale difference sequence with $\liminf_{t\to\infty} \mathbb{E}_{\mathcal{F}_{n,t-1}}|u_{n,t}| \geq \delta$ *a.s.* and $\left(u_{n,t}^2\right)_{t\in\mathbb{Z}}$ UI for each $n$ so Lemma 1(ii) implies that Assumption 5(i) is satisfied. The subsequence $\left(\theta'_{k_n}\right)_{n\in\mathbb{N}}$ lies in $\Theta^{\text{hom}}$ and satisfies both Assumptions 5 and 7, so Theorem 1(ii) implies that $\lim_{n\to\infty} p_{k_n}\left(\theta'_{k_n}, x\right) = \Phi(x)$ for all $x \in \mathbb{R}$. We have shown that $\{p_n(\theta'_n, x) : n \in \mathbb{N}\}$ satisfies (s) and the first part of (A.80).

For the second part of (A.80), consider $(\theta_n)_{n\in\mathbb{N}} \subseteq \Theta$ with $\theta_n = \left(\rho_n, (F_{n,t})_{t\in\mathbb{Z}}, F_{n,X_0}\right)$; if $(F_{n,t})_{t\in\mathbb{Z}} \in \Theta_u^{\text{hom}}$, then $\theta_n \in \Theta^{\text{hom}}$ and the result follows since $\lim_{n\to\infty} q_{k_n}(\theta_{k_n}, x) = \lim_{n\to\infty} p_{k_n}(\theta_{k_n}, x)$ when $(\theta_n)_{n\in\mathbb{N}} \subseteq \Theta^{\text{hom}}$ satisfies both Assumptions 5 and 7 by Theorem 1(ii). When $(\theta_n)_{n\in\mathbb{N}} \subseteq \Theta\diagdown\Theta^{\text{hom}}$, $(F_{n,t})_{t\in\mathbb{Z}} \in \Theta_u\diagdown\Theta_u^{\text{hom}}$ and $F_{n,t}(\cdot) = \mathbb{P}_{\theta_n}(u_{n,t} \leq \cdot)$ with $(u_{n,t})_{t\in\mathbb{Z}}$ satisfying Assumption 2(ii) for each $n$: $\left(u_{n,t}, \sigma_{n,t}^2\right)_{t\in\mathbb{Z}}$ is strictly stationary with $\sup_{n\in\mathbb{N}} \mathbb{E}\sigma_{n,0}^4 \leq B$, $\sigma_{n,t}^2 \leq \sup_{t\in\mathbb{Z}} \sigma_t^2 < \infty$ *a.s.* for all $t, n$ which implies that $\limsup_{n\to\infty} \sup_{t\in\mathbb{N}} \sigma_{n,t}^2 < \infty$ *a.s.*; also, $\sigma_{n,t}^2 \geq \delta$ by (13) which implies that $\sigma_n^2 \in \left[\delta, B^{1/2}\right]$, so there exists a subsequence $(k_n)_{n\in\mathbb{N}}$ of $(\nu_n)_{n\in\mathbb{N}}$ in (i) such that $\sigma_{k_n}^2 \to \sigma^2 > 0$. Since $(u_{k_n,t})_{n\in\mathbb{N}}$ is an ARCH($\infty$) process, Lemma 1(iii) implies that there exists a subsequence $(u_{h_n,t})_{n\in\mathbb{N}}$ of $(u_{k_n,t})_{n\in\mathbb{N}}$ such that $\sigma_{h_n,t}^2 = \mathbb{E}_{\mathcal{F}_{h_n,t-1}} u_{h_n,t}^2$ satisfies (14); letting $v_{1,n}(\rho) := \mathbb{E}u_{n,1}^2 \left(\sum_{j=0}^{\infty} \rho^j u_{n,-j}\right)^2$ when $\rho \in (-1,1)$, $\sup_{n\geq 1} v_{1,n}(\rho) \leq \left(\sum_{j=0}^{\infty} |\rho|^j\right)^2 \sup_{n\geq 1} \mathbb{E}u_{n,1}^4$, so BW implies that there exists a subsequence $\left\{v_{\beta_n}(\rho) : n \in \mathbb{N}\right\}$ of $\left\{v_{h_n}(\rho) : n \in \mathbb{N}\right\}$ such that $v_{1,\beta_n}(\rho) \to v_1(\rho)$; since

$$v_n(\rho) \geq \mathbb{E}u_{n,1}^2 \mathbf{1}\left\{u_{n,1}^2 \geq \delta\right\} \left(\sum_{j=0}^{\infty} \rho^j u_{n,-j}\right)^2 \geq \delta \mathbb{E}\left(\sum_{j=0}^{\infty} \rho^j u_{n,-j}\right)^2 = \delta\sigma_n^2 / \left(1-\rho^2\right) \geq \delta^2 / \left(1-\rho^2\right)$$

since $\sigma_n^2 \geq \delta$ by (13), $\liminf_{n\to\infty} v_n(\rho) > 0$, so $v_1(\rho) \in (0,\infty)$. We conclude that for arbitrary $(\theta_n)_{n\in\mathbb{N}} \subseteq \Theta$ there exists $\left(\theta_{\beta_n}\right)_{n\in\mathbb{N}} \subseteq (\theta_n)_{n\in\mathbb{N}}$ that satisfies Assumptions 5 and 7 and, hence,

$\lim_{n\to\infty} q_{\beta_n}\left(\theta_{\beta_n}, x\right) = \Phi(x)$ for all $x \in \mathbb{R}$ by Theorem 1(ii); hence $\{q_n(\theta_n, x) : n \in \mathbb{N}\}$ satisfies (s), completing the proof of (A.80) and of Theorem 2.

If $\left(\bar{\theta}'_n\right)_{n\in\mathbb{N}} \subseteq \bar{\Theta}^{\text{hom}}$, $(v_{n,t}, \mathcal{F}_{n,t})_{t\in\mathbb{Z}}$ with $v_{n,t} = (\varepsilon_{n,t}, e_{n,t})$ is a martingale difference sequence for each $n$ satisfying satisfying Assumption 5(i) for each $n$: $\mathbb{E}_{\mathcal{F}_{n,t-1}}\left(v_{n,t}v'_{n,t}\right) = \Sigma_n$ for all $t$ with $\lambda_{\min}(\Sigma_n) \geq \delta$, $\liminf_{t\to\infty} \mathbb{E}_{\mathcal{F}_{n,t-1}}|e_{n,t}| \geq \delta$ *a.s.* and $\left(\|v_{n,t}\|^2\right)_{t\in\mathbb{Z}}$ UI for all $n$; hence

$$\liminf_{n\to\infty}\liminf_{t\to\infty} \mathbb{E}_{\mathcal{F}_{n,t-1}}|e_{n,t}| \geq \delta \ a.s., \ \max_{1\leq t\leq n} \mathbb{E}\left(\|v_{n,t}\|^2 \mathbf{1}\left\{\|v_{n,t}\|^2 > \lambda_n\right\}\right) \to 0$$

when $\lambda_n \to \infty$. Also, the coefficients $c_{n,j}$ of $u_{n,t}$ satisfy $\sup_{n\geq1}\sum_{j=0}^{\infty} j^{1+\delta}c_{n,j}^2 \leq B$ which implies that $\sup_{n\geq1}\sum_{j=0}^{\infty} j^{\delta/4}|c_{n,j}| < \infty$ by (A.43). Since $\|\Sigma_n\| \leq \mathbb{E}\|v_{n,t}\|^2 \leq B^{\frac{2}{2+\delta}}$, $\left|\gamma_{u_n}(h)\right| \leq \mathbb{E}u_{n,1}^2 \leq B^{\frac{2}{2+\delta}}$ and $\sup_{n\geq1}\left|\sum_{j=0}^{\infty}\rho_n^{-j}c_{n,j}\right| \leq \sup_{n\geq1}\sum_{j=0}^{\infty}|c_{n,j}| < \infty$ when $|\rho_n| \to |\rho| \geq 1$, applying BW thrice implies that there exists a subsequence $(k_n)_{n\in\mathbb{N}}$ of $(\nu_n)_{n\in\mathbb{N}}$ in (ii) such that $\Sigma_{k_n} \to \Sigma$, $\lambda_{\min}(\Sigma) \geq \delta$, $\gamma_{u_{k_n}}(h) \to \gamma(h)$ and $\sum_{j=0}^{\infty}\rho_{k_n}^{-j}c_{k_n,j} \to \sum_{j=0}^{\infty}\rho^{-j}c_j \neq 0$ (Assumption 4 implies that $\inf_{n\geq1}\left|\sum_{j=0}^{\infty}\rho_n^{-j}c_{n,j}\right| \geq \delta$ when $|\rho_n| \to |\rho| \geq 1$). By (ii), $(\bar{\theta}'_{k_n})_{n\in\mathbb{N}} \subseteq \bar{\Theta}^{\text{hom}}$ satisfies Assumptions 6 and 7 so Theorem 1(iii) then implies that $\lim_{n\to\infty}\bar{p}_{k_n}(\bar{\theta}'_{k_n}, x) = \lim_{n\to\infty}p^*_{k_n}(\bar{\theta}'_{k_n}, x) = \Phi(x)$; hence, $\left\{\bar{p}_n(\bar{\theta}'_n, x) : n \in \mathbb{N}\right\}$ and $\left\{p^*_n(\bar{\theta}'_n, x) : n \in \mathbb{N}\right\}$ satisfy (s), proving the first two limits in (A.81).

If $\left(\bar{\theta}_n\right)_{n\in\mathbb{N}} \subseteq \bar{\Theta}$, $\lim_{n\to\infty}\bar{q}_n(\bar{\theta}_n, x) = \lim_{n\to\infty}q^*_n(\bar{\theta}_n, x) = \Phi(x)$ if $\left(\bar{\theta}_n\right)_{n\in\mathbb{N}} \subseteq \bar{\Theta}^{\text{hom}}$. If $\left(\bar{\theta}_n\right)_{n\in\mathbb{N}} \subseteq \bar{\Theta}\diagdown\bar{\Theta}^{\text{hom}}$, $\bar{F}_{n,t}(x) = \mathbb{P}_{\bar{\theta}_n}(\varepsilon_{n,t} \leq x, u_{n,t} \leq y)$ with $(\varepsilon_{n,t}, u_{n,t})_{t\in\mathbb{Z}}$ satisfying Assumption 4(ii) for each $n$: $\left(e_{n,t}, \mathbb{E}_{\mathcal{F}_{n,t-1}}e_{n,t}^2, \varepsilon_{n,t}, \mathbb{E}_{\mathcal{F}_{n,t-1}}\varepsilon_{n,t}^2\right)_{t\in\mathbb{Z}}$ is strictly stationary with $\sup_{n\in\mathbb{N}}\mathbb{E}e_{n,0}^4 \leq B$, $\sup_{n\in\mathbb{N}}\mathbb{E}\varepsilon_{n,0}^4 \leq B$, $\mathbb{E}_{\mathcal{F}_{n,t-1}}e_{n,t}^2 \leq \sup_{t\in\mathbb{Z}}\mathbb{E}_{\mathcal{F}_{t-1}}e_t^2 < \infty$ *a.s.* for each $t, n$, so $\limsup_{n\to\infty}\sup_{t\in\mathbb{N}}\mathbb{E}_{\mathcal{F}_{n,t-1}}e_{n,t}^2 < \infty$ *a.s.*; given the subsequence $(\nu_n)_{n\in\mathbb{N}}$ of $\mathbb{N}$ in (ii), $(\varepsilon_{\nu_n,t})_{n\in\mathbb{N}}$ is an ARCH($\infty$) process, so Lemma 1(iii) implies that there exists a subsequence $(\varepsilon_{k_n,t})_{n\in\mathbb{N}}$ of $(\varepsilon_{\nu_n,t})_{n\in\mathbb{N}}$ such that $\mathbb{E}_{\mathcal{F}_{k_n,t-1}}\varepsilon_{k_n,t}^2$ satisfies (14); letting $v_{2,n}(\rho) := \mathbb{E}\varepsilon_{n,1}^2\left(\sum_{j=0}^{\infty}\rho^j u_{n,-j}\right)^2$ when $\rho \in (-1,1)$, the same argument employed for $v_{1,n}(\rho)$ above with $u_{n,1}^2$ replaced by $\varepsilon_{n,1}^2$ ensures the existence of a subsequence $\left\{v_{2,\beta_n}(\rho) : n \in \mathbb{N}\right\}$ of $\{v_{2,k_n}(\rho) : n \in \mathbb{N}\}$ such that $v_{2,\beta_n}(\rho) \to v_2(\rho) \in (0,\infty)$. Finally, by the same argument used for $\bar{\Theta}^{\text{hom}}$ there exists a subsequence $(\alpha_n)_{n\in\mathbb{N}}$ of $(\beta_n)_{n\in\mathbb{N}}$ along which $\Sigma_{\alpha_n} \to \Sigma > 0$, $\gamma_{u_{\alpha_n}}(h) \to \gamma(h)$ and $\sum_{j=0}^{\infty}\rho_{\alpha_n}^{-j}c_{\alpha_n,j} \to \sum_{j=0}^{\infty}\rho^{-j}c_j \neq 0$ when $|\rho_n| \to |\rho| \geq 1$. Hence, Assumptions 6 and 7 are satisfied along the subsequence $\left(\bar{\theta}_{\alpha_n}\right)_{n\in\mathbb{N}}$ in $\bar{\Theta}$ so Theorem 1(iii) implies that $\lim_{n\to\infty}\bar{q}_{\alpha_n}\left(\bar{\theta}_{\alpha_n}, x\right) = \lim_{n\to\infty}q^*_{\alpha_n}\left(\bar{\theta}_{\alpha_n}, x\right) = \Phi(x)$; thus, $\left\{\bar{q}_n\left(\bar{\theta}_n, x\right) : n \in \mathbb{N}\right\}$ and $\left\{q^*_n(\bar{\theta}_n, x) : n \in \mathbb{N}\right\}$ satisfy the convergence criterion (s), completing the proof of (A.81) and of Theorem 3.

**Proof of Corollary 1.** Using the argument leading to (A.80) in the proof of Theorem 2, it is sufficient to prove that, for arbitrary sequences $(\theta_n)_{n\in\mathbb{N}} \subseteq \Theta$ and $(\theta'_n)_{n\in\mathbb{N}} \subseteq \Theta^{\text{hom}}$, $\mathbb{P}_{\theta'_n}(|T_{n,h}| \leq x)$ and $\mathbb{P}_{\theta_n}\left(\left|T_{n,h}^{EW}\right| \leq x\right)$ both converge to $\Phi(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$ and all $h$ satisfying $h/n \to 0$. By using the same subsequence extraction leading to (i) in the proof of Theorem 2 and that of the subsequent two paragraphs, it is sufficient to establish the above limits for the special case where $(\theta_n)_{n\in\mathbb{N}}$ and $(\theta'_n)_{n\in\mathbb{N}}$ satisfy Assumptions 5 and 7. This reduces the proof of the corollary to the verification of Theorem 1(ii) with $\left(T_n, T_n^{EW}\right)$ replaced by $\left(T_{n,h}, T_{n,h}^{EW}\right)$ and $h/n \to 0$. Letting $\Psi_{n,h} = \sum_{t=1}^{n-h} x_{n,t}\tilde{z}_{n,t}$, $U_{h,t} = \sum_{l=1}^{h} \rho_n^{h-l} u_{n,t+l}$ and $\upsilon_n = \hat{\sigma}_n^2 \left( |\phi_{2n}|^{-h} \sum_{j=0}^{h-1} (|\hat{\rho}_n| |\phi_{2n}|)^j \right)^2 \sum_{t=1}^{n-h} \tilde{z}_{n,t}^2$, we may write

$$T_{n,h} = \left( |\Psi_{n,h}| / \Psi_{n,h} \right) \left( \upsilon_n^{-1/2} \mathbf{1}_{F_n} + \upsilon_n^{-1/2} \mathbf{1}_{\bar{F}_n} \right) \sum_{t=1}^{n-h} \tilde{z}_{n,t} U_{h,t} = sign\left(\Psi_{n,h}\right) N_{n,h}, \tag{A.82}$$

where $N_{n,h} = (\upsilon_n^{-1/2} \mathbf{1}_{F_n} + \upsilon_n^{-1/2} \mathbf{1}_{\bar{F}_n}) \sum_{t=1}^{n-h} \tilde{z}_{n,t} U_{h,t}$ is the LP equivalent of (A.56) in the proof of Theorem 1. Since $h/n \to 0$, $sign\left(\Psi_{n,h}\right) = sign\left(\Psi_n\right) + o_p(1)$ where $\Psi_n$ is defined in (A.74). By comparing (A.82) to (A.73), (A.74), the argument following (A.74) in the proof of Theorem 2 implies that

$$N_{n,h} \to_d \mathcal{N}(0,1) \quad \text{when } n \to \infty \text{ and } h/n \to 0 \tag{A.83}$$

under Assumptions 5 and 7 is sufficient for the proof of the Corollary when $\theta \in \Theta^{\text{hom}}$. To prove (A.83), we employ a martingale approximation to $N_{n,h}$: changing the order of summation

$$\sum_{t=1}^{n-h} \tilde{z}_{n,t} U_{h,t} = \sum_{t=1}^{n-h} \tilde{z}_{n,t} \sum_{j=t+1}^{t+h} \rho_n^{t+h-j} u_{n,j} = A_n + B_n + C_n \tag{A.84}$$

where $A_n = \sum_{j=1}^{h} \rho_n^{h-j} u_{n,j} \sum_{t=1}^{j-1} \rho_n^t \tilde{z}_{n,t}$, $C_n = \sum_{j=n-2h+1}^{n-h} u_{n,j+h} \sum_{t=0}^{n-j-h} \rho_n^t \tilde{z}_{n,t+j}$ and the leading term is given by $B_n = \sum_{j=1}^{n-2h} u_{n,j+h} \sum_{t=0}^{h-1} \rho_n^t \tilde{z}_{n,t+j}$: since the outer sums of $A_n$ and $C_n$ have $h$ terms and $h/n \to 0$, the right side of (A.84) is dominated by $B_n$. As in the proof of Theorem 1, we consider the asymptotic behaviour of (A.84) under a moderately stationary instrument $\tilde{z}_{1t}$ and a mildly explosive instrument $\tilde{z}_{2t}$ separately and then combine the results to prove (A.83).

Under C$_+$(i)-C$_+$(ii), when $\sigma_t^2 = \sigma^2$ or $\rho_n \to 1$ or $h \to \infty$

$$n^{-1} b_n^{-1} \sum_{j=1}^{n-2h} \sigma_{n,j+h}^2 \left( \sum_{t=0}^{h-1} \rho_n^t \tilde{z}_{1t+j} \right)^2 \to_p \sigma^4 \tag{A.85}$$

where: (a) $b_n = ((1-\varphi_{1n}^2)^{-1} \wedge (1-\rho_n^2)^{-1}) h^2$ if $h/((1-\varphi_{1n}^2)^{-1} \wedge (1-\rho_n^2)^{-1}) \to 0$; (b) $b_n = 2^{-1}((1-\varphi_{1n}^2)^{-2} \wedge (1-\rho_n^2)^{-2}) (1-\rho_n^2)^{-1}$ if $((1-\varphi_{1n}^2)^{-1} \vee (1-\rho_n^2)^{-1})/h \to 0$; (c) $b_n = 4^{-1} (1-\varphi_{1n}^2)^{-2} h$ if $(1-\varphi_{1n}^2)^{-1}/h \to 0$ and $h/(1-\rho_n^2)^{-1} \to 0$; (d) $b_n = (1+\rho^2)^{-1} (1-\rho_n^2)^{-3}$ if $(1-\rho_n^2)^{-1}/h \to 0$ and $h/(1-\varphi_{1n}^2)^{-1} \to 0$.

Under C$_+$(ii)-C$_+$(iii), letting $\gamma_n = (\varphi_{2n}^2 - 1)^{-1} \varphi_{2n}^{n-2h} \sum_{j=0}^{h-1} (\rho_n \varphi_{2n})^j$,

$$\gamma_n^{-1} \sum_{t=1}^{n-h} \tilde{z}_{2t} U_{h,t} = Z_{n,h} Y_{n,h} + o_p(1) \tag{A.86}$$

where $Z_{n,h} = (\varphi_{2n}^2 - 1)^{1/2} \sum_{i=1}^{n-h} \varphi_{2n}^{-i} u_{n,i}$ and $Y_{n,h} = (\varphi_{2n}^2 - 1)^{1/2} \sum_{t=1}^{n-2h} \varphi_{2n}^{-(n-2h-t)} u_{n,t+h}$.

We include an abridged version of the proof of (A.85) and (A.86); the details of the remainder analysis follow the lines of the proof of Lemmata 3 and 4 and are omitted due to space restrictions. For (A.85), the IV identity (A.29) $\tilde{z}_{1t} = z_{1t} + \frac{\rho_n - 1}{\rho_n - \varphi_{1n}} (x_t - z_{1t})$ when $\mu = 0$ implies the identity

$$\sum_{t=0}^{h-1} \rho_n^t \tilde{z}_{1t+j} = \sum_{t=0}^{h-1} \rho_n^t z_{1t+j} + \frac{\rho_n - 1}{\rho_n - \varphi_{1n}} \left( \sum_{t=0}^{h-1} \rho_n^t x_{t+j} - \sum_{t=0}^{h-1} \rho_n^t z_{1t+j} \right). \tag{A.87}$$

Changing the order of summation in $\sum_{t=0}^{h-1} \rho_n^t z_{1t+j} = \sum_{t=0}^{h-1} \rho_n^t \sum_{i=1}^{t+j} \varphi_{1n}^{t+j-i} u_{n,i}$ and $\sum_{t=0}^{h-1} \rho_n^t x_{n,t+j} =$

$\sum_{t=0}^{h-1}\rho_n^t\sum_{i=1}^{t+j}\rho_n^{t+j-i}u_{n,i}$ we obtain expressions for the terms on the right of (A.87):

$$\sum_{t=0}^{h-1}\rho_n^t z_{1t+j} = z_{1j}\sum_{t=0}^{h-1}(\rho_n\varphi_{1n})^t+\sum_{i=1}^{h-1}\rho_n^i u_{i+j}\sum_{t=0}^{h-i-1}(\rho_n\varphi_{1n})^t \tag{A.88}$$

$$\sum_{t=0}^{h-1}\rho_n^t x_{t+j} = x_j\sum_{t=0}^{h-1}\rho_n^{2t}+\sum_{i=1}^{h-1}\rho_n^i u_{i+j}\sum_{t=0}^{h-i-1}\rho_n^{2t}. \tag{A.89}$$

The proof of (A.85) identifies the leading terms in (A.87)-(A.89) and employs the following result (see Lemma A1 of the Online Appendix in Kostakis, Magdalinos and Stamatogiannis (2023)): if $\sigma_{n,t}^2$ satisfies Assumption 5 and $\xi_{nt}$ is a triangular array of random variables that admits the decomposition $\xi_{nt}=\zeta_{nt}+r_{nt}$, $\max_{1\le t\le n}\left\|b_n^{-1/2}\zeta_{nt}\right\|_{L_\gamma}=O(1)$ and $\max_{1\le t\le n}\left\|b_n^{-1/2}r_{nt}\right\|_{L_\gamma}\to 0$ for some numerical sequence $b_n\to\infty$ and $\gamma=2$ under Assumption 5(i) and $\gamma=4$ under Assumption 5(ii), then

$$n^{-1}b_n^{-1}\left\|\sum_{t=0}^{n}\sigma_{n,t}^2\left(\xi_{nt}^2-\zeta_{nt}^2\right)\right\|_{L_1}\to 0. \tag{A.90}$$

For the cases (a)-(d) reported below (A.85), (A.90) applies with: (a) $\zeta_{nt}=hz_{1t}$ if $(1-\varphi_{1n}^2)^{-1}\ll(1-\rho_n^2)^{-1}$ and $\zeta_{nt}=hx_t$ if $(1-\rho_n^2)^{-1}\ll(1-\varphi_{1n}^2)^{-1}$; (b) $\zeta_{nt}=2^{-1}(1-\varphi_{1n})^{-1}\left(\sum_{i=1}^{h-1}\rho_n^i u_{n,i+t}-x_{n,t}\right)$ if $(1-\varphi_{1n}^2)^{-1}\ll(1-\rho_n^2)^{-1}$ and $\zeta_{nt}=(1-\rho_n^2)^{-1}\left(x_{n,t}+\sum_{i=1}^{h-1}\rho_n^i u_{n,i+t}\right)$ if $(1-\rho_n^2)^{-1}\ll(1-\varphi_{1n}^2)^{-1}$; (c) $\zeta_{nt}=(1-\rho_n\varphi_{1n})^{-1}\sum_{i=1}^{h-1}\rho_n^i u_{n,i+t}$; (d) $\zeta_{nt}=x_{n,t}\sum_{i=0}^{h-1}\rho_n^{2i}+(1-\rho_n^2)^{-1}\sum_{i=1}^{h-1}\rho_n^i u_{n,i+t}$. In all cases,

$$n^{-1}b_n^{-1}\sum_{t=0}^{n}\sigma_{n,t}^2\zeta_{nt}^2=\sigma^2 n^{-1}b_n^{-1}\sum_{t=0}^{n}\zeta_{nt}^2+o_p(1)\to_p\sigma^4$$

from the choice of $b_n$, showing (A.85).

To prove (A.86), we can employ similar methods to those used in the proof of Lemma 4(i) to show that $\epsilon_n=\gamma_n^{-1}\sum_{t=1}^{n-h}(\tilde z_{2t}-z_{2t})U_{h,t}$ and $\epsilon_n'=\sum_{j=1}^{n-2h}\varphi_{2n}^j u_{n,j+h}\sum_{t=0}^{h-1}(\varphi_{2n}\rho_n)^t\sum_{i=t+j+1}^{n-h}\varphi_{2n}^{-i}u_i$ satisfy $\epsilon_n=o_p(1)$ and $\epsilon_n'=o_p(1)$. The leading term $B_n$ of (A.84) satisfies

$$\begin{aligned}\gamma_n^{-1}B_n &= \gamma_n^{-1}\sum_{j=1}^{n-2h}u_{n,j+h}\sum_{t=0}^{h-1}\rho_n^t\sum_{i=1}^{t+j}\varphi_{2n}^{t+j-i}u_i+o_p(1)\\ &= \gamma_n^{-1}\left(\varphi_{2n}^2-1\right)^{-1/2}Z_{n,h}\sum_{j=1}^{n-2h}\varphi_{2n}^j u_{n,j+h}\sum_{t=0}^{h-1}(\varphi_{2n}\rho_n)^t+o_p(1)\\ &= Z_{n,h}Y_{n,h}+o_p(1)\end{aligned}$$

where the first asymptotic equivalence follows from $\epsilon_n=o_p(1)$ and the second from $\epsilon_n'=o_p(1)$.

Having established (A.85) and (A.86), it is straightforward to prove (A.83) under C$_+$(i) and C$_+$(iii): under C$_+$(i), Lemma 2 implies that $N_{n,h}=\upsilon_{1n}^{-1/2}B_{1n}+o_p(1)$, where $(\upsilon_{1n},B_{1n})$ denote $(\upsilon_n,B_n)$ in (19) and (A.84) with $\tilde z_{n,t}$ replaced by $\tilde z_{1t}$; since $\upsilon_n$ is asymptotically equivalent to the conditional variance of the martingale array $B_{1n}$ by law of large numbers in (A.85), a martingale CLT implies that $\upsilon_n^{-1/2}B_{1n}\to_d\mathcal N(0,1)$, showing (A.83) under C$_+$(i). Under C$_+$(iii), Lemma 2 implies that $N_{n,h}=\upsilon_{2n}^{-1/2}B_{2n}+o_p(1)$, $(\upsilon_{2n},B_{2n})$ denote $(\upsilon_n,B_n)$ with $\tilde z_{n,t}$ replaced by $\tilde z_{2t}$; by Lemma 4(i), $\sum_{t=1}^{n-h}\tilde z_{n,t}^2=(1+o_p(1))(\varphi_{2n}^2-1)^{-2}\varphi_{2n}^{2(n-h)}Z_{n,h}^2$, which implies that

$$N_{n,h}=\sigma^{-1}sign(Z_{n,h})Y_{n,h}+o_p(1)\to_d\mathcal N(0,1)$$

showing (A.83) under C$_+$(iii). Under C$_-$(i), applying the transformation $x\mapsto(-1)^{-t}x$ and denoting $v_{1n}^{+} = \hat{\sigma}_n^2 \sum_{i=1}^{n-2h+1} \left( \sum_{t=0}^{h-1} |\hat{\rho}_n|^t \tilde{z}_{1t+i}^{+} \right)^2$, $U_{h,t}^{+} = \sum_{l=1}^{h} |\rho_n|^{h-l} (-1)^{t+l} u_{n,t+l}$ and $N_{1,n,h}^{+} = \sum_{t=1}^{n-h} \tilde{z}_{1t}^{+} U_{h,t}^{+}$, we obtain

$$N_{n,h} = (1 + o_p(1)) (-1)^{-h} \left(v_{1n}^{+}\right)^{-1/2} \sum_{t=1}^{n-h} \tilde{z}_{1t}^{+} U_{h,t}^{+} \to_d \mathcal{N}(0,1)$$

from the $\mathrm{C}_+$(i) case. Under $\mathrm{C}_-$(iii), a similar argument yields

$$N_{n,h} = (1 + o_p(1)) (-1)^{-h} \left(v_{n}^{+}\right)^{-1/2} \sum_{t=1}^{n-h} \tilde{z}_{2t}^{+} U_{h,t}^{+} \to_d \mathcal{N}(0,1)$$

from the $\mathrm{C}_+$(iii) case (here $v_n^{+}$ denotes $v_n$ with $\tilde{z}_{n,t}$ replaced by $\tilde{z}_{2t}^{+}$). Having proved (A.83) under C(i) and C(iii), the proof of (A.83) under C(ii) follows from the argument following (A.73) in the proof of Theorem 1. This proves (A.83) and the corollary when $\theta \in \Theta^{\mathrm{hom}}$.

When $\theta \in \Theta$, we may show that $T_{n,h}^{EW} = T_{n,h} + o_p(1)$ when $\sigma_{n,t}^2 = \sigma_n^2$ or $|\rho_n| \to \rho \geq 1$ or $h \to \infty$ by using the same argument employed in Theorem 1 to show that $T_n^{EW} = T_n + o_p(1)$ under the same assumptions. When $|\rho_n| \to \rho \in (-1,1)$ and $h \to \infty$, $T_{n,h}^{EW} \to_d \mathcal{N}(0,1)$ follows from the law of large numbers $n^{-1} \sum_{j=1}^{n-2h} \sigma_{j+h}^2 \left( \sum_{t=0}^{h-1} \rho_n^t \tilde{z}_{1t+j} \right)^2 \to_p \mathbb{E}\sigma_h^2 \left( \sum_{t=0}^{h-1} \rho^t x_t \right)^2$ which shows (A.83) when $v_n$ is replaced by $v_n^{EW}$. The details of the proof of the last two statements are omitted and available upon request.

## 1.3 Additional Simulation Results

In this section, we present some additional simulation results. Table B1 below contains the empirical size and Figure B1 displays the power of the two-sided test of our procedure for the PR slope parameter $\beta$ based on $T_n^{*}$ in (23) for $n = 1,000$ based on $10,000$ replications for a grid of points for $b_1$ and $b_2$ for $\rho_{\varepsilon u} = 0.99$ and $\rho_{\varepsilon u} = -0.99$ respectively for the case $\rho = 1$, which we use for the instrument selection of Section 4.1 of the main paper[31]. Our task is to select the largest values for $b_1$ and $b_2$, subject to the size being close to the nominal 5%. Figures B2 and B3 contain the empirical size of our two-sided IV-based test for correlation $\rho_{\varepsilon u}$ of $-0.45$ and $0.45$ respectively. Figure B4 displays the proportion of times the different instruments are chosen. Figure B5 is a comparison of the length of CIs of IV and OLS under misspecification of the last observation (note, in this case, OLS has no valid coverage for the purely explosive specifications). Figures B6 and B7 present the coverage and length of CIs of our IV-based CIs and the equal-tailed two-sided intervals (ETCI) of Andrews and Guggenberger (2014) respectively. Figure B8 displays the empirical size of the OLS- and IV-based one-sided test under misspecification of the last observation. Finally,

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

Figures B9/B11 and B10/B12 contain the empirical size and power of our one-sided IV-based test in comparison with the size and power of the Elliott et al. (2015)'s procedure for correlation $\rho_{\varepsilon u}$ of $-0.45$ and 0.45 respectively.

| Table B1: Empirical size, $\rho_{\varepsilon u}$=−0.99,$n$=1,000 | | | | | | | | Empirical size, $\rho_{\varepsilon u}$=0.99,$n$=1,000 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $b_1/b_2$ | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.9 | 0.95 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 |
| 0.65 | 5.37% | 5.38% | 5.79% | 6.48% | 5.69% | 5.42% | 5.92% | 5.51% | 5.47% | 5.77% | 6.58% | 6.14% | 5.81% | 6.40% |
| 0.70 | 5.58% | 5.59% | 6.00% | 6.69% | 5.90% | 5.63% | 6.13% | 5.35% | 5.31% | 5.61% | 6.42% | 5.98% | 5.65% | 6.24% |
| 0.75 | 5.66% | 5.67% | 6.08% | 6.77% | 5.98% | 5.71% | 6.21% | 5.42% | 5.38% | 5.68% | 6.49% | 6.05% | 5.72% | 6.31% |
| 0.80 | 5.80% | 5.81% | 6.22% | 6.91% | 6.12% | 5.85% | 6.35% | 5.46% | 5.42% | 5.72% | 6.53% | 6.09% | 5.76% | 6.35% |
| 0.85 | 5.87% | 5.88% | 6.29% | 6.98% | 6.19% | 5.92% | 6.42% | 5.63% | 5.59% | 5.89% | 6.70% | 6.26% | 5.93% | 6.52% |
| 0.90 | 6.03% | 6.04% | 6.45% | 7.14% | 6.35% | 6.08% | 6.58% | 5.81% | 5.77% | 6.07% | 6.88% | 6.44% | 6.11% | 6.70% |
| 0.95 | 6.27% | 6.28% | 6.69% | 7.38% | 6.59% | 6.32% | 6.82% | 5.85% | 5.81% | 6.11% | 6.92% | 6.48% | 6.15% | 6.74% |

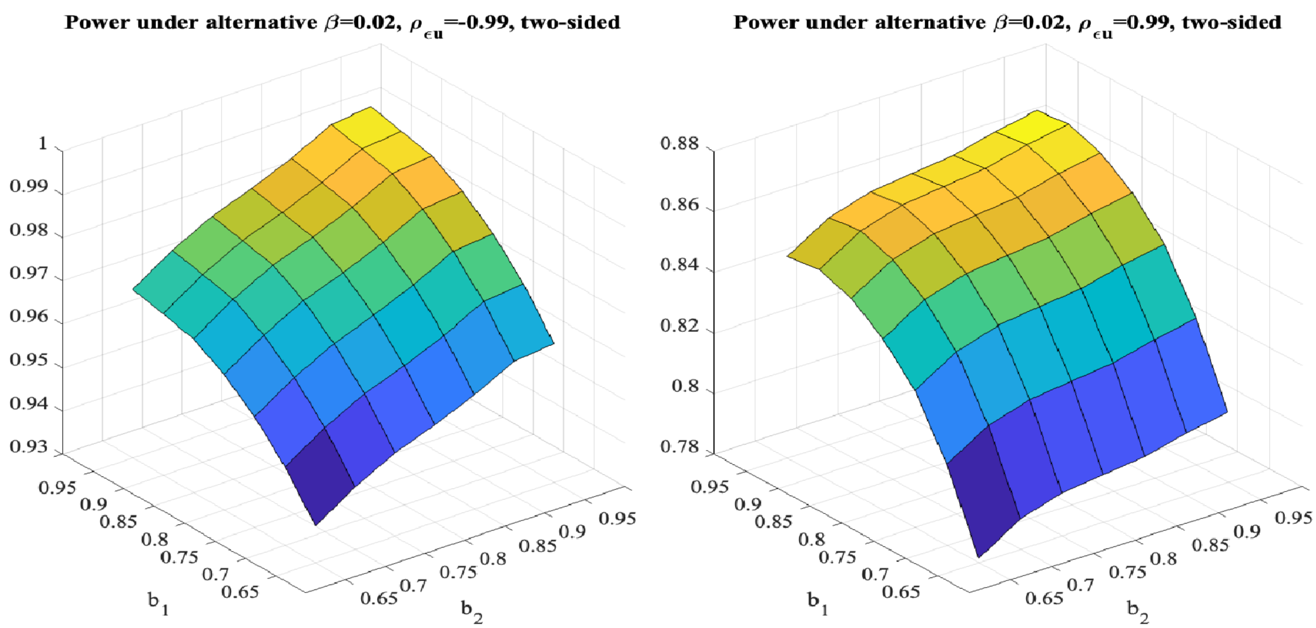



Figure B1: Power at $\rho = 1$ over grid for $b_1$ and $b_2$

# References


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

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

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

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

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

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

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

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

Magdalinos, T. (2022) Least Squares and IVX Limit Theory in Systems of Predictive Regressions with GARCH innovations. Econometric Theory, 38(5): 875-912.

Magdalinos, T. and Petrova, K. (2024) OLS limit theory for drifting sequences of parameters on the explosive side of unity, Working paper

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

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

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

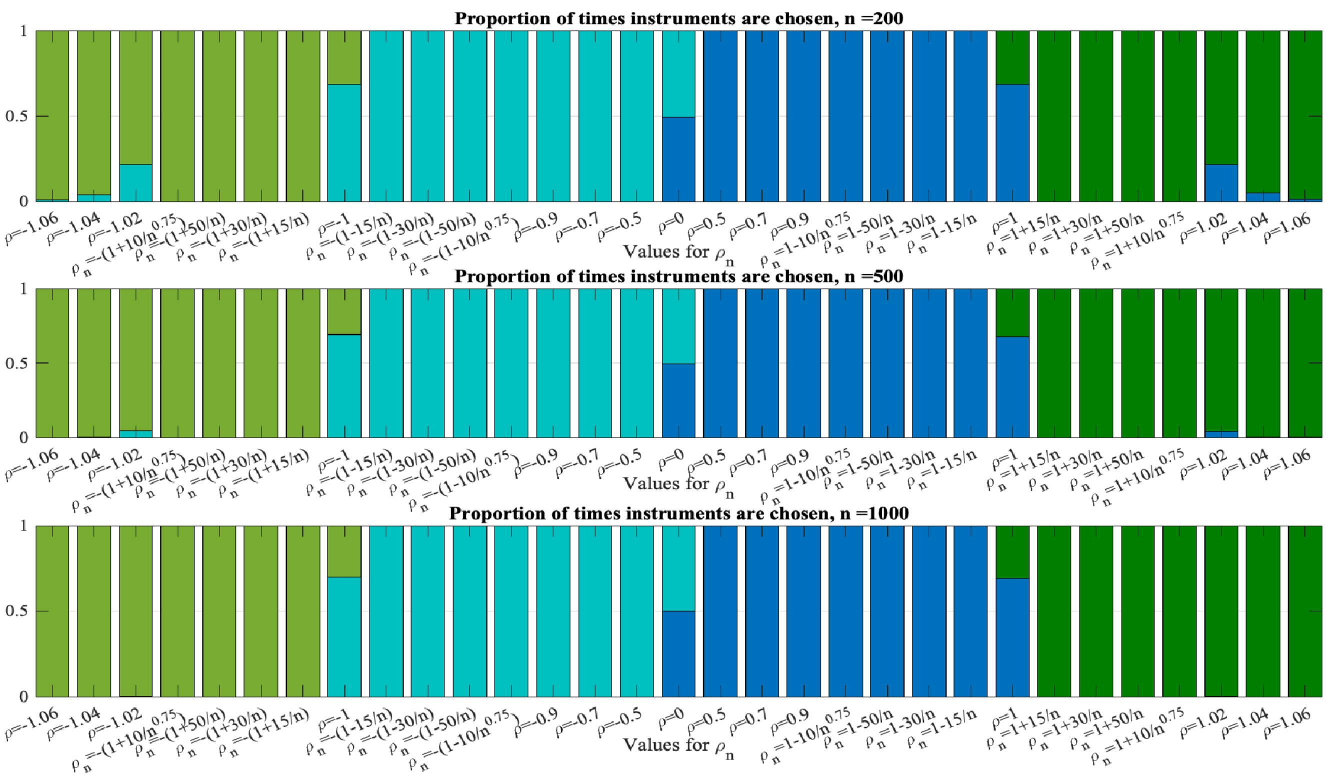




Figure B2: Proportion of times $\tilde{z}_{1t}$, $\tilde{z}_{1t}^{-}$, $\tilde{z}_{2t}$ and $\tilde{z}_{2t}^{-}$ are chosen