Giuseppe Cavaliere, Iliyan Georgiev*

Testing for unit roots in autoregressions with multiple level shifts

* Universidade Nova de Lisboa

Quaderni di Dipartimento

Serie Ricerche 2006, n.2

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Abstract

The asymptotic distributions of Augmented-Dickey-Fuller (ADF) unit root tests for autoregressive processes with a unit or near-unit root are discussed in the presence of multiple stochastic level shifts of large size occurring independently in time. The distributions depend on a Brownian motion and a Poisson-type jump process. Due to the latter, tests based on standard critical values experience power losses increasing rapidly with the number and the magnitude of the shifts. A new approach to unit root testing is suggested which requires no knowledge of either the location or the number of level shifts, and which dispenses with the assumption of independent shift occurrence. It is proposed to remove possible shifts from a time series by weighting its increments according to how likely it is, with respect to an ad hoc postulated distribution, a shift to have occurred in each period. If the number of level shifts is bounded in probability, the limiting distributions of the proposed test statistics coincide with those of ADF statistics under standard conditions. A Monte Carlo experiment shows that, despite their generality, the new tests perform well in finite samples.

Keywords: Unit roots, level shifts, compound Poisson process, random fixed point

JEL Classifications: C30, C32

1 Introduction

Since the seminal works by Perron (1989, 1990) it is well known that the performance of unit root tests is largely affected by the presence of structural level shifts which, if neglected, tend to inflate the evidence in favor of a unit root. Busetti and Harvey (2001) and Busetti and Taylor (2003) report similar findings for the stationarity tests of, inter alia Kwiatkowski et al. (1992). A major debate in this strand of the literature has been on whether the possible shift date should be regarded as known or unknown. Among others, unit root tests robust to a level shift at a known date have been developed by Perron (1989, 1990), Amsler and Lee (1995), Saikkonen and Lütkepohl (2001), Lanne et al. (2002). Tests which allow for unknown shift dates have been initially proposed by Banerjee et al. (1992), Perron and Vogelsang (1992), Zivot and Andrews (1992), and subsequently by Leybourne et al. (1998), Saikkonen and Lütkepohl (2002), Lütkepohl et al. (2004), among others; see Perron (2005) for a recent survey.

Few attempts have been made to robustify unit root tests in the presence of multiple level shifts. Lumsdaine and Papell (1997) and Clemente et al. (1998) generalize the tests proposed by Banerjee et al. (1992) by allowing for two level shifts at unknown dates. Unfortunately, there is little justification for fixing the number of shifts to one or two *a priori* (cf. Lumsdaine and Papell, 1997, p.218), and the above mentioned procedures can hardly be generalized to a larger number of shifts.[1] Recently, Kapetanios (2005) has shown how tests for unit roots can be obtained by estimating the shift dates using Bai and Perron's (1998) approach, and running a properly augmented Dickey–Fuller (ADF) regression; the critical values depend on the shift dates and the power of the tests declines as the maximum number of allowed shifts (which is assumed to be known) increases.

In this paper we take a novel approach to unit root testing in the presence of multiple level shifts. Specifically, we consider a rather general autoregressive data generating process with additive level shifts having the following features: (i) level shifts occur randomly over time; (ii) the num-

[1] An attempt has been made by Ohara (1999), who extended the Zivot and Andrews (1992) approach to the case of at most m breaks (m known). Size and power investigations (as well as critical values) are presented for the two break case only.

ber of shifts is unknown, and only needs to be bounded in probability; (iii) shifts need not occur independently over time, and in particular, may cluster together; (iv) shift sizes are random and of larger magnitude order than the shocks driving the autoregressive dynamics; (v) although shifts are exogenous throughout the paper, forms of dependence with the shocks driving the autoregressive dynamics could be allowed without affecting the results. These features differ in several respects from what has been previously considered in the literature. For instance, the investigator is not required to have any *a priori* knowledge about either the number or the location of shifts. The restrictions on the sequence of shift dates are mild, the main one being of technical nature, and requiring the total number of shifts not to grow with the sample size. This is in contrast with a strand of the literature where the number of shifts diverges together with the number of observations (cf. Balke and Fomby, 1991a, 1991b; Franses and Haldrup, 1994; Nelson et al., 2001), but we adopt it to preserve in the limit the distinction between ordinary shocks and level shifts.[2]

Despite the generality of (i)–(v), we are able to propose a family of ADF-type tests with null asymptotic distribution identical to that of ADF tests under standard conditions, and hence, without the need for new tables of critical values. Furthermore, the new tests have the same asymptotic local power function as standard ADF tests in the case of no level shifts.

Similarly to Amsler and Lee (1995) and Saikkonen and Lütkepohl (2002), the logic of the tests is to remove from the original time series, say X_t, the level shifts which might have occurred over a given sample, and then to apply standard ADF tests to the obtained 'de-jumped' time series, say $\tilde{X}_t^\delta$. In order to remove the shifts, for each observation we suggest to compute a probability $\tilde{\delta}_t$ (with respect to an ad hoc postulated distribution) that a level shift has occurred at time t, given the data. Then, shifts are removed by defining $\tilde{X}_t^\delta$ as $\tilde{X}_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s \Delta X_s$. ADF tests on $\tilde{X}_t^\delta$ are found to have the same (pivotal) limiting distributions as standard ADF tests under no level shifts – a property which holds both under the unit root null hypothesis and under local alternatives.

[2]In a recent paper, Perron and Qu (2004) adopted an approach similar to ours for evaluating the impact of structural breaks occurring randomly over time on log-periodogram estimation.

The idea of weighting each observation with a (pseudo) probability is present in the work of Franses and Lucas (1998) on the robustification of cointegration tests to innovational outliers. Despite the similarity, we depart from a rather different statistical model, which gives rise to estimators and tests with different asymptotic properties. For example, although not required for the definition of the unit root tests, in our framework consistent estimation of the shift dates is feasible, in contrast with the framework of Franses and Lucas.

To evaluate the relevance of the asymptotic theory developed in this paper, the agreement of its predictions with well-known finite-sample evidence is worth to be noted. Specifically, we show that neglecting the level shifts affects the asymptotic local power of ADF tests, which is consistent with widely documented findings that in finite samples unit root tests lose power in the presence of level shifts.

A specificity of our probability analysis is that it relies on random fixed point theory. Although it is a common practice to compute parameter estimates as fixed points of iterative algorithms, estimators themselves are rarely studied as fixed points; Aitchison and Silvey (1958) is a notable exception. By choosing a fixed-point approach, we can ensure that the object under analytical study is identical with the object which is actually computed - an obvious requirement that may otherwise be hard to check.[3]

The paper is organized as follows. In section 2 we present the reference data generating process. We also discuss the asymptotic distributions of standard unit root tests in the presence of level shifts occurring independently over time, and show that the distribution of the ADF statistics are characterized by a Poisson-type jump process, both under the null and under local alternatives. In section 3 the proposed tests for unit roots in the presence of level shifts and their asymptotic properties are introduced and discussed. In section 4 we refine our approach by proposing an algorithm for estimating the level shift dates jointly with the parameters governing the autoregressive dynamics. The small-sample properties of the new tests are analyzed through a set of Monte Carlo simulations in section 5. In section 6 we discuss how the new tests can be implemented in the case of linear time

[3]For example, numerical techniques may deliver a local maximum of a criterion function, whereas the asymptotic analysis is carried out for a global one.

trends and unknown autoregressive order. Section 7 concludes. Proofs are collected in the Appendix. The following notation is used: '$\overset{w}{\to}$' denotes weak convergence and '$\overset{P}{\to}$' convergence in P-probability, with $O_P(1)$ denoting boundedness in P-probability; $\mathbb{I}(\cdot)$ is the indicator function; $\mathbf{I}_k$ and $\mathbf{1}_k$ are the $k \times k$ identity matrix and the $k \times 1$ vector of ones. With '$x := y$' ('$x =: y$') we indicate that x is defined by y (y is defined by x), and $\lfloor \cdot \rfloor$ signifies the largest integer not greater than its argument. With $\mathcal{D}$[0,1] we denote the space of càdlàg functions on [0,1], endowed with the Skorohod topology. Finally, for any scalar sequence, e.g. $\{X_t\}$, $\Delta X_t := X_t - X_{t-1}$, $\nabla \mathbf{X}_t := (\Delta X_t, ..., \Delta X_{t-k+1})'$ and $\mathbf{X}_t := (X_t, (\nabla \mathbf{X}_t)')'$.

2 Model and preliminary results

2.1 Model and assumptions

We are interested in testing the unit root null hypothesis $\mathsf{H}_0 : \alpha = 1$ against local alternatives $\mathsf{H}_c : \alpha = 1 - c/T$ ($c > 0$), and fixed alternatives $\mathsf{H}_f : \alpha = \alpha_* \in (-1, 1)$ in the following additive level-shift model for the observable variable X_t:

$$\begin{aligned} X_t &= \varphi' Z_t + Y_t + \mu_t, \qquad t = -k, ..., T \\ Y_t &= \alpha Y_{t-1} + u_t, \\ u_t &= \textstyle\sum_{i=1}^{k} \gamma_i u_{t-i} + \varepsilon_t, \end{aligned} \tag{1}$$

where Y_t is an unobservable autoregressive process, Z_t is a $p \times 1$ vector of deterministic terms (e.g. a constant and a linear trend for $p = 2$), φ is a fixed vector conformable with Z_t, and μ_t is an unobservable level-shift component. The following assumption is maintained:

Assumption $\mathcal{M}$. (a) The roots of $\Gamma(z) := 1 - \sum_{i=1}^{k} \gamma_i z^i$ have modulus greater than 1; (b) $\{\varepsilon_t\}_{t=-k}^{\infty}$ is IID$(0, \sigma_\varepsilon^2)$ and $E|\varepsilon_1|^r < \infty$ for some $r \geq 4$.

Assumption $\mathcal{M}$ prevents Y_t from being I(2) or seasonally integrated, and ensures that the so-called long-run variance of u_t, hereafter $\sigma^2 := \sigma_\varepsilon^2 \Gamma(1)^{-2}$, is well-defined. The novelty of the paper lies in the way μ_t is specified and dealt with. To focus on this aspect, we regard the lag order k as known, and in the entire analytical part we consider the case $\varphi = 0$; the case $\varphi \neq 0$ as well as the case of unknown lag order are discussed in section 6. Initial values are set to $\mu_{-k}, ..., \mu_0 = 0$, and $(u_{-2k}, ..., u_{-k-1})'$ is given the stationary distribution induced by the difference equation $u_t = \sum_{i=1}^{k} \gamma_i u_{t-i} + \varepsilon_t$; [4] moreover, Y_{-k-1} may be any real random variable, including a constant, whose distribution is fixed and independent of T.

In the absence of level shifts and deterministic terms, $X_t = Y_t$ holds for all t; in this case, if $\hat{\alpha}$ and $\hat{\gamma}$ denote respectively the OLS estimators of α and $\gamma := (\gamma_1, ..., \gamma_k)'$ in the regression $X_t = \alpha X_{t-1} + \gamma' \nabla \mathbf{X}_{t-1} + \text{error}_t$, the well-known ADF unit root tests build on the statistics

$$\begin{aligned} ADF_{\widehat{\alpha}} &: = T(\hat{\alpha} - 1) / \hat{\Gamma}(1) \\ ADF_t &: = (\hat{\alpha} - 1) / s(\hat{\alpha}), \end{aligned}$$

where $\hat{\Gamma}(1) := 1 - \mathbf{1}_k' \hat{\gamma}$ and $s(\hat{\alpha})$ is the (OLS) standard error of $\hat{\alpha}$. Under Assumption $\mathcal{M}$ and for $\alpha = 1 - c/T$ ($c \geq 0$), it is known (see, e.g., Phillips, 1988; Chang and Park, 2002) that $\hat{\alpha} \xrightarrow{P} 1$, $\hat{\gamma} \xrightarrow{P} \gamma$ and

$$ADF_{\widehat{\alpha}} \xrightarrow{w} \frac{\int_0^1 B_c(s)\, dB_c(s)}{\int_0^1 B_c(s)^2\, ds}, \qquad ADF_t \xrightarrow{w} \frac{\int_0^1 B_c(s)\, dB_c(s)}{(\int_0^1 B_c(s)^2\, ds)^{1/2}} \tag{2}$$

as $T \to \infty$, where $B_c(s) := \int_0^s e^{-c(s-z)} dB(z)$ defines an Ornstein-Uhlenbeck process, and $B(\cdot)$ is a standard Brownian motion. Under the null hypothesis that $c = 0$, $B_c(\cdot) = B(\cdot)$ and the distributions in (2) are the univariate Dickey-Fuller distributions. Under the fixed alternative H_f:$\alpha = \alpha_* \in (0, 1)$, both $ADF_{\widehat{\alpha}}$ and ADF_t diverge to $-\infty$, and the unit root tests are consistent.

Now, suppose that the level-shift component μ_t is constant except for a few shifts. The simplest example is the single level shift model, with the level-shift component changing from μ_0 to μ_T at time T^*, $0 < T^* < T$ (cf. Perron, 1989; Amsler and Lee, 1995; Saikkonen and Lütkepohl, 2002, p.316). This model corresponds to μ_t being generated as

$$\mu_t := \mu_0 + \theta^* \mathbb{I}(t \geq T^*), \qquad \theta^* := \mu_T - \mu_0, \tag{3}$$

[4] The results of the paper remain valid under the more general specification $u_{-2k}, ..., u_{-k-1} = O_P(1)$.

with θ^* denoting the magnitude of the level shift, or equivalently, as

$$\mu_t := \mu_0 + \sum_{s=1}^{t} \delta_s \theta^*, \qquad \delta_s := \mathbb{I}(s = T^*), \tag{4}$$

with δ_s a dummy variable equal to unity at the time T^* of the level shift. Rather than assuming a single deterministic level shift, in this paper we consider shifts that occur randomly over time and have random magnitude. This is achieved by specifying the level-shift component as

$$\mu_t := \sum_{s=1}^{t} \delta_s \theta_s$$

where $\{\delta_t\}$ is an unobservable sequence of binary random variables indicating the occurrences of shifts, and $\{\theta_t\}$ is the (random) sequence of shift sizes[5]. The number of level shifts occurring up to time t is given by $N_t := \sum_{s=1}^{t} \delta_s$, with N_T denoting the total number of level shifts. The following assumption on the properties of the level shifts is required to hold jointly with Assumption $\mathcal{M}$.

Assumption $\mathcal{S}$. (a) N_T is bounded in probability conditionally on $N_T \geq 1$; (b) $\theta_t = T^{1/2}\eta_t$, where $\{\eta_t\}_{t=1}^{T}$ and $\{\eta_t^{-1}\}_{t=1}^{T}$ are sequences of $O_P(1)$ random variables as $T \to \infty$; (c) $\{\delta_t\}_{t=1}^{T}$ is independent of Y_{-k-1}, $(u_{-2k}, ..., u_{-k-1})'$, $\{\varepsilon_t\}_{t=-k}^{T}$ and $\{\eta_t\}_{t=1}^{T}$ for all T.

Several points are worth to note.

Remark 2.1. Formally, since $\{\delta_t\}$, $\{\theta_t\}$ and, under H_c, also α of (1) depend on T, we are considering a triangular array of the form $\{X_{T,t} = Y_{T,t} + \mu_{T,t}$; $t = -k, ..., T, T = 1, 2, ...\}$. The magnitudes of the level shifts could also constitute an array $\{\eta_{T,t}; t = 1, ..., T, T = 1, 2, ...\}$ with the property that, for every $\epsilon \in (0, 1)$, $P(M_\epsilon \geq |\eta_{T,t}| \geq m_\epsilon) > 1 - \epsilon$ for some constants $M_\epsilon > m_\epsilon > 0$ and $t = 1, ..., T, T = 1, 2, ...$; see Theorem 1 below. Unless differently specified, to keep notation simple we drop the 'T' subscript.

[5]As remarked above, we set $\mu_0 = 0$ in the following.

Remark 2.2. Assumption $\mathcal{S}$ generalizes the simple single shift model (3)–(4) in a number of directions. Specifically, it allows for multiple level shifts, whose number N_T only needs to be bounded in probability. This is a stochastic analogue of the deterministic setup in, among others, Perron (1989), where processes with a fixed number of structural breaks are studied. It is in contrast with the setup of, e.g., Balke and Fomby (1991a, 1991b) and Franses and Haldrup (1994), who let the number of level shifts diverge together with the sample size. Assumption $\mathcal{S}$ does not restrict the dependence structure of $\{\delta_t\}$, and in particular, allows for clusters of level shifts (including shifts at consecutive dates).

Remark 2.3. Assumption $\mathcal{S}$(b) fixes the stochastic magnitude order of level-shift sizes at $T^{1/2}$; it is not new, cf. Leybourne and Newbold (2000a, 2000b) and Perron (1989, p.1372). This choice matches the stochastic magnitude order of $\{Y_t\}$ under the hypothesis $\alpha = 1 - c/T$ and implies that level shifts have a non-negligible effect on the asymptotic distribution of ADF statistics: a desirable property given the broad evidence of a substantial effect on their finite-sample distributions. The rate of $T^{1/2}$ has also been used by Müller and Elliott (2003) to model the initial observation of processes with roots near unity. Note that Assumption $\mathcal{S}$ imposes no restriction on the dependence structure of $\{\eta_t\}$, and hence, of shift sizes.

Remark 2.4. Assumption $\mathcal{S}$(c) specifies the occurrence of level shifts as exogenous. This is not a strictly necessary assumption for the results of the paper. Thus, if P denotes the sequence of probability measures induced by model (1) under Assumptions $\mathcal{M}$ and $\mathcal{S}$, and conditional on the occurrence of at least one level shift (as will be imposed in section 3), it holds that $\max_{t:\delta_t=1} |\varepsilon_t| := \max_{1\le t\le T} |\delta_t \varepsilon_t|$, $\max_{t:\delta_t=1} |\Delta Y_t|$, $\max_{t:\delta_t=1} |\eta_t|$ and $\max_{t:\delta_t=1} |\eta_t^{-1}|$ are bounded in P-probability. What matters in the proofs of the main results is this boundedness, and as long as a relaxation of $\mathcal{S}$(c) (such as the case of endogenous level shifts) does not affect it, the results of the paper continue to hold.[6]

[6] Some kinds of dependence between δ_t and (ε_s, η_s) violate boundedness; consider, e.g., the case $\delta_t := \mathbb{I}\{|\varepsilon_t| = \max_{1\le s\le T} |\varepsilon_s|\}$, where $\max_{t:\delta_t=1} |\varepsilon_t| = \max_{1\le t\le T} |\varepsilon_t|$ and $\{\varepsilon_t\}$ is IID Gaussian.

Remark 2.5. Our model can equivalently be expressed in the multiple-break format (see, e.g., Bai and Perron, 1998) with $m := N_T$ breaks and

$$X_t = Y_t + \mu^{(j)}, \qquad t = T_{j-1} + 1, ..., T_j,\, j = 1, ..., m + 1, \qquad (5)$$

where the break points T_j are obtained as $T_j := \max\{t \in \{1, ..., T\} : N_t < j\}$, with $T_0 := -k$, $T_{m+1} := T$ and $\mu^{(j)} := \mu_{T_j}$. $\square$

2.2 Effect of level shifts on ADF tests

Our goal is to propose unit root tests with asymptotic distributions (2) under the full generality of Assumption $\mathcal{S}$. In this section we argue that these tests cannot be the standard ADF tests, since under Assumption $\mathcal{S}$ their asymptotic distributions depend on nuisance quantities. To make this point, we do not need the full generality of the assumption; instead, to render the problem more tractable, we introduce extra structure in the dependence and heterogeneity properties of the level shifts. The extra assumptions, which still represent a significant generalization of Leybourne and Newbold's (2000a) single-shift setup, are maintained only in the present sub-section.

Let $\{\delta_t\}$ be serially independent, and let $P(\delta_t = 1)$ satisfy $\sum_{s=1}^{t} P(\delta_s = 1) = G(t/T)$ for $t = 1, ..., T$, where $G : [0, 1] \to [0, \infty)$ is a continuous non-decreasing and not identically zero function such that $G(0) = 0$. For example, with $G(s) = \lambda s$ ($\lambda > 0$ and $T > \lambda$) the case of time-independent shift occurrence probability obtains: $P(\delta_t = 1) = \lambda/T$. In general, however, the probability of shift occurrence can vary in time, as, e.g., in the case $G(s) = \lambda(s - \frac{1}{2})\mathbb{I}(s > \frac{1}{2})$, where $P(\delta_t = 1)$ equals 0 and λ/T respectively in the first and the second half of the sample. Note that under the adopted specification the distance between consecutive level shifts increases with T.[7] Further, let $\theta_t = T^{1/2}\eta_t$, and let η_t be a rescaled linear process: $\eta_t := F(t/T)\sum_{i=0}^{\infty}\psi_i\omega_{t-i}$, where $F : [0, 1] \to \mathbb{R}$ is a continuous function, $\sum_{i=0}^{\infty}|\psi_i| < \infty$ and $\{\omega_t\}_{t=-\infty}^{T}$ is IID with $E|\omega_1| < \infty$.[8] For instance, similarly to u_t, $\sum_{i=0}^{\infty}\psi_i\omega_{t-i}$ could be the stationary solution of an

[7] As is needed for convergence of the level-shift process in $\mathcal{D}[0, 1]$; cf. Example VI.1.20 of Jacod and Shiryaev (2003). See also conditions (3.1)–(3.2) of Leipus and Viano (2003).

[8] Stochastic boundedness of $\{\eta_t^{-1}\}$, like in Assumption $\mathcal{S}$(b), requires further assumptions on F, $\{\varphi_i\}$ and $\{\omega_t\}$. These are relevant to the detection of level shifts, but not to Theorem 1, so we omit them.

autoregressive equation. The so defined $\{\eta_t\}$ is, in general, weakly serially dependent with time-varying variance driven by $F(\cdot)$ (see, e.g., Cavaliere, 2004), but can also be IID as a special case. Asymptotics for the ADF statistics under this specification are given in Theorem 1, where $\mathcal{P}$ denotes a Poisson process with intensity G. The process $\mathcal{P}$ is defined on $[0,1]$ and is constant apart from finitely many jumps equal to 1. Recall that σ^2 denotes the long-run variance of u_t.

Theorem 1 *Let Assumption $\mathcal{M}$ be satisfied, and let θ_t be specified as above, with $\{\delta_t\}_{t=1}^T$, $\{\omega_t\}_{t=-\infty}^T$ and $\{\varepsilon_t\}_{t=-k}^T$ jointly independent and independent of Y_{-k-1} and $(u_{-2k},...,u_{-k-1})'$. Then, under H_0 and H_c, the following convergence holds in $\mathcal{D}[0,1]$ as $T\to\infty$:*

$$\frac{1}{T^{1/2}}X_{\lfloor T\cdot\rfloor}\xrightarrow{w}\sigma\mathcal{H}_c(\cdot),\qquad \mathcal{H}_c(\cdot):=B_c(\cdot)+\frac{1}{\sigma}\mathcal{C}_F(\cdot),\tag{6}$$

where B_c and $\mathcal{C}_F$ are independent stochastic processes, B_c is an Ornstein-Uhlenbeck process with parameter c, $\mathcal{C}_F(\cdot)$ is a non-homogeneous compound Poisson process $\mathcal{C}_F(s)=\sum_{i=1}^{\mathcal{P}(s)}F(\tau_i)\eta_i^*$, with $\{\eta_i^*\}$ IID, distributed as η_1 and independent of $\mathcal{P}(\cdot)$, and with $\{\tau_i\}_{i=1}^{\mathcal{P}(1)}$ denoting the sequence of jump times of $\mathcal{P}$. Moreover,

$$ADF_{\widehat{\alpha}}\xrightarrow{w}\frac{\int_0^1\mathcal{H}_c(s)\,d\mathcal{H}_c(s)+\varkappa_0}{\int_0^1\mathcal{H}_c(s)^2\,ds},\qquad ADF_t\xrightarrow{w}\frac{\int_0^1\mathcal{H}_c(s)\,d\mathcal{H}_c(s)+\varkappa_0}{(\varkappa_1\int_0^1\mathcal{H}_c(s)^2\,ds)^{1/2}},\tag{7}$$

where $\varkappa_0$ and $\varkappa_1$ depend on $[\mathcal{C}_F]/\sigma^2$, with $[\mathcal{C}_F]:=\sum_{i=1}^{\mathcal{P}(1)}[F(\tau_i)\eta_i^*]^2$ standing for the quadratic variation of $\mathcal{C}_F$ over $[0,1]$. For $k=0$ we have $\varkappa_0=0$ and $\varkappa_1=1+[\mathcal{C}_F]/\sigma_\varepsilon^2$, while for any k, $\varkappa_0=0$ and $\varkappa_1=1$ conditionally on $\mathcal{P}(1)=0$.[9]

In Theorem 1 two effects of level shifts on the asymptotic distribution of ADF statistics under H_0 and H_c can be disentangled. First, in (6) the weak limit $\mathcal{C}_F(\cdot)$ of the level-shift process $T^{-1/2}\mu_{\lfloor T\cdot\rfloor}$ is superimposed on the weak limit $B_c(\cdot)$ of $T^{-1/2}Y_{\lfloor T\cdot\rfloor}$, in the same way as μ_t is superimposed

[9] The expressions for $\varkappa_0$ and $\varkappa_1$ for arbitrary k are given in the Appendix.

on Y_t in finite samples. As a consequence, $\mathcal{H}_c(\cdot)$ appears in (7) instead of $B_c(\cdot)$. Second, outliers that shifts in the levels of X_t induce in its first differences give rise to $\varkappa_0 \neq 0$ and $\varkappa_1 \neq 1$. For $k = 0$ the only effect is through outliers in the left-hand side regression variable ΔX_t; they lead to inconsistent estimation of $\sigma_\varepsilon^2 = Var(\varepsilon_t)$, and thus, to $\varkappa_1 \neq 1$. For $k > 0$, outliers in the lags of ΔX_t lead to inconsistent estimation of γ, and affect also the limiting distribution of $T(\hat{\alpha} - 1)$; hence, the appearance of $\varkappa_0 \neq 0$ and the more involved expression for $\varkappa_1$. The difference between (7) and the Dickey-Fuller distributions (2) arising in the absence of shifts agrees with the different finite-sample performance of ADF tests in the two cases, most notably under local alternatives, where level shifts lead to substantial power losses. The Monte Carlo simulations in section 5 illustrate this aspect.

Remark 2.6. A simple special case of distributions (7) obtains when $\eta_t = \omega_t = \eta$ is fixed (shift sizes are identical), $F \equiv 1$ and $G(s) = \lambda s$ for some $\lambda > 0$ and $T > \lambda$. Then $\mathcal{P}$ is a time-homogeneous Poisson process, $\mathcal{C}_F$ is replaced by $\eta\mathcal{P}$ in (6) and (7), and $[\mathcal{C}_F]$ is replaced by $\eta^2\mathcal{P}(1)$. The importance of the Poisson component depends on the size of the level shifts relative to the long run variance of the errors (through η/σ), as well as on the intensity of occurrence of level shifts. When the level shifts are small, i.e., η/σ is close to zero, the limiting distributions in (7) are close to the usual Dickey-Fuller type distributions. Conversely, when the shifts are large, the Poisson component becomes dominant.

Remark 2.7. The functional convergence $T^{-1/2}\mu_{\lfloor T\cdot \rfloor} \xrightarrow{w} \mathcal{C}_F(\cdot)$ in $D[0,1]$ generalizes Theorem 3.1 of Leipus and Viano (2003) to a setup where the jumps of the finite-sample process can be weakly dependent and heteroskedastic. The jumps of the limiting process $\mathcal{C}_F$ are anyway independent, since the jumps of $T^{-1/2}\mu_t$ are sampled from $\{\eta_t\}$ at intervals whose length diverges as $T \to \infty$, and since the dependence between the terms of $\{\eta_t\}$ vanishes as the time distance between them grows without bound. Functional convergence to $\mathcal{C}_F$ leads to Lemma A.3, which contains (univariate) generalizations of product-moment convergences of Georgiev (2005), and underlies (7). For $k = 0$, $\{\eta_t\}$ i.i.d. with $E\eta_1 = 0$ and $E\eta_1^2 < \infty$, $F \equiv 1$ and $G(s) = \lambda s$, in (7) convergence of the ADF_t statistic under the unit-root hypothesis follows by specializing the discussion of Georgiev

(2005, section 3) to the univariate case.

Remark 2.8. From the proof of Theorem 1 it follows that for a fixed $\alpha \in (-1, 1)$ the convergence $T^{-1/2}X_{\lfloor T\cdot \rfloor} \overset{w}{\to} \mathcal{C}_F(\cdot)$ holds in $D[0, 1]$; cf. (6). Thus, $X_{\lfloor T\cdot \rfloor}$ has the same stochastic magnitude order under H_f and H_0. As a consequence, by modifying the proof of Theorem 1 it can be shown that in our setup ADF tests are not consistent against fixed alternatives; cf. Leybourne and Newbold (2000b, Corollary 1). □

3 Tests which account for the level shifts

3.1 Overview

Given the conclusion of Theorem 1 that, in the presence of level shifts, the ADF statistics based on the observed time series X_t do not have the usual asymptotic distributions (2), here we propose modified statistics which *do* have (2) as their asymptotic distributions. Two powerful implications are that, first, standard tables of asymptotic critical values can be used for the proposed tests, and second, whatever the number and the size of the shifts, these tests have the same asymptotic local power as ADF tests in the absence of shifts.

Recall that, under Assumption $\mathcal{M}$, (2) are the asymptotic distributions of the ADF statistics from the regression $Y_t = \alpha Y_{t-1} + \gamma' \nabla \mathbf{Y}_{t-1} + \text{error}_t$, which, being Y_t unobservable, is not feasible empirically. Since $Y_t = X_t - \mu_t$ can be thought of as obtained form X_t by removing the (also unobservable) level shifts, we propose to conduct ADF tests on a process obtained by subtracting from X_t an estimator of μ_t. The idea is related to Saikkonen and Lütkepohl (2002), who suggest to adjust the original time series by removing an estimator of its deterministic component, including possible (deterministic) level shifts. Since in our case μ_t is a random jump process, in what follows our procedure is referred to as 'de-jumping'.

If δ_t were observable, but the shift sizes not, to optimize test power under alternatives close to the null we could estimate the shift sizes by pseudo-GLS under local alternatives, as suggested by Elliott et al. (1996). Furthermore, to keep under control the finite-sample size of the resulting tests, following the suggestion of Saikkonen and Lütkepohl (2002) and Lanne et al. (2002), we would set the localizing parameter to zero. This reduces to estimation of shift sizes under the unit-root null, i.e., to a regression of ΔX_t on impulse dummy variables, one per shift. The implied estimator of μ_t would be

$$\hat{\mu}_t := \sum_{s=1}^{t} \delta_s \Delta X_s = \sum_{s=1}^{t} \delta_s \theta_s + \sum_{s=1}^{t} \delta_s \Delta Y_s = \mu_t + \sum_{s=1}^{t} \delta_s \Delta Y_s.$$

The estimation error $\hat{\mu}_t - \mu_t = \sum_{s=1}^{t} \delta_s \Delta Y_s$ is bounded in probability uniformly in t (see Lemma A.2 in the Appendix), while μ_t (as well as Y_t under H_0 and H_c) has stochastic magnitude order $T^{1/2}$. This difference turns out to be sufficient for the ADF statistics based on the 'de-jumped' series $\hat{X}_t^{\delta} := X_t - \hat{\mu}_t = Y_t + (\mu_t - \hat{\mu}_t)$, $t = 1, ..., T$, ($\hat{X}_t^{\delta} := X_t = Y_t$ for $t = -k, ..., 0$) to have the null and local-to-null asymptotic distributions given in (2), and to diverge under fixed alternatives.

In the case of unobservable δ_t we imitate the above de-jumping scheme through the following three-step procedure:

1. the sequence of level shift indicators $\{\delta_t\}$ is consistently estimated by an estimator $\{\tilde{\delta}_t\}$; the level shift component is estimated by $\tilde{\mu}_t^{\delta} := \sum_{s=1}^{t} \tilde{\delta}_s \Delta X_s$;
2. the level shifts are removed by constructing the de-jumped series $\tilde{X}_t^{\delta} := X_t - \tilde{\mu}_t^{\delta}$ ($t \geq 1$) and $\tilde{X}_t^{\delta} := X_t$ ($t = -k, ..., 0$);
3. standard ADF tests, say $ADF_{\widehat{\alpha}}^{\delta}$ and ADF_t^{δ}, are computed using $\tilde{X}_t^{\delta}$ instead of the original time series; the unit root null hypothesis is evaluated by comparing $ADF_{\widehat{\alpha}}^{\delta}$ and ADF_t^{δ} to standard asymptotic critical values (see, e.g., Fuller, 1976).

In this framework, the key step is the consistent estimation of level shift indicators. Specifically, since $\tilde{X}_t^{\delta} = Y_t - \sum_{s=1}^{t} \delta_s \Delta Y_s - \sum_{s=1}^{t} (\tilde{\delta}_s - \delta_s)\theta_s = \hat{X}_t^{\delta} - \sum_{s=1}^{t} (\tilde{\delta}_s - \delta_s)\theta_s$, for the ADF test obtained from 1–3 above not to be influenced asymptotically by the fact that δ_t are estimated, $\sum_{t=1}^{T} |\tilde{\delta}_t - \delta_t|$ should converge to zero sufficiently fast. In the next sub-subsection we propose an estimator $\tilde{\delta}_t$ with this property; it is the expectation of δ_t conditional on the data, with respect to an ad hoc postulated distribution. This estimator offers the flexibility of not requiring an explicit decision rule about the location of shifts, although as a by-product it gives rise to such a rule with a traditional interpretation.

3.2 Level shift estimation and unit root tests

Our proposed level shift estimation method starts by noticing that the distribution of ΔX_t is a mixture, with mixing variable δ_t and mixture components[10] ΔY_t (when $\delta_t = 0$) and $\Delta Y_t + \theta_t$ (when $\delta_t = 1$). Therefore, estimating the shift indicators δ_t is equivalent to classifying the observable increments ΔX_t into such equal to ΔY_t and such contaminated by θ_t. Since under Assumption $\mathcal{S}$ the mixture components have different orders of magnitude, they are 'well-separated' and consistent classification is feasible.

In the spirit of quasi maximum likelihood [QML] estimation,[11] we replace (for the basic version of our estimator) the true probability distribution of the increments $\{\Delta X_t\}_{t=1}^T$ by the distribution of a sequence of T independent draws from the following mixture:

$$(\delta_t, \Delta X_t) \sim \begin{cases} (0, \sigma \cdot t\,(\nu)) & \text{with probability } 1 - \lambda/T \\ \big(1, (\sigma^2 + T\eta^2)^{1/2} \cdot t\,(\nu)\big) & \text{with probability } \lambda/T, \end{cases} \tag{8}$$

where $t\,(\nu)$ denotes a Student t distribution with ν degrees of freedom. The corresponding quasi-likelihood function for λ, when σ^2 and η^2 are regarded as known, is

$$\prod_{t=1}^T \Big(\frac{\lambda}{T}\phi_\nu(\Delta X_t; \sigma^2 + T\eta^2) + (1 - \frac{\lambda}{T})\phi_\nu(\Delta X_t; \sigma^2)\Big),$$

where, for $a > 0$, $\phi_\nu(e; a^2)$ is the probability density function of $a \cdot t\,(\nu)$.

[10] Although we write as if the two component distributions are the same for all t, in general they constitute a family indexed by t (e.g. η_t does not need to be identically distributed over time); cf. Remark 2.1.

[11] The level shift estimator is inspired by the specification in Remark 2.6; nevertheless, all the results in what follows hold in the more general setup of Assumption $\mathcal{S}$.

The QML estimator of λ then satisfies the equation

$$\lambda = \Phi_T^{\lambda}(\zeta) := \sum_{t=1}^{T} \tilde{\delta}(\Delta X_t; \zeta), \qquad \zeta := (\lambda, \eta^2, \sigma^2)', \tag{9}$$

with

$$\tilde{\delta}(e;\zeta) := \frac{\lambda}{T}\Big[\frac{\lambda}{T} + \Big(1 - \frac{\lambda}{T}\Big)\frac{\phi_\nu(e;\sigma^2)}{\phi_\nu(e;\sigma^2 + T\eta^2)}\Big]^{-1}. \tag{10}$$

We abbreviate $\tilde{\delta}(\Delta X_t;\zeta)$ to $\tilde{\delta}_t(\zeta)$. Note that $\tilde{\delta}_t(\zeta) = \tilde{E}(\delta_t | \{\Delta X_s\}_{s=1}^T) = \tilde{E}(\delta_t|\Delta X_t)$, with $\tilde{E}(\cdot| \{\Delta X_s\}_{s=1}^T)$ denoting expectation conditional on the data under distribution (8). Thus, $\tilde{\delta}_t(\zeta)$ measures how likely it is, given the data and the postulated distribution, a level shift to have occurred in period t.

From (9) it follows that, for fixed σ^2 and η^2, the QML estimator of λ is a *fixed point* of the map $\lambda \mapsto \sum_{t=1}^T \tilde{\delta}_t(\zeta)$. This allows to retrieve the estimator of the level shift indicators straightforwardly. First, the fixed point can be computed by iterating the map until convergence. Second, the resulting estimate of λ can be inserted into (10), yielding the estimates of δ_t needed to eliminate the level shifts from the original time series.

The pseudo parameters η^2 and σ^2 are related to the squared magnitude of the draws from the mixture components, respectively $T^{-1}\sum_{t=1}^T \delta_t(\Delta X_t)^2$ and $T^{-1}\sum_{t=1}^T (1-\delta_t)(\Delta X_t)^2$. Thus, instead of fixing η^2 and σ^2 *a priori*, we suggest to replace them with

$$\Phi_T^{\eta}(\zeta) \;:\; = \frac{1}{T}\sum_{t=1}^T \tilde{\delta}_t(\zeta)(\Delta X_t)^2 \tag{11}$$

$$\Phi_T^{\sigma}(\zeta) \;:\; = \frac{1}{T}\sum_{t=1}^T (1 - \tilde{\delta}_t(\zeta))(\Delta X_t)^2.$$

The estimator of $\{\delta_t\}_{t=1}^T$ is then obtained as $\{\tilde{\delta}_t(\zeta_T)\}_{t=1}^T$, where ζ_T is a random fixed point of the random map Φ_T obtained by combining eqs. (9) and (11):

$$\Phi_T(\zeta) := (\Phi_T^{\lambda}(\zeta), \Phi_T^{\eta}(\zeta), \Phi_T^{\sigma}(\zeta))'. \tag{12}$$

Once ζ_T has been found, level shifts can be removed from X_t by computing the series $\tilde{X}_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s(\zeta_T)\Delta X_s$ ($t \geq 1$), $\tilde{X}_t^\delta := X_t$ ($t = -k, ..., 0$), and the unit root hypothesis can be tested by conducting standard ADF tests, henceforth $ADF_{\widehat{\alpha}}^\delta$ and ADF_t^δ, on the ADF regression $\tilde{X}_t^\delta = \alpha\tilde{X}_{t-1}^\delta + \gamma'\nabla\tilde{\mathbf{X}}_{t-1}^\delta + \text{error}_t$ (recall that $\nabla\tilde{\mathbf{X}}_t^\delta := (\Delta\tilde{X}_t^\delta, ..., \Delta\tilde{X}_{t-k+1}^\delta)'$). In the next sub-section we show that $\tilde{\delta}_t(\zeta_T)$ estimate δ_t consistently at a rate fast enough for $ADF_{\widehat{\alpha}}^\delta$ and ADF_t^δ to have the usual Dickey-Fuller type asymptotic distributions under the null and under local alternatives. Under fixed alternatives the statistics give rise to a consistent test.

Remark 3.1. The plot of $\tilde{\delta}\,(e;\zeta)$ as a function of $|e|$ is an S-curve starting at λ/T for $e = 0$, and approaching unity as $|e| \to \infty$. For larger values of ν in (8), the curve is closer to that of a step function, in the sense that it remains closer to 0 for small $|e|$, exhibits steeper increase for intermediate values of $|e|$, and is closer to 1 for large $|e|$. As extensive simulations have shown, larger values of ν translate into a more sensitive estimator of $\{\delta_t\}$ in finite samples, and into unit root tests with better finite-sample power. We found no further power gains associated with the use of binary 0-1 estimators instead of our smooth estimators $\tilde{\delta}\,(\cdot;\zeta)$. That is, a 0-1 estimator defined as equal to unity if and only if $\tilde{\delta}\,(\cdot;\zeta) > \kappa$, $\kappa \in (0,1)$, induces for some values of κ tests with power comparable to that of a smooth estimator with large ν, whereas for other values of κ the 0-1 estimator yields lower power. Moreover, the optimal value of κ depends on the DGP. For these reasons, we prefer to base our tests on the smooth estimators and not on a 0-1 estimator.

Remark 3.2. Recall that, under model (1), a level shift occurring at time t generates a large outlier in the differenced series ΔX_t. Hence, an approach alternative to our smooth level shift estimator is the application to ΔX_t of traditional outlier detection methods (see, e.g., Chen and Tiao, 1990, Chen and Liu, 1993, and references therein). Simulations have shown that implementation of these methods in conjunction with de-jumping results in oversized tests, with no gains in size-adjusted power. Interestingly, our approach can, but need not, operate as a traditional method. Traditional methods typically classify as outliers regression residuals which, upon standardization, exceed some threshold. In our framework, the inequality $\tilde{\delta}(e;\zeta) > \kappa$, $\kappa \in (0,1)$, is equivalent to $|e/\sigma| > \tau_{T,\nu}(\kappa,\lambda,\eta/\sigma)$, for a function $\tau_{T,\nu}$ which can be found explicitly. Hence, if we evaluate $\tilde{\delta}(\cdot;\zeta)$ at a regression residual, as we essentially do in section 4, and compare the value to a threshold κ, this is equivalent to a comparison of the standardized residual to a threshold depending on κ, on the occurrence intensity of shifts, and on their relative size. As mentioned in the previous remark, in terms of test properties, there seems to be little payoff to the complication of choosing an appropriate κ.

Remark 3.3. As our model can be cast within the multiple break format (5), another approach to the estimation of level-shift dates is the procedure of Bai and Perron (1998). Again, extensive Monte Carlo experiments have shown that in our framework this procedure does very well in detecting the correct level shift dates under fixed alternatives, with unit root tests based on Bai-Perron de-jumping having power properties comparable to ours. However, in our experience unit root tests based on Bai-Perron de-jumping tend to have worse size properties and, under local alternatives, worse size-adjusted power.[12] Another distinct advantage of our approach over the Bai-Perron method (as well as over many traditional outlier detection methods) is that our tests work also when level shifts occur at consecutive dates; more generally, we do not need the distance between consecutive level shifts to grow with T. $\square$

3.3 Asymptotic results

We start this sub-section with a theorem about existence, uniqueness and asymptotics for a fixed point ζ_T of the mapping Φ_T, see (12), such that the estimator of the level shift indicators $\tilde{\delta}_t(\zeta_T)$ is consistent for δ_t uniformly in t.[13] The theorem also ensures that ζ_T can easily be computed by iterating Φ_T until convergence.

[12]A possible explanation for this result is that the Bai-Perron procedure delivers unsatisfactory results when the level shift component μ_t and the (near-) integrated component Y_t are both of order $O_P(T^{1/2})$, as in our model under the null hypothesis and under local alternatives.

[13]It is immediate that Φ_T also has the trivial fixed point $(0, 0, T^{-1}\sum_{t=1}^{T}(\Delta X_t)^2)'$, which however fails to satisfy the consistency requirement.

Theorem 2 *Let P be the probability measure induced, under Assumptions $\mathcal{M}$ and $\mathcal{S}$, by model (1) with either $\alpha = 1 - c/T$, $c \geq 0$, or $\alpha = \alpha_* \in (-1, 1)$, and conditional on the realization of at least one level shift. If ν of (8) satisfies $3 \leq \nu \leq r - 1$, then:*

EXISTENCE. *There exists a random sequence $\{\zeta_T\}$ such that $P(\Phi_T(\zeta_T) = \zeta_T) \to 1, \zeta_T = O_P(1)$ and ζ_T is component-wise bounded away from zero in P-probability.*

UNIQUENESS. *If ζ_T and ς_T both have the above properties, then $P(\zeta_T = \varsigma_T) \to 1$.*

COMPUTABILITY. *For every non-random ζ_{T0} with positive coordinates, the sequence of iterates $\zeta_{Ti} := \Phi_T(\zeta_{T,i-1})$, $i \geq 1$, satisfies $P(\zeta_{Ti} \to_{i\to\infty} \zeta_T) \to 1$ as $T \to \infty$.*

CONSISTENCY. *$\sum_{t=1}^T |\delta_t - \tilde{\delta}_t(\zeta_T)| = O_P(T^{-1/2})$, $\sum_{t=1}^T \delta_t(1 - \tilde{\delta}_t(\zeta_T)) = O_P(T^{-(\nu-2)/2})$ and $\zeta_T = (N_T, H_T, \sigma_Y^2) + o_P(1)$, where $H_T := \sum_{t=1}^T \delta_t \eta_t^2$ and $\sigma_Y^2 := P\lim_{T\to\infty} Var(\Delta Y_T)$.*

Thus, ζ_T is consistent for (N_T, H_T, σ_Y^2), and $\{\tilde{\delta}_t(\zeta_T)\}_{t=1}^T$ is consistent for the sequence of level-shift indicators $\{\delta_t\}_{t=1}^T$ at the rate of $T^{1/2}$. As a consequence (not used in the sequel), for any sequence $a_T = const \in (0, 1)$ or $a_T = T^{\epsilon - 1/2}$, $0 < \epsilon < 1/2$, the set of periods with $\tilde{\delta}_t(\zeta_T) > 1 - a_T$ coincides, with P-probability approaching one, with the set of periods where level shifts occur. This is a consistent explicit level-shift detection rule.

Remark 3.4. Adopt the notation $z = (z^\lambda, z^\eta, z^\sigma)'$ for vectors $z \in \mathbb{R}^3$. According to the consistency part of Theorem 2, $(\zeta_T^\lambda, \zeta_T^\eta) = (N_T, H_T) + o_P(1)$. This introduces a small complication into the argument for existence of ζ_T, for in general no compact in $\mathbb{R}^3$ contains N_T and H_T (and hence, $(\zeta_T^\lambda, \zeta_T^\eta)$) with probability approaching one, while existence theorems for random fixed points are typically formulated for self-maps of compacts. One consequence is that to prove the theorem we resort to the trick of looking for a fixed point of the form $\zeta_T = \zeta(z_T)$, where ζ is the random function $\zeta(z) = (N_T + z^\lambda, H_T z^\eta, z^\sigma)'$, and $(z_T^\lambda, z_T^\eta) = (0, 1) + o_P(1)$. For $H_T \neq 0$ the inverse function ζ^{-1} is well-defined, and z_T can be found as a fixed point of $\zeta^{-1} \circ \Phi_T \circ \zeta$. To ensure that we can avail of this fact, Theorem 2 is stated conditionally on the presence of level shifts $(N_T \geq 1)$, so that $H_T \neq 0$ a.s. The case $N_T = 0$ is considered in the simulation exercise. $\square$

Before introducing in Theorem 3 our main result, we present an important lemma which forms the basis of the proofs of both Theorems 2 and 3. The lemma contains several convergence statements uniform on compacts. One regards the rates of consistent estimation of $\{\delta_t\}$; another one, the distance between the OLS estimators of α and γ from ADF regressions for de-jumped series and for the unobserved Y_t; yet another statement is an evaluation of the Jacobian matrix of Φ_T that underlies contraction arguments. In the lemma, for compacts $\mathbb{A}_T$ in $\mathbb{R}^3$ and real functions f defined on $\mathbb{A}_T$, we write $\sup_{z\in\mathbb{A}_T} f(\zeta(z))$ as $\sup_{\mathbb{A}_T} f(\zeta)$.

Lemma 1 *Let ζ be the random function defined by $\zeta(z) := (N_T + z^\lambda, H_T z^\eta, z^\sigma)'$, or the identity function on $\mathbb{R}^3$, and let the compact $\mathbb{A}_T \subset \mathbb{R}^3$ be such that $\max_{z\in\mathbb{A}_T} \zeta^\upsilon(z) = O_P(1)$ and $\min_{z\in\mathbb{A}_T} \zeta^\upsilon(z)$ are bounded away from zero in P-probability, $\upsilon \in \{\lambda, \eta, \sigma\}$. Define $X_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s(\zeta)\Delta X_s$ ($t \geq 1$), $X_t^\delta := X_t$ ($t = -k, ..., 0$), and let $\nabla \mathbf{X}_t^\delta$ be defined accordingly. Under the conditions of Theorem 2, it holds that:*

a. $\sup_{\mathbb{A}_T} \sum_{t=1}^T |\delta_t - \tilde{\delta}_t(\zeta)| = O_P(T^{-1/2})$ *and* $\sup_{\mathbb{A}_T} \sum_{t=1}^T \delta_t(1 - \tilde{\delta}_t(\zeta)) = O_P(T^{-(\nu-2)/2})$*;*

b. $\sup_{\mathbb{A}_T}(|N_T - \Phi_T^\lambda(\zeta)|, |H_T - \Phi_T^\eta(\zeta)|, |\sigma_Y^2 - \Phi_T^\sigma(\zeta)|) = o_P(1)$ *component-wise;*

c. $\sup_{\mathbb{A}_T} \|(T(\hat{\alpha} - \alpha_\delta), T^{1/2}(\hat{\gamma} - \gamma_\delta)')\| = o_P(1)$ *and* $\sup_{\mathbb{A}_T} |\hat{\sigma}_\varepsilon^2 - \sigma_{\varepsilon,\delta}^2| = o_P(1)$ *if* $\alpha = 1 - c/T$ ($c \geq 0$)*, where $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\sigma}_\varepsilon^2$ (α_δ, γ_δ and $\sigma_{\varepsilon,\delta}^2$) are the OLS estimators of α, γ and σ_ε^2 from the regression $Y_t = \alpha Y_{t-1} + \gamma'\nabla \mathbf{Y}_{t-1} + error_t$ ($X_t^\delta = \alpha X_{t-1}^\delta + \gamma'\nabla \mathbf{X}_{t-1}^\delta + error_t$);*

d. $\sup_{z\in\mathbb{A}_T} \|(\Phi_T)'_\varsigma|_{\varsigma=\zeta(z)}\| = o_P(1)$*, where $(\Phi_T)'_\varsigma$ is the Jacobian matrix of $\Phi_T(\varsigma)$.*

Remark 3.5. Throughout, the random functions appearing in suprema, infima, maxima and minima are continuous and have separable domains (subsets of $\mathbb{R}^n$). Hence, the extrema in question are random variables.

Remark 3.6. The lemma states that the OLS estimators of α and γ based on de-jumped data and on the unobservable Y_t are asymptotically equivalent under the null and under local alternatives. Asymptotic equivalence holds for any z belonging to any $\mathbb{A}_T$, not only for a fixed point of $\zeta^{-1} \circ \Phi_T \circ \zeta$. However, since the fixed point matches relevant sample characteristics, it can be expected that a test based on it has better finite-sample properties. As can be seen from the proof of Lemma A.7 in the Appendix, under fixed alternatives α_δ and γ_δ are asymptotically equivalent to the OLS estimators from the unfeasible regression for data de-jumped with the true $\{\delta_t\}$ instead of $\{\tilde{\delta}_t(\zeta)\}$ ($\hat{X}_t^\delta$ in the notation of section 3.1), but not to the estimators from a regression for Y_t. This is not surprising given that de-jumping was aimed at optimizing test properties close to the unit-root null. A similar phenomenon, which does not compromise the consistency of the tests, can be seen in the proof of Theorem 4.2 of Saikkonen and Lütkepohl (2002). □

We are now ready to state our main result.

Theorem 3 *Suppose that the conditions of Theorem 2 are satisfied. Let ζ_T be as in Theorem 2, and $\tilde{X}_t^\delta$ denote X_t^δ of Lemma 1 evaluated at ζ_T. Under H_0 and H_c, the ADF statistics $ADF_{\widehat{\alpha}}^\delta$, ADF_t^δ based on the regression $\tilde{X}_t^\delta = \alpha \tilde{X}_{t-1}^\delta + \gamma' \nabla \tilde{\mathbf{X}}_{t-1}^\delta + error_t$ have asymptotic distributions given by (2), i.e., as $T \to \infty$*

$$ADF_{\widehat{\alpha}}^\delta \xrightarrow{w} \frac{\int_0^1 B_c(s)\, dB_c(s)}{\int_0^1 B_c(s)^2\, ds}, \qquad ADF_t^\delta \xrightarrow{w} \frac{\int_0^1 B_c(s)\, dB_c(s)}{(\int_0^1 B_c(s)^2\, ds)^{1/2}},$$

where weak convergence refers to the sequence of measures conditional on $N_T \geq 1$. Under H_f, $ADF_{\widehat{\alpha}}^\delta \xrightarrow{P} -\infty$ and $ADF_t^\delta \xrightarrow{P} -\infty$ as $T \to \infty$.

According to Theorem 3, in the presence of level shifts the asymptotic distribution of the ADF statistics based on the de-jumped time series $\tilde{X}_t^\delta$ is the same as that of standard ADF statistics computed from series with no level shifts, both under the null and under local alternatives. This result allows to refer to well-known tables of critical values (cf. e.g. Fuller, 1976). Moreover, under fixed alternatives the proposed tests are consistent.

4 Joint estimation of model parameters and level-shift indicators

The unit root tests described earlier are based on a sequential procedure: initially, level-shift indicators δ_t are estimated by looking at the observable increments $\Delta X_t = \Delta Y_t + \theta_t \delta_t$; later, the parameters of model (1) are estimated by an ADF regression on the de-jumped series. In this section we discuss a procedure for joint estimation of level-shift indicators and model parameters. Interest is not in parameter estimation *per se*, but rather in the possibility to obtain estimates of δ_t based on estimates of the innovations $\varepsilon_t + \delta_t \theta_t$, instead of $\Delta Y_t + \delta_t \theta_t$. As ε_t has smaller variance than ΔY_t, ε_t and $\varepsilon_t + \theta_t$ are better separated than ΔY_t and $\Delta Y_t + \theta_t$, so the joint procedure can be expected to have better finite-sample properties than the sequential one.

In our setup, the joint estimation problem can be solved computationally by means of a simple iterative procedure. Specifically, the following steps can be followed: (i) estimation of level shift indicators (where, at the first iteration, the estimates obtained from the basic estimation procedure, see section 3.2, are used); (ii) construction of a de-jumped time series; (iii) estimation of the regression coefficients α and γ by a standard ADF regression on the de-jumped series; (iv) computation of estimates of $\varepsilon_t + \theta_t \delta_t$. These estimates serve as a new input for step (i), and the procedure is iterated until convergence. The unit root test statistics are computed within this procedure at step (iii).

As for the basic de-jumping of section 3.2, the above procedure delivers a fixed point, say ξ_T, of a random map, say Ψ_T. Compared to Φ_T from (12), Ψ_T is of larger and sample-dependent dimension, as it has components for the parameters α and γ, and also for each δ_t, $t = 1, ..., T$. The definition follows.

For a given $d = (d_1, ..., d_T)'$, denote the process X_t de-jumped with weights d by $X_t^d := X_t - \sum_{s=1}^t d_s \Delta X_s \,(t \geq 1)$ and $X_t^d := X_t$ ($t = -k, ..., 0$), so that $\mathbf{X}_t^d := (X_t^d, (\nabla \mathbf{X}_t^d)')'$ ($t \geq 0$) as agreed in section 1. Further, for a given (a, γ'), let $e_t^d := \Delta X_t - (a, \gamma')\, D_T \mathbf{X}_{t-1}^d$ ($t \geq 1$), where a is inserted for $T^{1/2}(\alpha - 1)$, and D_T is the diagonal $(k+1) \times (k+1)$ matrix $\mathrm{diag}(T^{-1/2}, 1, ..., 1)$. Define Ψ_T, with argument $\xi := (d', a, \gamma', \lambda, \eta^2, \sigma^2)'$

and with components $\Psi_T^\delta, \Psi_T^{a,\gamma}, \Psi_T^\lambda, \Psi_T^\eta$ and Ψ_T^σ, as follows. First, let $\Psi_T^\delta := (\Psi_{T1}^\delta, ..., \Psi_{TT}^\delta)'$, $\Psi_{Tt}^\delta(\xi) := \tilde{\delta}(e_t^d; \lambda, \eta^2, \sigma^2)$ for $t = 1, ..., T$, see (10), and $\Psi_T^\lambda(\xi) := \sum_{t=1}^T \Psi_{Tt}^\delta(\xi)$. Denote the series X_t de-jumped with $\Psi_T^\delta(\xi)$ by $X_t^\Psi := X_t - \sum_{s=1}^t \Psi_{Ts}^\delta(\xi) \Delta X_s$ $(t \geq 1)$ and $X_t^\Psi := X_t$ $(t = -k, ..., 0)$; for $\mathbf{X}_t^\Psi$ defined accordingly, define $\Psi_T^{a,\gamma}(\xi)$ as the estimator of $(a, \gamma')'$ from the regression

$$\Delta X_t^\Psi = (a, \gamma') D_T \mathbf{X}_{t-1}^\Psi + \text{error}_t. \tag{13}$$

Finally, using $e_t^\Psi := \Delta X_t - (\Psi_T^{a,\gamma})' D_T \mathbf{X}_{t-1}^\Psi$ as an updated estimator of $\varepsilon_t + \delta_t \theta_t$, define

$$\Psi_T^\eta(\xi) := T^{-1} \sum_{t=1}^T \Psi_{Tt}^\delta(\xi) (e_t^\Psi)^2$$

and

$$\Psi_T^\sigma(\xi) := T^{-1} \sum_{t=1}^T (1 - \Psi_{Tt}^\delta(\xi))(e_t^\Psi)^2 .$$

For a fixed point ξ_T of Ψ_T, consider the ADF statistics, say $ADF_{\hat{\alpha}}^\Psi$ and ADF_t^Ψ, from regression (13) for X_t^Ψ evaluated at $\xi = \xi_T$. The next theorem discusses existence and uniqueness of a fixed point ξ_T such that $\Psi_{Tt}^\delta(\xi_T)$ estimate δ_t consistently. It also states conditions ensuring that $ADF_{\hat{\alpha}}^\Psi$ and ADF_t^Ψ have the usual asymptotic distributions (2) both under the null hypothesis and under local alternatives.

Theorem 4 *Let the conditions of Theorem 2 hold. Under* H_0 *and* H_c, $c \geq 0$, *then:*

EXISTENCE. *There exists a random sequence* $\{\xi_T\}$ *such that:*

i. $P(\Psi_T(\xi_T) = \xi_T) \to 1$, *i.e.* ξ_T *is a fixed point of* Ψ_T *with* P-*probability tending to 1;*

ii. CONSISTENCY. $\{\Psi_{Tt}^\delta(\xi_T)\}$ *estimate* $\{\delta_t\}$ *consistently:* $\sum_{t=1}^T |\delta_t - \Psi_{Tt}^\delta(\xi_T)| = O_P(T^{-1/2})$ *and* $\sum_{t=1}^T \delta_t(1 - \Psi_{Tt}^\delta(\xi_T)) = O_P(T^{-(\nu-2)/2})$, *whereas* $(\Psi_T^\lambda(\xi_T), \Psi_T^\eta(\xi_T), \Psi_T^\sigma(\xi_T), (\Psi_T^\gamma(\xi_T))') = (N_T, H_T, \sigma_\varepsilon^2, \gamma') + o_P(1)$*;*

iii. UNIT ROOT TESTS. *The limiting distributions of $ADF^{\Psi}_{\tilde{\alpha}}$ and ADF^{Ψ}_{t} are given by*

$$ADF^{\Psi}_{\tilde{\alpha}} \xrightarrow{w} \frac{\int_0^1 B_c(s)\,dB_c(s)}{\int_0^1 B_c(s)^2\,ds}, \qquad ADF^{\Psi}_{t} \xrightarrow{w} \frac{\int_0^1 B_c(s)\,dB_c(s)}{(\int_0^1 B_c(s)^2\,ds)^{1/2}},$$

where weak convergence refers to the sequence of measures conditional on $N_T \geq 1$.

Furthermore, if part (b) of Assumption $\mathcal{M}$ is replaced by the requirement that $E|\varepsilon_1|^r < \infty$ for some $r \geq 5$, then for $4 \leq \nu \leq r-1$ the following facts hold too.

UNIQUENESS. *If ξ_T and ς_T have properties (i) to (iii), then $P(\xi_T = \varsigma_T) \to 1$.*

COMPUTABILITY. *Let $\tilde{\delta}(\zeta_T) := (\tilde{\delta}_1(\zeta_T), ..., \tilde{\delta}_T(\zeta_T))'$ and $\xi_{T0} := (\tilde{\delta}(\zeta_T)', T^{1/2}(\tilde{\alpha}_\delta - 1), \tilde{\gamma}'_\delta, \zeta'_T)'$, with ζ_T as in Theorems 2 and $(\tilde{\alpha}_\delta, \tilde{\gamma}'_\delta)$ obtained from the regression in Theorem 3. Let also $\xi_{Ti} := \Psi_T(\xi_{T,i-1})$ for $i \geq 1$. Then $P(\xi_{Ti} \to_{i\to\infty} \xi_T) \to 1$ as $T \to \infty$.*

Remark 4.1. The theorem is similar to Theorems 2 and 3, but weaker in some respects. Uniqueness is established within a smaller class of sequences than in Theorem 2, but this has no practical implications, given that the iterative algorithm from the computability part above converges to the fixed point ξ_T with properties (i)-(iii). The choice of initial value for the iteration is important: Theorems 2 and 3 ensure that ξ_{T0}, defined through the fixed point of the basic de-jumping procedure, is sufficiently close to ξ_T for the iteration to converge.

Remark 4.2. Theorem 4 does not cover fixed alternatives, which are studied in the next section by simulation. □

5 Monte Carlo results

In this section the finite-sample size and power properties of standard ADF tests ($ADF_{\widehat{\alpha}}$, ADF_t) and of the proposed ADF tests based on level-shifts removal are investigated by Monte Carlo simulation, for DGPs either with or without level shifts. We need to establish two main facts: first, that allowing for multiple level shifts does not result in deteriorated size properties of the new tests; second, that the power properties of the new tests are close to those of the usual ADF tests under standard conditions (i.e., without level shifts), at least as the sample size increases, under fixed and local alternatives. Additional aspects of interest are how the new tests behave when the magnitude of the level shifts is independent of T, and how the properties of the tests are affected by the choice of degrees of freedom ν in (8). In particular, we want to assess whether the bounds on ν in Theorems 2, 3 and 4 are strict. Finally, the section ends with a study of the single shift case, in which our tests are compared to some well-known unit root tests especially developed to deal with a single shift. We aim at showing that not using the prior information on the number of shifts (which in practice is available only by exception), does not necessarily entail power losses for our tests.

The employed DGPs are as follows. Data are generated for sample sizes of $T = 100, 200, 400$ observations according to model (1) with $k = 1$, $\gamma := \gamma_1 \in \{-0.5, 0, 0.5\}$, $\varphi = 0$, $Y_{-1} = 0$ and zero-mean unit-variance IID innovations, following either a Gaussian or a standardized $t(10)$ distribution. We consider the unit root case, which obtains by setting $\alpha = 1$ in (1), the sequence of local alternatives $\alpha = 1 - c/T$ with $c := 7$, and the fixed alternative $\alpha = 0.9$.

Three specifications of the level shift component are initially employed. First, the case of no level shifts ($\mu_t = 0$ for all t) is considered, with the resulting model denoted by S_0. Second, with S_4 we denote the case of four shifts occurring at fixed sample fractions t_i, $i = 1, ..., 4$, with $t_1 := \lfloor 0.2T \rfloor$, $t_2 := \lfloor 0.35T \rfloor$, $t_3 := \lfloor 0.6T \rfloor$ and $t_4 := \lfloor 0.8T \rfloor$, and with size magnitudes $\eta_1 = \eta_4 := 0.4$ and $\eta_2 = -\eta_3 := 0.35$; consequently, the level shift component is[14]

$$\mu_t \quad : \quad = T^{1/2}[0.4\mathbb{I}_{\{t\geq\lfloor 0.2T\rfloor\}} + 0.35\mathbb{I}_{\{t\geq\lfloor 0.35T\rfloor\}} \\ -0.35\mathbb{I}_{\{t\geq\lfloor 0.6T\rfloor\}} + 0.4\mathbb{I}_{\{t\geq\lfloor 0.8T\rfloor\}}] \; . \tag{14}$$

Third, we consider a case, S_r in the following, with a random number of level shifts $N_T \sim 2 + B(T, 2/T)$ (B denoting a Binomial distribution); i.e., at least two, and on average four, level shifts occur over the sample. The shift dates t_i, $i = 1, ..., N_T$, are generated as $t_i := \lfloor \tau_i T \rfloor$, where the relative locations τ_i are independent and uniformly distributed on $[0, 1]$; the (independent) shift sizes η_t are drawn from a uniform distribution on $[-4, -0.35] \cup [0.35, 4]$.

We consider tests based both on the basic de-jumping procedure ($ADF^{\delta}_{\widehat{\alpha}}$, ADF^{δ}_{t}), see sections 3, and on the 'finer' version of section 4 ($ADF^{\Psi}_{\widehat{\alpha}}$,$ADF^{\Psi}_{t}$). For the degrees of freedom parameter ν of (8), in the Gaussian case Theorems 2, 3 and 4 allow us to choose arbitrarily large finite values. We have checked that the simulation results for large finite ν are, first, virtually indistinguishable from those for $\nu = \infty$ (i.e., standard Gaussian densities instead of t densities in (8)), and second, yield best test performance. Thus, in what follows we set $\nu = \infty$ in the case of Gaussian innovations. In the case of standardized $t\,(10)$ innovations, to investigate the relevance of the upper bound on ν in Theorems 2, 3 and 4, we use both $\nu = 8$ (which satisfies the bound $\nu \leq r - 1$) and $\nu = \infty$ (which violates it). All tests are performed at the 5% (asymptotic) nominal level, with critical values taken

[14] For our selection of T, this model generates level shifts of size between 4 and 8 standard deviations of the errors. These shift magnitudes, although large, are not unrealistic. For instance, Vogelsang and Perron (1998, p.1090) report, for the long GNP series considered in Perron (1992), level shifts 'generally no larger (in absolute value) than 8 (...) relative to the standard deviation of the innovation errors'. Furthermore, in some recent papers (see Papell et al., 2000, *inter alia*), unit root tests under level shifts are applied to unemployment series where level shifts are as large as *10* times the standard deviation of the innovations. Using, as in Perron (1990), quarterly US unemployment data from 1948 to 1988 with a break at 1974:1, we have estimated the size of the shift as 6.9 times the residual standard deviation from a 9 lags ADF regression on the dejumped data (cf. Perron, 1990, Table 1). With $T = 163$ this corresponds to $\eta = 0.54$, which is actually larger than the values we have used in our Monte Carlo design. Similarly, using monthly unemployment over the same range ($T = 485$) we find that the size of the shift is 10.4 times the residual standard deviation from the second-stage ADF regression; this corresponds to $\eta = 0.47$.

from Fuller (1976, Tables 10.A.1 and 10.A.2). Computations are based on $10,000$ Monte Carlo replications and are carried out in Ox v. 3.40, Doornik (2001).

Size. In Table 1 we report the finite-sample size of the six considered tests for Gaussian innovations. Two results are worth to note. First, in the case of no level shifts (Model S_0) the size properties of the proposed tests are accurate, with sizes ranging from 4.8% to 5.5% for the ADF^{δ} tests (basic de-jumping) and from 4.9% to 6.0% for the ADF^{Ψ} tests (finer de-jumping). Hence, allowing for multiple level shifts does not result in spurious rejection of the unit root hypothesis.[15] Second, in the case of multiple level shifts (Modes S_4 and S_r) standard ADF tests tend to be slightly undersized, in particular for $\gamma = \pm 0.5$.[16] Conversely, despite their generality the ADF^{δ} and ADF^{Ψ} tests display good size properties, with ADF^{Ψ} being only slightly oversized for moderate sample sizes. Hence, also for models S_4 and S_r allowing for multiple level shifts does not lead to spurious rejection of the null hypothesis.

Power against local alternatives. In Table 2 the size-adjusted power of the six tests is investigated under the local alternative $\alpha = 1 - 7/T$, still for Gaussian innovations. In the case of no level shifts the new tests have roughly the same power properties as standard ADF tests. That is, allowing for level shifts when there are actually none, does not deteriorate the power properties of the tests. The picture changes when level shifts occur. For Model S_4, standard ADF tests have extremely low power: with $T = 100$, power is about 0 for $\gamma = -0.5$, about 0.2% for $\gamma = 0$ and about 10% for $\gamma = 0.5$ (in agreement with Theorem 1, the distribution of the ADF tests depends both on γ and on the level-shift process). Conversely, the tests based on de-jumping have good power properties, with power growing toward the asymptotic power envelope as the sample size increases. For

[15]This is an important conclusion given that we did not present analytical results for model S_0.

[16]This result agrees with the asymptotics of Section 2, where it is shown (i) that in the presence of level shifts standard unit root tests do not behave according to the standard asymptotic theory even under the null hypothesis, and (ii) that the ADF test statistics are no longer invariant to the short run parameters γ.

$T = 100$, power is about 10% for $\gamma = -0.5$, 20% for $\gamma = 0$ and 30% for $\gamma = 0.5$; for $T = 200$ power grows above 20% for $\gamma = -0.5$, 30% for $\gamma = 0$ and 40% for $\gamma = 0.5$. For $T = 400$ the rejection rate is about 38% for negative γ and 45% for $\gamma \in \{0, 0.5\}$. The unit root tests based on finer de-jumping (ADF^{Ψ}) seem to be preferable over the tests based on the basic de-jumping (ADF^{δ}) for small sample sizes, as expected. Notice also that the dependence of finite-sample power on γ diminishes as T increases, as predicted by the asymptotic results of sections 3 and 4. Results for the random shift model S_r do not qualitatively differ from those for the S_4 model; in general, the power loss experienced by standard ADF tests is less severe than for S_4, and the power of our tests is closer to the asymptotic power envelope. Overall, the power results are promising, as they show that we are able to distinguish between unit root processes and processes which are stationary apart from several level shifts even in samples of moderate dimension.

Power against fixed alternatives. Table 3 reports the (size-adjusted) power of both standard and our unit root tests against the fixed alternative $\alpha = 0.9$. In the case of no level shifts (model S_0), the power of our test is close to the power of a standard unit root test; hence, there is no particular disadvantage in using the unit root tests which are robust to level shifts even if no shift occur. For model S_4, standard ADF tests have (size-adjusted) power close to 0, even for large sample sizes. Conversely, ADF^{δ} and ADF^{Ψ} tests have power which significantly increases with the sample size. As in the case of local alternatives, large sample sizes are needed for the tests to achieve a power level close to that of ADF tests under no level shifts. Interestingly, the tests based on the joint estimation of the level shift indicators and the autoregressive parameters (ADF^{Ψ}) have higher power than the tests based on the basic de-jumping (ADF^{δ}) when γ is negative. Results for the random level shifts model S_r do not differ substantially, except that standard ADF test now have non-zero power. For fixed alternatives which are further away from the null hypothesis, these results remain unchanged except that the joint estimation of the level shift indicators and the autoregressive parameters now allows to achieve better power properties in samples of moderate size[17]. In summary, the power properties of the proposed tests appear to be good also against the fixed alternatives, with power converging toward one as the sample size grows, as predicted by Theorem 3.

Fixed shift sizes. Although our paper deals with large level shifts, modelled as proportional to $T^{1/2}$, here we briefly compare the properties of standard unit root tests and of the proposed tests when the size of the level shifts is kept fixed as T increases. To this aim, we consider the following modification of model S_4:

$$\begin{aligned}\mu_t \quad : \quad &= 4\mathbb{I}_{\{t\geq\lfloor 0.2T\rfloor\}} + 3.5\mathbb{I}_{\{t\geq\lfloor 0.35T\rfloor\}} \\ &-3.5\mathbb{I}_{\{t\geq\lfloor 0.6T\rfloor\}} + 4.0\mathbb{I}_{\{t\geq\lfloor 0.8T\rfloor\}}.\end{aligned} \tag{15}$$

For $T = 100$, it generates exactly the level shift component (14) analyzed so far, whereas for larger sample sizes, (15) generates shifts of smaller magnitude with respect to (14). Table 4 reports size, local power and non-local power for $T = 200, 400$; results for $T = 100$ are given in Tables 1–3, panel S_4. The following results are worth to note. First, although the impact of the level shifts decreases as the sample size increases, the power of standard ADF tests is still very low. Size distortions, however, are ameliorated. Second, the local power of the proposed ADF^{δ}, ADF^{Ψ} tests still tends to the power envelope in the case of no shifts, although more slowly than in the case of large level shifts, as expected. Third, the power of the proposed tests grows as the sample size increases.

Student-t innovations and the choice of ν. We now examine the properties of the tests when the innovations, instead of being Gaussian, follow

[17]It has been recently noticed that for some unit root tests, power can decrease as α gets away from 1; see Perron and Qu (2006) and references therein. In order to assess whether this drawback affects out tests as well, we have run some Monte Carlo simulations based on model S_4 for various values of α, γ and T. The results show that the power reversal problem does not seriously affect our tests. More specifically, although the size-adjusted power is not monotonic in α, power tends to stabilize as α declines, in particular for the test based on the finer de-jumping method of Section 4. Moreover, there is no evidence that, for values of α far away from the null, power tends to be very small. In general, the tests based on finer de-jumping have the best power properties (the full set of results is reported in Appendix S).

a standardized Student $t(10)$ distribution. The ADF^{δ} and ADF^{Ψ} statistics are computed using $\nu = 8$ (satisfying the assumption $\nu \leq r - 1$ of Theorems 3 and 4) and $\nu = \infty$ (violating this assumption). Only the size-adjusted power under Model S_4 is reported in Table 5.[18] First, it is seen that the power of standard tests is almost as poor as for the case of Gaussian innovations, ranging from 0 ($\gamma = -0.5$) to 11.5% ($\gamma = 0.5, T = 400$). Regarding the ADF^{δ} and ADF^{Ψ} tests, for $\nu = 8$ only $T > 200$ ensures a substantial power gain over standard tests. To the contrary, the tests employing $\nu = \infty$ have power properties close to those observed for Gaussian errors, with only a slightly stronger dependence on the short run coefficient γ. Given also that the tests exhibited good size properties, for sample sizes comparable to those considered in the Monte Carlo experiment, we suggest to use large values of ν, even for non-Gaussian innovations.[19]

Single level shift and comparison with existing tests. As anticipated at the beginning of the section, we aim at comparing the proposed tests against existing methods in the well studied case of a single level shift; see Perron (2005) and Vogelsang and Perron (1998) for a review. Specifically, we consider the following tests for a unit root under a level shift at an unknown date: (i) the test proposed by Perron and Vogelsang (1992), ADF_t^{PV} hereafter; (ii) its 'efficient' generalization based on GLS removal of the level shifts, as proposed by Perron and Rodriguez (2003), denoted as ADF_t^{PV}.[20] Notice that these tests are not similar in the presence of a level shift of large size[21], and hence, they might experience some size distortions. Although in our framework the shift date is unknown, we also compare our tests to tests employing a known date of shift occurrence. Namely, we consider Per-

[18] Extended tables covering both size and power under models S_0, S_4 and S_r are reported in Appendix S.

[19] This choice may not be optimal in the absence of level shifts; see the S_0-panel referred to in the previous footnote.

[20] We do not compare our tests with tests designed for the case of two level shifts (e.g., Lumsdaine and Papell, 1997; Clemente et al., 1998) since these tests do not outperform their single-break counterparts (i.e, the ADF_t^{PV} test) when there is only a level shift. We also avoid comparison with the tests by Ohara (1999) and Kapetanios (2005) since these are designed for the case of innovational outliers only.

[21] The reason is that, when the size of the level shifts is large, the level shift component is no longer 'slowly evolving' in the sense of Elliott et al. (1996), Condition B.

ron's (1990) unit root test, $ADF^{P}_{t.\tau}$ (τ denoting the date of the level shift expressed as percentage of the sample size) hereafter, and a 'known date' version of Perron and Rodriguez (2003) GLS-based test, $ADF^{PR}_{t.\tau}$ in what follows. Our aim is to evaluate to what extent allowing for multiple breaks at unknown dates results in a power loss with respect to tests especially designed for the case of a single level shift. We consider the case of a shift of size $\eta T^{1/2}$, $\eta = 0.4$, occurring at the middle of the sample, i.e., at time $[\tau T]$, $\tau = 0.5$; hence, the level shift component

$$\mu_t := T^{1/2}(0.4\mathbb{I}_{\{t \geq \lfloor 0.5T \rfloor\}}) \tag{16}$$

is of magnitude comparable to those considered earlier.

Results are reported in Table 6, where the first panel contains simulated test sizes. The proposed tests have good size properties, comparable to the size of the tests assuming a known shift date, namely $ADF^{P}_{t.\tau}$ and $ADF^{PR}_{t.\tau}$. Conversely, due to the magnitude of the shift, the ADF^{PV}_{t} and ADF^{PR}_{t} tests are slightly oversized for $\gamma = 0$ and $\gamma = -0.5$. Regarding the (size-adjusted) power against local alternatives (second panel in the table), we observe that in the presence of large shifts, the power of the proposed tests is above the power of the ADF^{PR}_{t} and the ADF^{PV}_{t} test (the former being more powerful than the latter, as expected; cf. Perron and Rodriguez, 2003), in particular for larger sample sizes ($T = 200, 400$). Compared to tests assuming a known shift date, although the proposed tests dominates Perron's (1990) test,[22] they are dominated by the known-date version of Perron and Rodriguez (2003) test, $ADF^{PR}_{t.\tau}$. Nevertheless, for $T > 100$ the power of our tests is not too far from the power of the $ADF^{PR}_{t.\tau}$ test. Overall, the tests based on finer de-jumping, $ADF^{\Psi}_{\hat{\alpha}}$ and ADF^{Ψ}_{t}, show better power than the test based on the basic de-jumping, $ADF^{\delta}_{\hat{\alpha}}$ and ADF^{δ}_{t}, in particular for moderate sample sizes and $\gamma < 0$. Moreover, our 'finer' method does pretty well compared to unit root tests assuming a level shift at an unknown date and, for samples of at least 200 observations, also with the tests assuming a known shift date. The results do not qualitatively differ when the fixed alternative $\alpha = 0.9$ (third panel in the table) is considered.

[22] This result is expected since Perron's (1990) test is not based on an efficient de-jumping procedure under local alternatives; cf. Perron and Rodriguez (2003).

We conclude with a discussion of the fixed alternative $\alpha = 0.9$ when the size of the single level shift is kept constant at 4 times the standard deviation of the innovations, corresponding to (16) with $T = 100$. Results for $T = 200, 400$ are reported in the lowest panel of Table 6. Under a fixed alternative and fixed level-shift size, as the sample size gets larger our test tend to be outperformed by some of the existing tests, in particular for $\gamma \leq 0$. For $\gamma \geq 0$, the proposed tests do better than the ADF_t^P test and, for $\gamma = 0.5$, they are comparable with the ADF_t^{PR} test which, however, does better for $\gamma \leq 0$. Notice that here the effect of the (fixed) level shift on the power of standard ADF tests becomes negligible as the sample size grows.

6 Further issues

In this section we briefly show how the proposed tests can be used in the case of a linear time trend and in the case of unknown autoregressive order.

6.1 Linear time trends

Although all the analytical results given so far are derived under the assumption of no deterministics in the DGP, in the presence of linear time trends the proposed tests can be successfully applied in conjunction with a proper detrending procedure.

The approach we suggest is to combine pseudo-GLS detrending (see Elliott et al., 1996) with de-jumping in the computation of the ADF^{δ} and ADF^{Ψ} statistics.[23] Specifically, in the presence of a linear trend, the basic de-jumping procedure can be combined with deterministic corrections in the following way:

1. δ_t is estimated using basic de-jumping (i.e., neglecting the presence of the linear trend);
2. level shifts are removed by computing $\tilde{X}_t^{\delta}$ as in section 3.1;

[23]Given a time series X_t, the pseudo-GLS detrended series at $\overline{\alpha} := 1 - \overline{c}/T$ ($\overline{c} \geq 0$) is defined as $\tilde{X}_t^{\overline{\alpha}} := X_t^{\overline{\alpha}} - \hat{\varphi}^{\overline{\alpha}\prime} Z_t^{\overline{\alpha}}$, where $(X_0^{\overline{\alpha}}, X_t^{\overline{\alpha}}) := (X_0, (1 - \overline{\alpha}L)\, X_t)$, $(Z_0^{\overline{\alpha}}, Z_t^{\overline{\alpha}}) := (Z_0, (1 - \overline{\alpha}L)\, Z_t)$ and $\hat{\varphi}^{\overline{\alpha}}$ minimizes $S(\hat{\varphi}^{\overline{\alpha}}) := \sum_t (X_t^{\overline{\alpha}} - \hat{\varphi}^{\overline{\alpha}\prime} Z_t^{\overline{\alpha}})^2$.

3. pseudo-GLS de-trending is applied to $\tilde{X}_t^\delta$;

4. the ADF^δ statistics are obtained from an ADF regression for the de-trended $\tilde{X}_t^\delta$.

As in section 4, via an iterative procedure it is also possible to obtain the estimates of the level shift indicators jointly with those of the autoregressive coefficients and the coefficients of the linear trend $Z_t = (1, t)'$. To this end, the vector ξ and the mapping Ψ_T of section 4 are augmented with components $\varphi = (\varphi_c, \varphi_\tau)'$ and Ψ_T^φ for the coefficients of the linear trend. In place of e_t^d, the estimates $e_t^{d,\tau} := e_t^d - \varphi_\tau$ of $\varepsilon_t + \delta_t\theta_t$ are used in the definition of Ψ_{Tt}^δ and Ψ_T^λ. The updated de-jumped series is now $X_t^\Psi := X_t - \sum_{s=1}^t \Psi_{Ts}^\delta(\Delta X_s - \varphi_\tau)$; it is GLS-detrended to get $X_t^{\Psi,\tau}$ and the updated estimate Ψ_T^φ. ADF regression is performed on $X_t^{\Psi,\tau}$ instead of X_t^Ψ, and e_t^Ψ are replaced by $e_t^{\Psi,\tau} := e_t^\Psi - \Psi_T^\varphi$. The iteration is initialized with the outcome of the basic de-jumping procedure, see $1 - 4$ above.

As is standard, we evaluate the properties of the tests using pseudo-GLS detrending at $\bar{\alpha} := 1 - \bar{c}/T$, with $\bar{c} = 13.5$. Results are reported in Table 7 (size) and Table 8 (local power). All conclusions obtained for the case of no deterministic terms carry over: first, in the absence of level shifts our tests behave as the standard ADF tests, while in the presence of shifts our tests do not experience the serious power loss of standard ADF tests. The only difference from the results in Tables 1–2 is that in the presence of shifts the (still severe) power loss of standard ADF tests is partially mitigated by detrending the data.[24] As to our tests, GLS detrending of the de-jumped series does not seem to affect their size and power properties.

6.2 Unknown AR order

So far, we have assumed that the autoregressive order k of the errors is known to the econometrician. In more realistic situations, k is unknown and needs to be estimated.

[24]Further simulations (not reported) have shown that OLS detrending could be used instead of pseudo-GLS detrending. Obviously, OLS detrending leads to different (asymptotic) power properties (cf. Elliott et al., 1996; Müller and Elliott, 2003).

To deal with unknown lag order we suggest the following strategy, which combines our test procedure with standard methods for the determination of the lag order. At the first round, the basic de-jumping algorithm of section 3 (which does not require to specify k) is applied, and the de-jumped time series, $\tilde{X}_t^\delta$, is obtained; see steps $1-2$ in section 3.1. At the second round, using $\tilde{X}_t^\delta$ a standard criterion for the determination of the lag order is employed; this delivers an estimate of k, say $\tilde{k}$. At the final round, the ADF^δ statistics are computed (see step 3 in section 3.1) fixing the number of lags at $\tilde{k}$.

Tests based on the finer de-jumping of section 4 can be implemented as well: once k has been estimated as described above, ADF^Ψ statistics can be constructed by fixing k at $\tilde{k}$.

In Table 9 we report the results from a small Monte Carlo experiment based on model S_4 (four level shifts at fixed locations). Two criteria for lag order determination are used: a sequential t test and the BIC criterion; see Ng and Perron (1995). The size of the tests is largely acceptable, with the tests based on the sequential t-tests being slightly more liberal than the tests based on the BIC. Also, ADF^Ψ tests are slightly oversized when compared with the basic tests ADF^δ. The (local) power of the tests is not too distant from the power obtained by fixing k at its true value. Again, the joint estimation procedure of section 4 leads to better power for negative values of γ.

7 Conclusions

We have proposed a modification of the well-known augmented Dickey-Fuller (ADF) tests which allows to test for unit roots in the presence of multiple level shifts. Differently from existing work, we do not restrict the number of level shifts – which occur at random dates and have random sizes – apart from requiring it to be bounded in probability. The proposed tests have a limiting null distribution for which critical values are well-known; moreover, they have the same asymptotic power functions as standard ADF tests under no level shifts. A Monte Carlo simulation has shown that the new tests behave well in finite samples, and that they can account for linear time trends and unknown lag orders as well.

The results of the paper can be extended in various directions, which we reserve for further research. For instance, the finite autoregression assumption may be replaced by a general linear process assumption, along the work of Chang and Park (2002). A further important extension is to multivariate cointegrated models.

A Mathematical appendix

The appendix contains proofs of the results given so far. The following notation is used: as previously, P denotes the sequence of probability measures induced by model (1), conditionally on $N_T \geq 1$, and under Assumptions $\mathcal{M}$ and $\mathcal{S}$. Under the hypotheses H_0 and H_c (null and local alternatives), the measure P specializes to P_c, while under H_f (fixed stable alternative), P specializes to P_f. For a sequence $\{y_t\}$ of random variables, $\max_{t\leq T}\{y_t\} := \max_{-k\leq t\leq T}\{y_t\}$, $\max_{t:\delta_t=1}\{y_t\} := \max_{1\leq t\leq T}\{\delta_t y_t\}$, and $\min_{t:\delta_t=1}\{y_t\} := (\max_{t:\delta_t=1}\{y_t^{-1}\})^{-1}$. An asterisk denotes true parameter values in cases of ambiguity; $\|.\|$ denotes the Euclidean norm.

A.1 Preliminary lemmas

Lemma A.1 *The following magnitude orders hold:*

a. $\max_{t\leq T}|Y_t| = O_{P_c}(T^{1/2})$*;*

b. If $\max_{t\leq T}|\varepsilon_t| = O_P(T^\tau)$, $\tau \geq 0$, *then* $\max_{t\leq T}|u_t| = O_P(T^\tau)$, $\max_{t\leq T}|Y_t| = O_{P_f}(T^\tau)$ *and* $\max_{t\leq T}|\Delta Y_t| = O_P(T^\tau)$*;*

c. If $E_P|\varepsilon_1|^{\nu+1} < \infty$, $\nu \geq 0$, *then* $\sum_{t=1}^T |\Delta Y_t|^{\nu+1} = O_P(T)$.

PROOF. As ε_t, Y_{-k-1} and $(u_{-2k}, ..., u_{-k-1})$ are assumed independent of δ_s for all t and s, conditioning on $N_T \geq 1$ in the definition of P is irrelevant for this proof. Under P_c we have $\max_{t\leq T}|T^{-1/2}Y_t| \xrightarrow{w} \sigma \max_{s\in[0,1]}|B_c(s)|$ (see (2) for the definition of B_c); hence (a). In (b) the evaluation of $\max_{t\leq T}|u_t|$ (and also of $\max_{t\leq T}|Y_t|$ and $\max_{t\leq T}|\Delta Y_t|$ under P_f) obtains from the moving average representation of u_t (resp. Y_t and ΔY_t) with exponentially decreasing coefficients. Under P_c, from $\Delta Y_t = (-c/T)Y_{t-1} + u_t$ and (a) it follows that $\max_{t\leq T}|\Delta Y_t|$ inherits the magnitude order of $\max_{t\leq T}|u_t|$.

As to part (c), since u_t is the stationary solution of $u_t = \sum_{i=1}^k \gamma_i u_{t-i} + \varepsilon_t$, $(u_t)^{\nu+1}$ is stationary and ergodic by Theorem 3.35 of White (2001). Then $E_P|\varepsilon_1|^{\nu+1} < \infty$ implies, by the Marcinkiewicz-Zygmund inequality, that $E_P|u_t|^{\nu+1} < \infty$. By an ergodic LLN (Theorem 3.34 of White, 2001), $\sum_{t=1}^T |u_t|^{\nu+1} = O_P(T)$. Under P_f a similar argument applies to $\sum_{t=1}^T |\Delta Y_t|^{\nu+1}$. Under P_c, the evaluation $\sum_{t=1}^T |\Delta Y_t|^{\nu+1} = O_{P_c}(T)$ follows from

$\sum_{t=1}^{T} |\Delta Y_t|^{\nu+1} \leq 2^{\nu}(\sum_{t=1}^{T} |u_t|^{\nu+1} + c^{\nu+1}T^{-(\nu+1)/2} \sum_{t=1}^{T} |T^{-1/2}Y_{t-1}|^{\nu+1})$

and $\sum_{t=1}^{T} |T^{-1/2}Y_{t-1}|^{\nu+1} \leq T(\max_{t\leq T} |T^{-1/2}Y_t|)^{\nu+1} = O_{P_c}(T)$; see (a). ■

Lemma A.2 *If $y_t = O_P(1)$ is a sequence of random variables independent of δ_s for all t, s under P, then $\sum_{t=1}^{T} \delta_t |y_t| = O_P(1)$ and $\max_{t:\delta_t=1} |y_t| = O_P(1)$.*

PROOF. Direct from the definitions of the measure P (by Assumption $\mathcal{S}(a)$) and of boundedness in probability. ■

A.2 Proof of Theorem 1 and related results

To avoid confusion with the conditional measure P, we use Q for the measure in Theorem 1.

A.2.1 Proof of Theorem 1 (Part I)

Here we derive the weak limit of the process $T^{-1/2}X_{\lfloor T\cdot\rfloor}$. The limiting distributions of the ADF statistics will be obtained at the end of the subsection.

Recall that $X_t = Y_t + \mu_t$, and define $\eta_t^* := \sum_{i=0}^{\infty} \psi_i \omega_{t-i}$. To obtain the weak limit of $T^{-1/2}\mu_{\lfloor T\cdot\rfloor}$, we start from the convergence of $\mu^*_{\lfloor T\cdot\rfloor} := \sum_{t=1}^{\lfloor T\cdot\rfloor} \delta_t \eta_t^*$, and then use the representation $T^{-1/2}\mu_{\lfloor T\cdot\rfloor} = \sum_{t=1}^{\lfloor T\cdot\rfloor} F(t/T)\Delta\mu_t^*$.

First, we show that the serial dependence of $\{\eta_t^*\}$ is lost in the limit of $\mu^*_{\lfloor T\cdot\rfloor}$. Consider, possibly upon enlarging the underlying probability space, a doubly infinite matrix $\{\omega_{it}\}_{i\in\mathbb{N},t\in\mathbb{Z}}$ of i.i.d. random variables, with $\omega_{1t} := \omega_t$, and which is independent of $\{\delta_t\}_{t=1}^{T}$ and $\{Y_t\}_{t=-k-1}^{T}$. Denote by r_t the time of the last shift occurrence before time t ($r_t := -\infty$ if there are no shifts before time t), and for $t = 1, ..., T$ define

$$\eta_t^+ := \delta_t\Big(\sum_{i=0}^{t-r_t-1} \psi_i \omega_{t-i} + \sum_{i=t-r_t}^{\infty} \psi_i \omega_{t,t-i}\Big) + (1-\delta_t)\sum_{i=0}^{\infty} \psi_i \omega_{t,t-i},$$

where $\sum_{i=\infty}^{\infty}(\cdot) := 0$. For a fixed T the vector $\{\eta_t^+\}_{t=1}^T$ is distributed as T independent draws from the distribution of η_1^* (no ω_{it} is present in two different η_t^+'s), and is independent of $\{\delta_t\}_{t=1}^T$ (direct by writing its distribution conditional on $\{\delta_t\}_{t=1}^T$ and observing that it does not depend on $\{\delta_t\}_{t=1}^T$). Thus, the weak limit of $\sum_{t=1}^{\lfloor T\cdot\rfloor}\delta_t\eta_t^+$ obtains from Theorem 3.1 of Leipus and Viano (2003). Since $E|\eta_t^+| = E|\eta_1^*| < \infty$ (as implied by $E|\omega_{t,i}| = E|\omega_1| < \infty$ and $\sum_{i=1}^{\infty}|\psi_i| < \infty$), we have that $\sum_{t=1}^{\lfloor T\cdot\rfloor}\delta_t\eta_t^+ \xrightarrow{w} \sum_{i=1}^{\mathcal{P}(\cdot)}\eta_i^+ =: \mathcal{C}(\cdot)$ in $\mathcal{D}[0,1]$, where $\mathcal{P}$ is the Poisson process defined in Theorem 1.

Next, we argue that also $\mu_{\lfloor T\cdot\rfloor}^*$ has $\mathcal{C}(\cdot)$ as its weak limit. For $s \in [0,1]$ we have that

$$\sum_{t=1}^{\lfloor Ts\rfloor}\delta_t\eta_t^* = \sum_{t=1}^{\lfloor Ts\rfloor}\delta_t\eta_t^+ + \vartheta_s, \qquad \text{where } \vartheta_s := \sum_{t=1}^{\lfloor Ts\rfloor}\delta_t\sum_{i=t-r_t}^{\infty}\psi_i(\omega_{t-i} - \omega_{t,t-i}).$$

By the independence of ω's and δ's, and with $m_T := \min_{1\leq t\leq T}\{t - r_t\}$ denoting the minimum distance between two consecutive shift dates (∞ if at most one shift occurs),

$$\begin{aligned}
E(\max_{s\in[0,1]}|\vartheta_s|\big|\,\{\delta_t\}_{t=1}^T) &\leq E(\sum_{t=1}^T\delta_t\sum_{i=t-r_t}^{\infty}|\psi_i|\left(|\omega_{t-i}| + |\omega_{t,t-i}|\right)\big|\,\{\delta_t\}_{t=1}^T) \\
&= 2(E\,|\omega_0|)(\sum_{t=1}^T\delta_t\sum_{i=t-r_t}^{\infty}|\psi_i|) \\
&\leq 2(E\,|\omega_0|)N_T(\sum_{i=m_T}^{\infty}|\psi_i|).
\end{aligned}$$

Since $\sum_{i=1}^{\infty}|\psi_i| < \infty$ by assumption and $m_T \xrightarrow{Q} \infty$, it follows that $\sum_{i=m_T}^{\infty}|\psi_i| \xrightarrow{Q} 0$, and further, that $E(\sum_{i=m_T}^{\infty}|\psi_i|)^2 \to 0$ by dominated convergence. Thus, by taking expectations in the above display,

$$E(\max_{s\in[0,1]}|\vartheta_s|) \leq 2(E\,|\omega_0|)[EN_T^2 E(\textstyle\sum_{i=m_T}^{\infty}|\psi_i|)^2]^{1/2} \to 0$$

since $EN_T^2 \to G\,(1) + G\,(1)^2$ by a direct calculation (G denoting the Poisson intensity), so that $\max_{s\in[0,1]}|\vartheta_s| = o_Q\,(1)$ by Markov's inequality.

The last relation is equivalent to $\max_{s\in[0,1]} |\mu^*_{\lfloor Ts \rfloor} - \sum_{t=1}^{\lfloor Ts \rfloor} \delta_t \eta_t^+| = o_Q(1)$, and hence, $\mu^*_{\lfloor T\cdot \rfloor}$ has the same weak limit as$\sum_{t=1}^{\lfloor T\cdot \rfloor} \delta_t \eta_t^+$: $\mu^*_{\lfloor T\cdot \rfloor} \xrightarrow{w} \mathcal{C}(\cdot)$ in $\mathcal{D}[0,1]$. Due to the equality $T^{-1/2}\mu_{\lfloor T\cdot \rfloor} = \sum_{t=1}^{\lfloor T\cdot \rfloor} F(t/T)\Delta\mu_t^*$, the convergence $T^{-1/2}\mu_{\lfloor T\cdot \rfloor} \xrightarrow{w} \int_0^{\cdot} F(s)d\mathcal{C}(s) = \mathcal{C}_F$ is expected. For brevity and consistency with further references, we note that it follows, e.g., from Theorem 2.7 of Kurtz and Protter (1991). On the other hand, using standard local-to-unity asymptotics, $T^{-1/2}Y_{\lfloor T\cdot \rfloor} \xrightarrow{w} \sigma B_c(\cdot)$ in $\mathcal{D}[0,1]$. Due to the stochastic independence of μ_t and Y_t, the joint convergence

$$T^{-1/2}(Y_{\lfloor T\cdot \rfloor}, \mu_{\lfloor T\cdot \rfloor}) \xrightarrow{w} (\sigma B_c(\cdot), \mathcal{C}_F(\cdot)) \tag{A.1}$$

in the product space $(\mathcal{D}[0,1])^{\times 2}$ obtains. Although this is not a topological vector space, the functional $(x,y) \to x+y$ is continuous on the support $\mathcal{C}[0,1]\times\mathcal{D}[0,1]$ of $(\sigma B_c(\cdot), \mathcal{C}_F(\cdot))$, and $T^{-1/2}X_{\lfloor T\cdot \rfloor} \xrightarrow{w} \sigma B_c(\cdot)+\mathcal{C}_F(\cdot) = \sigma\mathcal{H}_c(\cdot)$ by the continuous mapping theorem (CMT).

Note next that (1) implies the following representation of ΔX_t:

$$\Delta X_t = \sum_{i=1}^{k} \gamma_i \Delta X_{t-i} + \tilde{\varepsilon}_t,\ \tilde{\varepsilon}_t := \varepsilon_t + \Gamma(L)\Delta\mu_t - (c/T)\Gamma(L)Y_{t-1}$$

for $t = 1, ..., T$. Let $\nabla \mathbf{X}_{t-1} := (\Delta X_{t-1}, ..., \Delta X_{t-k})'$, and similarly for $\nabla \mathbf{Y}_{t-1}$ and $\nabla \boldsymbol{\mu}_{t-1}$. The numerator of the $ADF_{\widehat{\alpha}}$ statistic based on X_t can be expressed as $T(\widehat{\alpha}-1) = (A_T/T)(B_T/T^2)^{-1}$, where

$$\begin{aligned} A_T &:= \sum_{t=1}^{T} X_{t-1}\tilde{\varepsilon}_t \\ &\quad -(\sum_{t=1}^{T} X_{t-1}\nabla\mathbf{X}'_{t-1})(\sum_{t=1}^{T} \nabla\mathbf{X}_{t-1}\nabla\mathbf{X}'_{t-1})^{-1}(\sum_{t=1}^{T} \nabla\mathbf{X}_{t-1}\tilde{\varepsilon}_t), \\ B_T &:= \sum_{t=1}^{T} X_{t-1}^2 \\ &\quad -(\sum_{t=1}^{T} X_{t-1}\nabla\mathbf{X}'_{t-1})(\sum_{t=1}^{T} \nabla\mathbf{X}_{t-1}\nabla\mathbf{X}'_{t-1})^{-1}(\sum_{t=1}^{T} \nabla\mathbf{X}_{t-1}X_{t-1}), \end{aligned}$$

see Chang and Park (2002). We write the limits of these and of related quantities using stochastic integrals. The limits follow from Theorem 2.7

of Kurtz and Protter (1991), in view of the relation $N_T = O_Q(1)$ and of the convergence

$$T^{-1/2}((Y_{\lfloor T\cdot\rfloor}, \mu_{\lfloor T\cdot\rfloor}), (\textstyle\sum_{t=1}^{\lfloor T\cdot\rfloor} \varepsilon_t, \mu_{\lfloor T\cdot\rfloor}), \mu_{\lfloor T\cdot\rfloor})' \qquad \text{(A.2)}$$
$$\xrightarrow{w} ((\sigma B_c, \mathcal{C}_F), (\sigma_\varepsilon B, \mathcal{C}_F), \mathcal{C}_F)'$$

on the space $\mathcal{D}_5[0, 1]$ of 5×1-vector càdlàg functions, endowed with the Skorohod topology. Convergence (A.2) holds on the product space $(\mathcal{D}[0, 1])^{\times 5}$ similarly to (A.1), and can be extended to $\mathcal{D}_5[0, 1]$ by Proposition VI.2.2(b) of Jacod and Shiryaev (2003). Thus, Kurtz and Protter's theorem delivers that, jointly with (A.2),

$$T^{-1}\sum_{t=1}^{T}(Y_{t-1}, \mu_{t-1})'(\varepsilon_t, \Delta\mu_t) \xrightarrow{w} \int_0^1 (\sigma B_c\,(s)\,, \mathcal{C}_F\,(s))' d(\sigma_\varepsilon B\,(s)\,, \mathcal{C}_F\,(s)). \qquad \text{(A.3)}$$

Another useful limit is that of $\mathbf{M} := T^{-1}\sum_{t=1}^{T}(\Delta\mu_t, \nabla\boldsymbol{\mu}'_{t-1})'(\Delta\mu_t, \nabla\boldsymbol{\mu}'_{t-1})$ and its continuous transformations. As $\sum_{t=1}^{T}\delta_{t-i}\delta_{t-j} = o_Q\,(1)$ for $i \neq j$, $\delta_{T-i} = 0$ with probability approaching 1 for $i = 0, ..., k$, and $\max_{t:\delta_t=1}|\eta_t| = O_Q(1)$ by Lemma A.2, it holds that

$$\mathbf{M} = \mathbf{I}_{k+1}(T^{-1}\sum_{t=1}^{T}(\Delta\mu_t)^2) + o_Q\,(1) \xrightarrow{w} \mathbf{I}_{k+1}\,[\mathcal{C}_F]\,, \qquad \text{(A.4)}$$

the limit as $T^{-1}\sum_{t=1}^{T}(\Delta\mu_t)^2 = T^{-1}\mu_T^2 - 2(T^{-1}\sum_{t=1}^{T}\mu_{t-1}\Delta\mu_t) \xrightarrow{w} \mathcal{C}_F(1)^2 - 2\int_0^1 \mathcal{C}_F(s)d\mathcal{C}_F(s) = [\mathcal{C}_F]$ by (A.2), (A.3) and CMT. Convergence is joint with (A.2) and (A.3) again by CMT. ■

A.2.2 Proof of Theorem 1 (Part II)

To complete the proof of Theorem 1 we need to introduce the following lemma.

Lemma A.3 *Let $\Omega_{\nabla\nabla}$ denote the probability limit of $T^{-1}\sum_{t=1}^{T}\nabla\mathbf{Y}_{t-1}\nabla\mathbf{Y}'_{t-1}$, and let $\Omega_{1\nabla}$ be the constant matrix defined through $T^{-1}\sum_{t=1}^{T}Y_{t-1}\nabla\mathbf{Y}'_{t-1} \xrightarrow{w}$*

$\sigma^2 \mathbf{1}_k' \int_0^1 B_c(s)\, dB_c(s) + \Omega_{1\nabla}$ *(see, e.g., Hansen, 1992; Phillips, 1987, Lemma 1), where* $\mathbf{1}_k$ *is a* $k \times 1$ *vector of ones. Then, as* $T \to \infty$*, the following converge jointly:*

(i) $T^{-2} \sum_{t=1}^T X_{t-1}^2 \xrightarrow{w} \sigma^2 \int_0^1 \mathcal{H}_c(s)^2 ds$;

(ii) $T^{-1} \sum_{t=1}^T X_{t-1} \tilde{\varepsilon}_t \xrightarrow{w} \Gamma(1)\, \sigma^2 \int_0^1 \mathcal{H}_c(s)\, d\mathcal{H}_c(s) + (\Gamma(1) - 1)\, [\mathcal{C}_F]$;

(iii) $T^{-1} \sum_{t=1}^T \nabla \mathbf{X}_{t-1} \nabla \mathbf{X}_{t-1}' \xrightarrow{w} \Omega_{\nabla\nabla} + [\mathcal{C}_F]\, \mathbf{I}_k =: \Omega_{\nabla\nabla}^{\mathcal{C}}$;

(iv) $T^{-1} \sum_{t=1}^T X_{t-1} \nabla \mathbf{X}_{t-1}' \xrightarrow{w} \mathbf{1}_k' \{\sigma^2 \int_0^1 \mathcal{H}_c(s)\, d\mathcal{H}_c(s) + [\mathcal{C}_F]\} + \Omega_{1\nabla}$;

(v) $T^{-1} \sum_{t=1}^T \nabla \mathbf{X}_{t-1} \tilde{\varepsilon}_t \xrightarrow{w} -\gamma\, [\mathcal{C}_F]$;

(vi) $T^{-1} \sum_{t=1}^T \tilde{\varepsilon}_t^2 \xrightarrow{w} \sigma_\varepsilon^2 + [\mathcal{C}_F]\, (1 + \gamma'\gamma)$.

PROOF. Convergence (i) obtains from (6) and CMT. Further, the left side of (ii) equals $\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3$, with the following $\mathbf{A}_i$'s. First, $\mathbf{A}_1 := T^{-1} \sum_{t=1}^T X_{t-1} \varepsilon_t \xrightarrow{w} \sigma^2 \Gamma(1) \int_0^1 \mathcal{H}_c(s)\, dB(s)$ by (A.3) and CMT, since $\sigma = \sigma_\varepsilon / \Gamma(1)$. Second, $\mathbf{A}_2 := T^{-1} \sum_{t=1}^T X_{t-1} \Gamma(L) \Delta\mu_t$, which can be written, with $\gamma_0 := -1$, as

$$\begin{aligned} \mathbf{A}_2 &= -T^{-1} \sum_{i=0}^k \gamma_i \Big[\sum_{t=1}^T X_{t-i-1} \Delta\mu_{t-i} + \sum_{t=1}^T \big(\sum_{j=1}^i \Delta\mu_{t-j}\big) \Delta\mu_{t-i} \\ &\quad + \sum_{t=1}^T \big(\sum_{j=1}^i \Delta Y_{t-j}\big) \Delta\mu_{t-i} \Big]. \end{aligned}$$

The first summation in brackets contributes to $\mathbf{A}_2$ with

$$\Gamma(1) \big(T^{-1} \sum_{t=1}^T X_{t-1} \Delta\mu_t\big) + o_Q(1) \xrightarrow{w} \Gamma(1)\, \sigma \int_0^1 \mathcal{H}_c(s)\, d\mathcal{C}_F(s) \ .$$

The contribution of the second one is a continuous transformation of $\mathbf{M}$ in (A.4) and tends to $(\Gamma(1) - 1)\, [\mathcal{C}_F]$. Since $T^{-1} \sum_{t=1}^T \Delta Y_{t-j} \Delta\mu_{t-i} = o_Q(1)$ by Lemma A.2, we can conclude that $\mathbf{A}_2 \xrightarrow{w} \Gamma(1)\, \sigma \int_0^1 \mathcal{H}_c(s)\, d\mathcal{C}_F(s) + (\Gamma(1) - 1)\, [\mathcal{C}_F]$. Third,

$$\mathbf{A}_3 := -T^{-2} c \sum_{t=1}^T X_{t-1} \Gamma(L) Y_{t-1} \xrightarrow{w} -c\sigma^2 \Gamma(1) \int_0^1 \mathcal{H}_c(s)\, B_c(s)\, ds$$

by (A.2) and CMT. Combining the limits of the $\mathbf{A}_i$'s, and recalling that $B_c(s)$ satisfies $dB_c(s) = -cB_c(s)\,ds + dB(s)$, gives the limit asserted in (ii).

With $\mathbf{B} := T^{-1}\sum_{t=1}^{T}\nabla\mathbf{Y}_{t-1}\nabla\boldsymbol{\mu}'_{t-1}$, convergence (iii) follows from the identity

$$T^{-1}\sum_{t=1}^{T}\nabla\mathbf{X}_{t-1}\nabla\mathbf{X}'_{t-1} = T^{-1}\sum_{t=1}^{T}\nabla\mathbf{Y}_{t-1}\nabla\mathbf{Y}'_{t-1} + T^{-1}\sum_{t=1}^{T}\nabla\boldsymbol{\mu}_{t-1}\nabla\boldsymbol{\mu}'_{t-1} + \mathbf{B} + \mathbf{B}'.$$

Here $\mathbf{B} = o_Q(1)$ by Lemma A.2 applied element-wise, whereas the other two terms have limits $\Omega_{\nabla\nabla}$ and $[\mathcal{C}_F]\,\mathbf{I}_k$ respectively, the latter by (A.4). For (iv), note that

$$T^{-1}\sum_{t=1}^{T}X_{t-1}\nabla\mathbf{X}'_{t-1} = T^{-1}\sum_{t=1}^{T}Y_{t-1}\nabla\mathbf{Y}'_{t-1} + T^{-1}\sum_{t=1}^{T}\mu_{t-1}\nabla\mathbf{Y}'_{t-1} + T^{-1}\sum_{t=1}^{T}X_{t-1}\nabla\boldsymbol{\mu}'_{t-1},$$

where the limit of the first term on the right side is given in the hypothesis, the second term equals $T^{-1}\sum_{t=1}^{T}(\mu_{t-2}\Delta Y_{t-1}, ..., \mu_{t-k-1}\Delta Y_{t-k}) + o_Q(1)$ by Lemma A.2, and converges weakly to $\sigma\mathbf{1}'_k\int_0^1\mathcal{C}_F(s)\,dB_c(s)$ by partial summation and (A.3), while, again by Lemma A.2, the third term equals

$$T^{-1}\sum_{t=1}^{T}(X_{t-2}\Delta\mu_{t-1}, ..., X_{t-k-1}\Delta\mu_{t-k}) + T^{-1}\sum_{t=1}^{T}((\Delta\mu_{t-1})^2, ..., (\sum_{i=1}^{k}\Delta\mu_{t-i})\Delta\mu_{t-k}) + o_Q(1),$$

and can be seen to tend to $\mathbf{1}'_k\{\sigma\int_0^1\mathcal{H}_c(s)\,d\mathcal{C}_F(s) + [\mathcal{C}_F]\}$ by (A.3), (A.4) and CMT.

For items (v) and (vi) we have

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\nabla\mathbf{X}_{t-1}\tilde{\varepsilon}_t &= T^{-1}\sum_{t=1}^{T}\nabla\boldsymbol{\mu}_{t-1}\Gamma(L)\Delta\mu_t+\mathbf{C}+\mathbf{D}\\
&\quad+\mathbf{E}+\mathbf{F}\overset{w}{\to}-\gamma\left[\mathcal{C}_F\right],\\
T^{-1}\sum_{t=1}^{T}\tilde{\varepsilon}_t^2 &= T^{-1}\sum_{t=1}^{T}\varepsilon_t^2+T^{-1}\sum_{t=1}^{T}\left(\Gamma(L)\Delta\mu_t\right)^2\\
&\quad+\mathbf{G}+\mathbf{H}\overset{w}{\to}\sigma_\varepsilon^2+(1+\gamma'\gamma)\left[\mathcal{C}_F\right],
\end{aligned}
$$

since two kinds of $o_Q(1)$ terms appear in these expressions. First,

$$
\begin{aligned}
\mathbf{C} &: =T^{-1}\sum_{t=1}^{T}\nabla\boldsymbol{\mu}_{t-1}\varepsilon_t\\
\mathbf{D} &: =T^{-1}\sum_{t=1}^{T}\nabla\mathbf{Y}_{t-1}\Gamma(L)\Delta\mu_t
\end{aligned}
$$

and

$$
\mathbf{G}:=2T^{-1}\sum_{t=1}^{T}\varepsilon_t\Gamma(L)\Delta\mu_t
$$

are $o_Q(1)$ by Lemma A.2. Second, $\mathbf{E}: =T^{-1}\sum_{t=1}^{T}\nabla\mathbf{Y}_{t-1}\varepsilon_t\overset{Q}{\to}0$ by an LLN for $T^{-1}\sum_{t=1}^{T}(u_{t-1},...,u_{t-k})'\varepsilon_t$ and by (A.3) for $cT^{-2}\sum_{t=1}^{T}\left(Y_{t-2},...,Y_{t-k-1}\right)'\varepsilon_t$, and similarly for $\mathbf{F}$ and $\mathbf{H}$ which contain overnormalized contributions of $cT^{-1}Y_{t-1}$. The terms with non-zero limits are $T^{-1}\sum_{t=1}^{T}\varepsilon_t^2\overset{Q}{\to}\sigma_\varepsilon^2$ by an LLN,

$$
T^{-1}\sum_{t=1}^{T}\nabla\boldsymbol{\mu}_{t-1}\Gamma(L)\Delta\mu_t=(0_{k\times1},\mathbf{I}_k)\mathbf{M}(1,-\gamma')'\overset{w}{\to}-\gamma\left[\mathcal{C}_F\right]
$$

by (A.4) and CMT, and

$$
T^{-1}\sum_{t=1}^{T}\left(\Gamma(L)\Delta\mu_t\right)^2=tr((1,-\gamma')'(1,-\gamma')\mathbf{M})\overset{w}{\to}\left[\mathcal{C}_F\right]\left(1+\gamma'\gamma\right)
$$

by (A.4) and CMT again.

Convergence is joint since, up to terms with constant probability limits, the left sides (i)-(vi) can be collected in a continuous transformation of the left sides of (A.2)-(A.3). ■

From Lemma A.3 it then follows that, with $\Gamma_\infty := \Gamma(1) + \mathbf{1}_k'(\Omega_{\nabla\nabla}^{\mathcal{C}})^{-1}\gamma\,[\mathcal{C}_F]$,

$$T^{-1}A_T \xrightarrow{w} \Gamma_\infty\sigma^2 \int_0^1 \mathcal{H}_c(s)\, d\mathcal{H}_c(s) + \left(\Gamma_\infty + \Omega_{1\nabla}(\Omega_{\nabla\nabla}^{\mathcal{C}})^{-1}\gamma - 1\right)[\mathcal{C}_F],$$
$$T^{-2}B_T \xrightarrow{w} \sigma^2 \int_0^1 \mathcal{H}_c(s)^2\, ds.$$

Since $T(\hat{\alpha}-1) = (A_T/T)(B_T/T^2)^{-1}$, it follows that $\hat{\alpha} \xrightarrow{Q} 1$, and hence,

$$\hat{\Gamma}(1) \;:\; = 1 - \mathbf{1}_k'\hat{\gamma} = \Gamma(1) - \mathbf{1}_k'(\sum_{t=1}^T \nabla\mathbf{X}_{t-1}\nabla\mathbf{X}_{t-1}')^{-1}(\sum_{t=1}^T \nabla\mathbf{X}_{t-1}\tilde{\varepsilon}_t) + o_Q(1).$$

From Lemma A.3 we find $\hat{\Gamma}(1) \xrightarrow{w} \Gamma_\infty$. Taken together with $ADF_{\widehat{\alpha}} = T(\hat{\alpha}-1)/\hat{\Gamma}(1)$, these results yield the limit of $ADF_{\widehat{\alpha}}$ in (7) with $\varkappa_0 := (1 + (\Omega_{1\nabla}(\Omega_{\nabla\nabla}^{\mathcal{C}})^{-1}\gamma - 1)\Gamma_\infty^{-1})[\mathcal{C}_F]/\sigma^2$.

The proof for the t statistic is similar. Specifically, as $A_T^2/T^2 = O_Q(1)$, see above, the following equalities hold:

$$\begin{aligned} ADF_t &= A_T/T(B_TC_T/T^3 - A_T^2/T^3)^{-1/2} \\ &= \frac{A_T/T}{B_T/T^2}(B_T/T^2)^{1/2}(C_T/T)^{-1/2} + o_Q(1) \\ &= T(\widehat{\alpha}-1)(B_T/T)^{1/2}(C_T/T)^{-1/2} + o_Q(1) \qquad \text{(A.5)} \end{aligned}$$

with (Chang and Park, 2002)

$$C_T := \sum_{t=1}^T \tilde{\varepsilon}_t^2 - (\sum_{t=1}^T \tilde{\varepsilon}_t\nabla\mathbf{X}_{t-1}')(\sum_{t=1}^T \nabla\mathbf{X}_{t-1}\nabla\mathbf{X}_{t-1}')^{-1}(\sum_{t=1}^T \nabla\mathbf{X}_{t-1}\tilde{\varepsilon}_t). \qquad \text{(A.6)}$$

According to Lemma A.3,

$$T^{-1}C_T \xrightarrow{w} \sigma_\varepsilon^2 + [\mathcal{C}_F]\left(1 + \gamma'\gamma\right) - \gamma'(\Omega_{\nabla\nabla}^{\mathcal{C}})^{-1}\gamma\,[\mathcal{C}_F]^2 =: C_\infty\,.$$

Hence, (A.5) and (A.6) ensure that

$$ADF_t \overset{w}{\to} \frac{\Gamma_\infty \sigma^2 \int_0^1 \mathcal{H}_c(s)\, d\mathcal{H}_c(s) + (\Gamma_\infty + \Omega_{1\nabla}(\Omega^{\mathcal{C}}_{\nabla\nabla})^{-1}\gamma - 1)\,[\mathcal{C}_F]}{\sigma C_\infty^{1/2}(\int_0^1 \mathcal{H}_c(s)^2\, ds)^{1/2}},$$

which is the limit in (7) with $\varkappa_1 := C_\infty \Gamma_\infty^{-2}\sigma^{-2}$. In the special case $k = 0$, we have $\Gamma_\infty = 1$, $C_\infty = \sigma_\varepsilon^2 + [\mathcal{C}_F]$ and $\sigma^2 = \sigma_\varepsilon^2$, from where the simpler expressions for $\varkappa_0$ and $\varkappa_1$. ■

A.3 Proof of Theorem 3 and related results

A.3.1 Uniform evaluations related to de-jumping

Lemma A.4 *Suppose that $\{\hat{e}_t\}$ is a sequence such that $\max_{1\le t\le T}|\hat{e}_t - \varepsilon_t - \delta_t\theta_t| = o_P(T^{1/2})$. Then $\max_{t:\delta_t=1}\hat{e}_t^2 = O_P(T)$ and $T^{-1}\min_{t:\delta_t=1}\hat{e}_t^2$ is bounded away from zero in P-probability.*

PROOF. Recall that $\theta_t = T^{1/2}\eta_t$. The evaluation of $\max_{t:\delta_t=1}\hat{e}_t^2$ follows from the inequality

$$\max_{t:\delta_t=1}\hat{e}_t^2 \le 3(T\max_{t:\delta_t=1}\eta_t^2 + \max_{t:\delta_t=1}\varepsilon_t^2 + \max_{1\le t\le T}(\hat{e}_t - \varepsilon_t - T^{1/2}\delta_t\eta_t)^2)$$

by applying Lemma A.2 to the first two maxima on the right side, and the hypothesis of the lemma to the last one. The fact that $T^{-1}\min_{t:\delta_t=1}\hat{e}_t^2$ is bounded away from zero in P-probability obtains from $T^{-1}\min_{t:\delta_t=1}\hat{e}_t^2 \ge (\min_{t:\delta_t=1}\eta_t^2)\min_{t:\delta_t=1}\left|1 - |\eta_t^{-1}(T^{-1/2}\hat{e}_t - \eta_t)|\right|^2$. The first minimum on the right side equals $(\max_{t:\delta_t=1}\eta_t^{-2})^{-1} = (O_P(1))^{-1}$ by Lemma A.2 and the assumption that $\eta_t^{-1} = O_P(1)$, while the second one is $1 + o_P(1)$. Indeed, since

$$\begin{aligned}\max_{t:\delta_t=1}|\eta_t^{-1}(\hat{e}_t - T^{1/2}\eta_t)| &\le \max_{t:\delta_t=1}|\eta_t^{-1}|\big(\max_{1\le t\le T}|\hat{e}_t - \varepsilon_t - T^{1/2}\delta_t\eta_t| \\ &\quad + \max_{t:\delta_t=1}|\varepsilon_t|\big) = o_P(T^{1/2}),\end{aligned}$$

it follows that, with P-probability approaching one,

$$\min_{t:\delta_t=1} |1 - |\eta_t^{-1}(T^{-1/2}\hat{e}_t - \eta_t)|| = 1 - T^{-1/2} \max_{t:\delta_t=1} |\eta_t^{-1}(\hat{e}_t - T^{1/2}\eta_t)| = 1 + o_P(1). \quad \blacksquare$$

Remark A.1. In the discussion of joint de-jumping and parameter estimation, instead of a sequence of random variables $\{\hat{e}_t\}$, a sequence of random functions $\{\hat{e}_t(\cdot)\}$ of a vector variable with domain $\mathbb{Y}_T$ (say) will appear. If we suppose that $\sup_{y\in\mathbb{Y}_T;1\leq t\leq T} |\hat{e}_t(y) - \varepsilon_t - \delta_t\theta_t| = o_P(T^{1/2})$, the conclusions of Lemma A.4 hold for $\sup_{y\in\mathbb{Y}_T;t:\delta_t=1} \hat{e}_t^2(y)$ and $\inf_{y\in\mathbb{Y}_T;t:\delta_t=1} \hat{e}_t^2(y)$. The proof carries over with only notational changes. □

The next lemma provides conditions for uniformly consistent detection of level shifts.

Lemma A.5 *Let $\mathbb{A}_T$ and $\zeta(\cdot)$ be as in Lemma 1, $\{\hat{e}_t\}$ be as in Lemma A.4, and $\tilde{\delta}(\hat{e}_t;\zeta)$ be as in (10). Then $\sup_{\mathbb{A}_T} \sum_{t=1}^T \delta_t(1 - \tilde{\delta}(\hat{e}_t;\zeta)) = O_P(T^{-(\nu-2)/2})$ and $\sup_{\mathbb{A}_T} \sum_{t=1}^T |\delta_t - \tilde{\delta}(\hat{e}_t;\zeta)| = O_P(T^{-\min\{1/2,(\nu-2)/2\}})$, provided that $\sum_{t=1}^T (1-\delta_t)|\hat{e}_t|^{\nu+1} = O_P(T)$.*

PROOF. We write $\sup_{\mathbb{A}_T} f(\zeta)$ for $\sup_{z\in\mathbb{A}_T} f(\zeta(z))$, and similarly for suprema involving the components of $\zeta = (\zeta^\lambda, \zeta^\eta, \zeta^\sigma)'$. First, as $0 < \zeta^\lambda < T$ on $\mathbb{A}_T$ with P-probability tending to 1,

$$\begin{aligned} 1 - \tilde{\delta}(\hat{e}_t;\zeta) &= (1 - \frac{\zeta^\lambda}{T})\Big[1 - \frac{\zeta^\lambda}{T} + \frac{\zeta^\lambda}{T}\frac{\phi_\nu(\hat{e}_t;\zeta^\sigma + T\zeta^\eta)}{\phi_\nu(\hat{e}_t;\zeta^\sigma)}\Big]^{-1} \\ &< \frac{T}{\zeta^\lambda}\frac{\phi_\nu(\hat{e}_t;\zeta^\sigma)}{\phi_\nu(\hat{e}_t;\zeta^\sigma + T\zeta^\eta)} \\ &= \frac{T}{\zeta^\lambda}\Big(1 + T\frac{\zeta^\eta}{\zeta^\sigma}\Big)^{1/2}\Big(1 + \frac{\hat{e}_t^2}{\nu\zeta^\sigma}\Big)^{-(\nu+1)/2} \\ &\quad \times\Big(1 + \frac{\hat{e}_t^2}{\nu(\zeta^\sigma + T\zeta^\eta)}\Big)^{(\nu+1)/2}, \end{aligned}$$

where $\phi_\nu(e;a^2) = ga^{-1}(1 + e^2/(\nu a^2))^{-(\nu+1)/2}$ has been inserted, g being a normalization constant independent of a. As $\sup_{\mathbb{A}_T}\{(\zeta^\lambda)^{-1}(1 +$

$T\zeta^\eta/\zeta^\sigma)^{1/2}\} = O_P(T^{1/2})$, we find that

$$\begin{aligned}\delta_t(1-\tilde{\delta}\,(\hat{e}_t;\zeta)) &\leq O_P(T^{3/2})\delta_t\Big(1+\frac{\min_{t:\delta_t=1}\hat{e}_t^2}{\nu\zeta^\sigma}\Big)^{-(\nu+1)/2}\\ &\quad\times\Big(1+\frac{\max_{t:\delta_t=1}\hat{e}_t^2}{\nu(\zeta^\sigma+T\zeta^\eta)}\Big)^{(\nu+1)/2}.\end{aligned}$$

Summing over t and accounting for the assumption that $N_T = O_P(1)$ yields the evaluation

$$\begin{aligned}\sum_{t=1}^{T}\delta_t(1-\tilde{\delta}\,(\hat{e}_t;\zeta)) &\leq O_P(T^{3/2})\Big(1+\frac{\min_{t:\delta_t=1}\hat{e}_t^2}{\nu\zeta^\sigma}\Big)^{-(\nu+1)/2}\\ &\quad\times\Big(1+\frac{\max_{t:\delta_t=1}\hat{e}_t^2}{\nu(\zeta^\sigma+T\zeta^\eta)}\Big)^{(\nu+1)/2}.\end{aligned}$$

Since $\sup_{\mathbb{A}_T}\zeta^\sigma = O_P(1)$ and, by Lemma A.4, $T^{-1}\min_{t:\delta_t=1}\hat{e}_t^2$ is bounded away from 0 in P-probability, we have $(1+(\nu\zeta^\sigma)^{-1}\min_{t:\delta_t=1}\hat{e}_t^2)^{-1} \leq T^{-1}(\nu\zeta^\sigma)(T^{-1}\min_{t:\delta_t=1}\hat{e}_t^2)^{-1} \leq O_P(T^{-1})$ uniformly on $\mathbb{A}_T$. Further, $\max_{t:\delta_t=1}\hat{e}_t^2 = O_P(T)$ again by Lemma A.4, and as $\inf_{\mathbb{A}_T}\zeta^\eta$ is bounded away from 0 in P-probability, it follows that $\nu^{-1}(\zeta^\sigma+T\zeta^\eta)^{-1}\max_{t:\delta_t=1}\hat{e}_t^2 \leq O_P(1)$ uniformly on $\mathbb{A}_T$. Combining the two conclusions yields

$$\sup_{\mathbb{A}_T}\sum_{t=1}^{T}\delta_t(1-\tilde{\delta}\,(\hat{e}_t;\zeta)) = O_P(T^{-(\nu-2)/2})\,.$$

The order of magnitude of $\sum_{t=1}^T(1-\delta_t)\,\tilde{\delta}\,(\hat{e}_t;\zeta)$ is addressed next. By evaluating from below the denominator of (10), it can be concluded that

$$\begin{aligned}\tilde{\delta}\,(\hat{e}_t;\zeta) &< \frac{T^{-1}\zeta^\lambda}{(1-T^{-1}\zeta^\lambda)}\frac{\phi_\nu(\hat{e}_t;\zeta^\sigma+T\zeta^\eta)}{\phi_\nu(\hat{e}_t;\zeta^\sigma)}\\ &= \frac{T^{-1}\zeta^\lambda}{(1-T^{-1}\zeta^\lambda)}\Big(1+T\frac{\zeta^\eta}{\zeta^\sigma}\Big)^{-1/2}\Big(1+\frac{\hat{e}_t^2}{\nu\zeta^\sigma}\Big)^{(\nu+1)/2}\\ &\quad\times\Big(1+\frac{\hat{e}_t^2}{\nu(\zeta^\sigma+T\zeta^\eta)}\Big)^{-(\nu+1)/2}.\end{aligned}$$

The contribution of the factors containing only ζ^λ, ζ^η and ζ^σ is $O_P(T^{-3/2})$ uniformly on $\mathbb{A}_T$, while the last factor does not exceed unity. Further, by applying to $x = (\nu\zeta^\sigma)^{-1/2}|\hat{e}_t|$ the inequality $(1+x^2)^{(\nu+1)/2} < 2^{(\nu+1)/2}(1+|x|^{\nu+1})$, it is seen that $(1+(\nu\zeta^\sigma)^{-1}\hat{e}_t^2)^{(\nu+1)/2} < 2^{(\nu+1)/2} + O_P(1)|\hat{e}_t|^{\nu+1}$ since $\sup_{\mathbb{A}_T}(\zeta^\sigma)^{-1} = O_P(1)$. Hence,

$$\begin{aligned} (1-\delta_t)\,\tilde{\delta}(\hat{e}_t;\zeta) &< T^{-3/2}(1-\delta_t)(O_P(1) + O_P(1)|\hat{e}_t|^{\nu+1}) \quad \text{(A.7)} \\ \sum_{t=1}^{T}(1-\delta_t)\,\tilde{\delta}(\hat{e}_t;\zeta) &< O_P(T^{-1/2}) + O_P(T^{-3/2})\sum_{t=1}^{T}(1-\delta_t)|\hat{e}_t|^{\nu+1}, \end{aligned}$$

where in the second line the inequality $\sum_{t=1}^{T}(1-\delta_t) \le T$ has been used. Under the hypothesis that $\sum_{t=1}^{T}(1-\delta_t)|\hat{e}_t|^{\nu+1} = O_P(T)$, we get

$$\sum_{t=1}^{T}(1-\delta_t)\,\tilde{\delta}(\hat{e}_t;\zeta) \le O_P(T^{-1/2})$$

uniformly on $\mathbb{A}_T$. This result and the first part of the lemma give

$$\begin{aligned} \sum_{t=1}^{T}|\delta_t - \tilde{\delta}(\hat{e}_t;\zeta)| &= \sum_{t=1}^{T}(1-\delta_t)\,\tilde{\delta}(\hat{e}_t;\zeta) + \sum_{t=1}^{T}\delta_t(1-\tilde{\delta}(\hat{e}_t;\zeta)) \\ &\le O_P(T^{-1/2}) + O_P(T^{-(\nu-2)/2}) \end{aligned}$$

on $\mathbb{A}_T$, again uniformly. ■

Remark A.2. If $\hat{e}_t(\cdot)$ are random functions satisfying the condition of Remark A.1, and if also $\sup_{y\in\mathbb{Y}_T}\sum_{t=1}^{T}(1-\delta_t)|\hat{e}_t(y)|^{\nu+1} = O_P(T)$, then the conclusions of Lemma A.5 hold for $\sup_{(y,z)\in\mathbb{Y}_T\times\mathbb{A}_T}\sum_{t=1}^{T}\delta_t(1-\tilde{\delta}(\hat{e}_t(y);\zeta(z)))$ and $\sup_{(y,z)\in\mathbb{Y}\times\mathbb{A}}\sum_{t=1}^{T}|\delta_t - \tilde{\delta}(\hat{e}_t(y);\zeta(z))|$. Again, the proof is only notationally more involved. □

Next, let $d = (d_1, ..., d_T)'$ be a (continuous) $\mathbb{R}^T$-valued random function with domain $\mathbb{G}_T$ (say). For instance, in the proof of Lemma 1, d will have components $d_t(\cdot) = \tilde{\delta}(\Delta X_t; \zeta(\cdot))$ $(t = 1, ..., T)$ and domain $\mathbb{G}_T = \mathbb{A}_T$. In the next lemma we consider as a function on $\mathbb{G}_T$ the de-jumped series $X_t^d := X_t - \sum_{s=1}^{t} d_s\Delta X_s$ $(t = 1, ..., T)$, $X_t^d := X_t$ $(t = -k, ..., 0)$ and, among other things, evaluate the distance between $\{Y_t\}$ and $\{X_t^d\}$ depending on how close d is to $\delta = (\delta_1, ..., \delta_T)'$. The notation $\sup_{\mathbb{G}_T} f := \sup_{w\in\mathbb{G}_T} f(w)$ is used.

Lemma A.6 *Let d be defined on a set $\mathbb{G}_T$ such that* $\sup_{\mathbb{G}_T} \sum_{t=1}^{T} |d_t - \delta_t| = O_P(T^{\beta-1/2})$ *for some* $\beta \in [0, 1/2)$*. If* $\max_{t\leq T} |\varepsilon_t| = O_P(T^\tau)$*, with* $\beta + \tau < 1/2$ *and* $\beta < \tau$*, then:*

a. $\sup_{\mathbb{G}_T; t\leq T} |X_t^d - Y_t| = O_P(T^\beta)$, $\sup_{\mathbb{G}_T;t\leq T} |X_t^d| = O_{P_c}(T^{1/2}) = O_{P_f}(T^\tau)$*;*

b. $\sup_{\mathbb{G}_T} \sum_{t=1}^{T} |\Delta X_t^d - \Delta Y_t| = O_P(T^\beta)$, $\sup_{\mathbb{G}_T} |\Delta X_t^d - \Delta Y_t| \leq a_T \delta_t + b_T$ *with* $a_T = O_P(T^\beta)$ *and* $b_T = O_P(T^{\beta+\tau-1/2})$*, and* $\sup_{\mathbb{G}_T;t\leq T} |\Delta X_t^d| = O_P(T^\tau)$*;*

c. $\sup_{\mathbb{G}_T} \|D_T \mathbf{X}_t^d - D_T \mathbf{Y}_t^d\| \leq \tilde{a}_T \boldsymbol{\delta}_t + \tilde{b}_T$ *with* $\boldsymbol{\delta}_t := \sum_{i=0}^{k-1} \delta_{t-i}$, $\tilde{a}_T = O_P(T^\beta)$ *and* $\tilde{b}_T = O_P(T^{\beta+\tau-1/2})$, $\sup_{\mathbb{G}_T;0\leq t\leq T} \|D_T \mathbf{X}_t^d\| = O_P(T^\tau)$ *and* $\sup_{\mathbb{G}_T;0\leq t\leq T} \|\mathbf{X}_t^d\| = O_{P_f}(T^\tau)$*.*

d. $\sup_{\mathbb{G}_T} \| \sum_{t=1}^{T-1} \mathbf{Y}_t (X_t^d - Y_t)\| = O_{P_f}(T^{1/2+\beta})$, $\sup_{\mathbb{G}_T} \| \sum_{t=1}^{T-1} \nabla \mathbf{X}_t^d (X_t^d - Y_t)\| = O_{P_f}(T^{\tau+\beta})$ *and, if* $\sup_{\mathbb{G}_T} \sum_{t=1}^{T} \delta_t |d_t - \delta_t| = O_{P_f}(T^{-1})$*, then* $\sup_{\mathbb{G}_T} \| \sum_{t=1}^{T-1} X_t^d (X_t^d - Y_t) - T \Upsilon_T\| = O_{P_f}(T^{1/2+\tau+\beta})$*, where* $\Upsilon_T := T^{-1} \sum_{t=1}^{T-1} (\sum_{s=1}^{t} \delta_s \Delta Y_s)^2$*.*

PROOF. For $t = 1, ..., T$ we have $X_t = Y_t + T^{1/2} \sum_{s=1}^{t} \delta_s \eta_s$, and for such t we find

$$X_t^d - Y_t = -\sum_{s=1}^{t} \delta_s \Delta Y_s + \sum_{s=1}^{t} (\delta_s - d_s)(\Delta Y_s + T^{1/2} \delta_s \eta_s), \quad \text{(A.8)}$$

$$\Delta X_t^d - \Delta Y_t = -\delta_t \Delta Y_t + (\delta_t - d_t)(\Delta Y_t + T^{1/2} \delta_t \eta_t), \quad \text{(A.9)}$$

while for $t = -k, ..., 0$, $X_t^d - Y_t = 0$. The first relation in (a) follows using (A.8):

$$\sup_{\mathbb{G}_T;t\leq T} |X_t^d - Y_t| \leq \sum_{s=1}^{T} \delta_s |\Delta Y_s| \quad \text{(A.10)}$$

$$+ \big[\max_{s\leq T} |\Delta Y_s| + T^{1/2} \sum_{s=1}^{T} \delta_s |\eta_s| \big] \sup_{\mathbb{G}_T} \sum_{s=1}^{T} |\delta_s - d_s| \quad \text{(A.11)}$$

$$\leq O_P(1) + [O_P(T^\tau) + O_P(T^{1/2})] O_P(T^{\beta-1/2})$$

$$= O_P(T^\beta) \quad \text{(A.12)}$$

with magnitude orders respectively from Lemmas A.2, A.1(b), A.2, and from the hypothesis. Together with $\max_{t\leq T} |Y_t| = O_{P_c}(T^{1/2})$ $(= O_{P_f}(T^\tau))$, this implies the other relations in (a).

The right side of (A.10) is a uniform upper bound for $\sum_{t=1}^{T} |\Delta X_t^d - \Delta Y_t|$, see (A.9); hence, the first relation in (b). Again from (A.9), $|\Delta X_t^d - \Delta Y_t|$ for $t = 1, ..., T$ does not exceed

$$\delta_t \big(\max_{s:\delta_s=1} |\Delta Y_s| + T^{1/2} \max_{s:\delta_s=1} |\eta_s| \sum_{s=1}^{T} |\delta_s - d_s| \big) + \max_{s\leq T} |\Delta Y_s| \sum_{s=1}^{T} |\delta_s - d_s| ,$$

from where, upon taking suprema, the definitions of a_T and b_T in (b) are obvious. Their magnitude orders follow as in (a). Further, $\sup_{\mathbb{G}_T;t\leq T} |\Delta X_t^d| \leq \sup_{\mathbb{G}_T;t\leq T} |\Delta X_t^d - \Delta Y_t| + \max_{t\leq T} |\Delta Y_t| = O_P(T^\tau)$ by the above evaluation, Lemma A.1(b), and the inequality $\beta < \tau$.

A vector version $\sup_{\mathbb{G}_T} \|\nabla \mathbf{X}_t^d - \nabla \mathbf{Y}_t\| \leq \boldsymbol{\delta}_t a_T + k b_T$ obtains readily. Together with (a), it gives for $\mathbf{Y}_t = (Y_t, (\nabla \mathbf{Y}_t)')'$ and $\mathbf{X}_t^d = (X_t^d, (\nabla \mathbf{X}_t^d)')'$ the first relation in (c). Since $\max_{0\leq t\leq T} \|D_T \mathbf{Y}_t\| \leq \max_{t\leq T} |T^{-1/2} Y_t| + k \max_{t\leq T} |\Delta Y_t| = O_P(T^\tau)$, it holds further that

$$\begin{aligned} \sup_{\mathbb{G}_T;0\leq t\leq T} \|D_T \mathbf{X}_t^d\| &\leq \max_{0\leq t\leq T} \|D_T \mathbf{Y}_t\| + \sup_{\mathbb{G}_T;0\leq t\leq T} \|D_T \mathbf{X}_t^d - D_T \mathbf{Y}_t\| \\ &\leq O_P(T^\tau) + \tilde{a}_T N_T + \tilde{b}_T, \end{aligned}$$

which is $O_P(T^\tau)$ for $\beta < \tau$. Similarly, $\max_{t\leq T} |Y_t| = O_{P_f}(T^\tau)$ leads to $\max_{0\leq t\leq T} \|\mathbf{Y}_t\| = O_{P_f}(T^\tau)$ and $\sup_{\mathbb{G}_T;0\leq t\leq T} \|\mathbf{X}_t^d\| = O_{P_f}(T^\tau)$.

For item (d), note that $\sum_{t=s}^{T-1} \mathbf{Y}_t = (\sum_{t=s}^{T-1} Y_t, Y_{T-1} - Y_{s-1}, ..., Y_{T-1-k} - Y_{s-1-k})'$, so that

$$\begin{aligned} \max_{1\leq s\leq T} \| \sum_{t=s}^{T-1} \mathbf{Y}_t\| &\leq | \sum_{t=1}^{T-1} Y_t| + \max_{s\leq T} | \sum_{t=1}^{s-1} Y_t| + 2k \max_{s\leq T} |Y_s| \\ &= O_{P_f}(T^{1/2}) \end{aligned} \quad \text{(A.13)}$$

by the weak convergence of $\max_{s\leq T} |T^{-1/2} \sum_{t=1}^{s-1} Y_t|$ under P_f, and by Lemma A.1(b). Using (A.8), we find the decomposition $\sum_{t=1}^{T-1} \mathbf{Y}_t (X_t^d - Y_t) = \mathbf{K}_1 + T^{1/2} \mathbf{K}_2 + \mathbf{K}_3$, where, first, $\mathbf{K}_1 := -\sum_{t=1}^{T-1} (\mathbf{Y}_t \sum_{s=1}^{t} \delta_s \Delta Y_s) = -\sum_{s=1}^{T-1} (\delta_s \Delta Y_s \sum_{t=s}^{T-1} \mathbf{Y}_t)$ satisfies

$$\|\mathbf{K}_1\| \leq N_T \max_{s:\delta_s=1} |\Delta Y_s| \max_{1\leq s\leq T} \| \sum_{t=s}^{T-1} \mathbf{Y}_t\| = O_{P_f}(T^{1/2})$$

by Assumption $\mathcal{S}$(a), Lemma A.2 and (A.13). Second, as $\sum_{s=1}^{T}|\delta_s - d_s| \leq O_P(T^{\beta-1/2})$ on $\mathbb{G}_T$, we see that $\mathbf{K}_2 := \sum_{t=1}^{T-1}(\mathbf{Y}_t \sum_{s=1}^{t}\delta_s(\delta_s - d_s)\eta_s) = \sum_{s=1}^{T-1}(\delta_s(\delta_s - d_s)\eta_s \sum_{t=s}^{T-1}\mathbf{Y}_t)$ satisfies

$$\begin{aligned}\sup_{\mathbb{G}_T}\|\mathbf{K}_2\| &\leq (\sup_{\mathbb{G}_T}\textstyle\sum_{s=1}^{T}|\delta_s - d_s|)\max\limits_{s:\delta_s=1}|\eta_s|\max\limits_{1\leq s\leq T}\|\sum_{t=s}^{T-1}\mathbf{Y}_t\| \\ &= O_{P_f}(T^{\beta}),\end{aligned}$$

and similarly, $\mathbf{K}_3 := \sum_{t=1}^{T-1}(\mathbf{Y}_t \sum_{s=1}^{t}(\delta_s - d_s)\Delta Y_s) = \sum_{s=1}^{T-1}((\delta_s - d_s)\Delta Y_s \sum_{t=s}^{T-1}\mathbf{Y}_t)$ satisfies

$$\begin{aligned}\sup_{\mathbb{G}_T}\|\mathbf{K}_3\| &\leq (\sup_{\mathbb{G}_T}\textstyle\sum_{s=1}^{T}|\delta_s - d_s|)\max\limits_{s\leq T}|\Delta Y_s|\max\limits_{1\leq s\leq T}\|\sum_{t=s}^{T-1}\mathbf{Y}_t\| \\ &= O_{P_f}(T^{\tau+\beta}),\end{aligned}$$

using Lemma A.1(b) and again (A.13). The first relation in (d) follows by combining the above magnitude orders. The second relation obtains upon partial summation; with $\vec{X}_t^d := (X_t^d, ..., X_{t-k+1}^d)'$, so that $\nabla\mathbf{X}_t^d = \vec{X}_t^d - \vec{X}_{t-1}^d$, we find

$$\begin{aligned}\|\textstyle\sum_{t=1}^{T-1}\nabla\mathbf{X}_t^d(X_t^d - Y_t)\| &\leq \|\vec{X}_{T-1}^d\||X_{T-1}^d - Y_{T-1}| \\ &\quad + \max_{t\leq T}\|\vec{X}_t^d\|\textstyle\sum_{t=1}^{T-1}|\Delta X_t^d - \Delta Y_t| \\ &\leq k\max_{t\leq T}|X_t^d|(|X_{T-1}^d - Y_{T-1}| \\ &\quad + \textstyle\sum_{t=1}^{T}|\Delta X_t^d - \Delta Y_t|) \\ &\leq O_{P_f}(T^{\tau+\beta})\end{aligned}$$

uniformly on $\mathbb{G}_T$ by (a) and (b). Finally, $\sup_{\mathbb{G}_T}|\sum_{t=1}^{T-1}X_t^d(X_t^d - Y_t) - \sum_{t=1}^{T-1}(X_t^d - Y_t)^2| = \sup_{\mathbb{G}_T}|\sum_{t=1}^{T-1}Y_t(X_t^d - Y_t)| = O_{P_f}(T^{1/2+\beta})$ as shown earlier, whereas, using (A.8), we can see that

$$\sup_{\mathbb{G}_T}|\textstyle\sum_{t=1}^{T-1}(X_t^d - Y_t)^2 - T\Upsilon_T| = O_{P_f}(T^{1/2+\tau+\beta}),$$

since $\Upsilon_T = T^{-1}\sum_{t=1}^{T-1}(\sum_{s=1}^{t}\delta_s\Delta Y_s)^2 \leq N_T \max_{s:\delta_s=1}(\Delta Y_s)^2 = O_{P_f}(1)$,

$$\begin{aligned}\sum_{t=1}^{T-1}(\sum_{s=1}^{t}(\delta_s - d_s)\Delta Y_s)^2 &\leq \max_{s\leq T}(\Delta Y_s)^2\sum_{t=1}^{T}(\sum_{s=1}^{t}|\delta_s - d_s|)^2 \\ &= O_{P_f}(T^{2\tau+2\beta})\end{aligned}$$

and

$$T\sum_{t=1}^{T-1}(\sum_{s=1}^{t}(\delta_s-d_s)\delta_s\eta_s)^2 \leq \max_{s:\delta_s=1}(\eta_s)^2 T\sum_{t=1}^{T}(\sum_{s=1}^{t}\delta_s|\delta_s-d_s|)^2 = O_{P_f}(1)$$

using Lemma A.2 and the assumed magnitude orders of $\sum_{s=1}^{T}|\delta_s - d_s|$ and $\sum_{s=1}^{T}\delta_s|\delta_s-d_s|$, while the magnitude orders of cross-products in the expansion of $(X_t^d - Y_t)^2$, summed over t, follow from the above three evaluations and the Cauchy-Schwartz inequality. ■

We conclude this sub-section with an important result on the properties of OLS estimators in the ADF regression based on X_t^d, with X_t^d as in Lemma A.6.

Lemma A.7 *Let $(\hat{\alpha}_d - 1, \hat{\gamma}_d')'$ and $(\hat{\alpha}-1,\hat{\gamma}')'$ be the OLS estimators from the regressions $X_t^d = (\alpha,\gamma')\mathbf{X}_{t-1}^d + error_t$ and $Y_t = (\alpha,\gamma')\mathbf{Y}_{t-1} + error_t$, and $\hat{\sigma}_{\varepsilon,d}^2$ and $\hat{\sigma}_{\varepsilon}^2$ be the corresponding residual variances. Under the assumptions of Lemma A.6 with $\beta = 0$, it holds that:*

a. $\sup_{\mathbb{G}_T}\|(T(\hat{\alpha}_d-\hat{\alpha}), T^{1/2}(\hat{\gamma}_d-\hat{\gamma})', \hat{\sigma}_{\varepsilon,d}^2 - \hat{\sigma}_{\varepsilon}^2)\| = O_{P_c}(T^{\tau-1/2})$.

Furthermore, if additionally $\sup_{\mathbb{G}_T}\sum_{s=1}^{T}\delta_s|d_s-\delta_s| = O_{P_f}(T^{-1})$, *then:*

b. $\sup_{\mathbb{G}_T}\|(\hat{\alpha}_d - \bar{\alpha}_*, (\hat{\gamma}_d - \bar{\gamma}_*)')' + (\bar{\alpha}_* - 1)(S_{11}^T\Upsilon_T^{-1} + \mathbf{i}\mathbf{i}')^{-1}\mathbf{i}\| = O_{P_f}(T^{\tau-1/2})$ *and* $\sup_{\mathbb{G}_T}\hat{\sigma}_{\varepsilon,d}^2 = O_{P_f}(1)$, *where $(\bar{\alpha}_*, \bar{\gamma}_*')$ under H_f is defined through the representation $Y_t = (\bar{\alpha}_*, \bar{\gamma}_*')\mathbf{Y}_{t-1} + \varepsilon_t$ implied by (1), $S_{11}^T := T^{-1}\sum_{t=1}^{T-1}\mathbf{Y}_t\mathbf{Y}_t'$ and $\mathbf{i} := (1, 0_{1\times k})'$.*

PROOF. For a joint preparation for (a) and (b), we extend the definition of $(\bar{\alpha}_*, \bar{\gamma}_*')$ to H_0 and H_c by $(\bar{\alpha}_*, \bar{\gamma}_*') := (1, \gamma_*')$, so that $\alpha_* = \bar{\alpha}_* - c_*/T$ under H_0, H_c and H_f, where c_* is the true localizing parameter under H_c, and is zero otherwise. With this notation,

$$\Delta Y_t = (\bar{\alpha}_* - 1, \bar{\gamma}_*')\mathbf{Y}_{t-1} + \tilde{e}_t$$

and

$$\Delta X_t^d = (\bar{\alpha}_* - 1, \bar{\gamma}_*')\mathbf{X}_{t-1}^d + \tilde{e}_t + \check{e}_t,$$

where $\tilde{e}_t := \varepsilon_t - (c_*/T)\Gamma_*(L)\,Y_{t-1}$ and $\check{e}_t := (\bar{\alpha}_* - 1)(Y_{t-1} - X^d_{t-1}) + \bar{\Gamma}_*(L)\,(\Delta X^d_t - \Delta Y_t)$. The representation of ΔX^d_t obtains by inserting $\Delta Y_t = \Delta X^d_t + (\Delta Y_t - \Delta X^d_t)$ and $\mathbf{Y}_{t-1} = \mathbf{X}^d_{t-1} + (\mathbf{Y}_{t-1} - \mathbf{X}^d_{t-1})$ into that of ΔY_t.

Define $\bar{D}_T := D_T$, $\bar{B}_T := \mathrm{diag}(T, T^{1/2})$ under H_0 and H_c, and $\bar{D}_T := I_{k+1}$, $\bar{B}_T = I_2$ under H_f. With $S^T_{11} := T^{-1}\sum_{t=0}^{T-1}\bar{D}_T\mathbf{Y}_t(\bar{D}_T\mathbf{Y}_t)'$, $S^T_{1e} := T^{-h}\sum_{t=1}^{T}(\bar{D}_T\mathbf{Y}_{t-1})\tilde{e}_t$, $R^T_{11} := T^{-1}\bar{D}_T[\sum_{t=1}^{T-1}\mathbf{X}^d_t(\mathbf{X}^d_t)' - \sum_{t=1}^{T-1}\mathbf{Y}_t(\mathbf{Y}_t)']\bar{D}_T$, $R^T_{1e} := T^{-h}\bar{D}_T\sum_{t=1}^{T}(\mathbf{X}^d_{t-1} - \mathbf{Y}_{t-1})\tilde{e}_t$ and $R^T_{1\nabla} := T^{-h}\sum_{t=1}^{T}(\bar{D}_T\mathbf{X}^d_{t-1})\check{e}_t$, where $h = 1/2$ under H_0 and H_c, and $h = 1$ under H_f, we have

$$\bar{B}_T(\hat{\alpha} - \bar{\alpha}_*, (\hat{\gamma} - \bar{\gamma}_*)')' = (S^T_{11})^{-1}S^T_{1e}, \tag{A.14}$$

$$\bar{B}_T(\hat{\alpha}_d - \bar{\alpha}_*, (\hat{\gamma}_d - \bar{\gamma}_*)')' = (S^T_{11} + R^T_{11})^{-1}(S^T_{1e} + R^T_{1e} + R^T_{1\nabla}). \tag{A.15}$$

We show that $r^T_{1j} := \sup_{\mathbb{G}_T}\|R^T_{1j}\| = O_{P_c}(T^{\tau-1/2})$, $j \in \{1, e, \nabla\}$, while $\sup_{\mathbb{G}_T}\|R^T_{11} - \mathbf{i}\mathbf{i}'\Upsilon_T\|$, $\sup_{\mathbb{G}_T}\|R^T_{1e}\|$ and $\sup_{\mathbb{G}_T}\|R^T_{1\nabla} + (\bar{\alpha}_* - 1)\mathbf{i}\Upsilon_T\|$ are $O_{P_f}(T^{\tau-1/2})$. Then we use the standard facts that, under P, S^T_{11} has a positive definite, and S^T_{1e} has a finite (zero, under P_f) weak limit.

First,

$$\|R^T_{11}\| \leq (\max_{1\leq t\leq T}\|D_T\mathbf{X}^d_t\| + \max_{1\leq t\leq T}\|D_T\mathbf{Y}_t\|)(T^{-1}\sum_{t=1}^{T-1}\|D_T\mathbf{X}^d_t - D_T\mathbf{Y}_t\|)$$

under H_0 and H_c, where the maxima are $O_{P_c}(T^{\tau})$, and the summation is bounded by $T^{-1/2}\sum_{t=1}^{T}|X^d_t - Y_t| + k\sum_{t=1}^{T}|\Delta X^d_t - \Delta Y_t| \leq O_{P_c}(T^{1/2})$, all uniformly on $\mathbb{G}_T$; see Lemmas A.1(b) and A.6 (with $\beta = 0$). Hence, $r^T_{11} = O_{P_c}(T^{\tau-1/2})$. Under H_f, $\|TR^T_{11} - \sum_{t=1}^{T-1}\mathbf{X}^d_t(\mathbf{X}^d_t - \mathbf{Y}_t)'\|$ equals

$$\begin{aligned}\|\sum_{t=1}^{T}(\mathbf{Y}_t - \mathbf{X}^d_t)\mathbf{Y}'_t\| &\leq \|\sum_{t=1}^{T}(Y_t - X^d_t)\mathbf{Y}'_t\| \\ &\quad +(\sum_{t=1}^{T}\|\nabla\mathbf{X}^d_t - \nabla\mathbf{Y}_t\|)\max_{1\leq t\leq T}\|\mathbf{Y}_t\| \\ &\leq O_{P_f}(T^{1/2})\end{aligned}$$

uniformly on $\mathbb{G}_T$ by Lemmas A.6(d,b) and A.1(b); $\| \sum_{t=1}^{T-1} \mathbf{X}_t^d (\mathbf{X}_t^d - \mathbf{Y}_t)' - \sum_{t=1}^{T-1} (X_t^d, 0_{1\times k})' (\mathbf{X}_t^d - \mathbf{Y}_t)' \|$ equals

$$\begin{aligned} \| \sum_{t=1}^{T-1} \nabla \mathbf{X}_t^d (\mathbf{X}_t^d - \mathbf{Y}_t)' \| &\leq \| \sum_{t=1}^{T-1} \nabla \mathbf{X}_t^d (X_t^d - Y_t)' \| \\ &+ \max_{1\leq t\leq T} \| \nabla \mathbf{X}_t^d \| \sum_{t=1}^{T} \| \nabla \mathbf{X}_t^d - \nabla \mathbf{Y}_t \| \\ &\leq O_{P_f}(T^\tau) \end{aligned}$$

uniformly on $\mathbb{G}_T$ by Lemma A.6(d,b), and $\| \sum_{t=1}^{T-1} (X_t^d, 0_{1\times k})' (\mathbf{X}_t^d - \mathbf{Y}_t)' - \mathbf{ii}' \sum_{t=1}^{T-1} X_t^d (X_t^d - Y_t) \|$ equals

$$\| \sum_{t=1}^{T-1} X_t^d (\nabla \mathbf{X}_t^d - \nabla \mathbf{Y}_t)' \| \leq \max_{t\leq T} |X_t^d| \sum_{t=1}^{T} \| \nabla \mathbf{X}_t^d - \nabla \mathbf{Y}_t \| \leq O_{P_f}(T^\tau)$$

by Lemma A.6(a,b). From the last three inequalities,

$$\| R_{11}^T - T^{-1} \mathbf{ii}' \sum_{t=1}^{T-1} X_t^d (X_t^d - Y_t) \| \leq O_{P_f}(T^{-1/2})$$

on $\mathbb{G}_T$, and from Lemma A.6(d), $\sup_{\mathbb{G}_T} \| R_{11}^T - \mathbf{ii}' \Upsilon_T \| = O_{P_f}(T^{\tau+\beta-1/2})$.

Second, $\| R_{1e}^T \| \leq T^{-1} | \sum_{t=1}^{T} (X_{t-1}^d - Y_{t-1}) \tilde{e}_t | + T^{-h} \| \sum_{t=1}^{T} (\nabla \mathbf{X}_{t-1}^d - \nabla \mathbf{Y}_{t-1}) \tilde{e}_t \|$, where

$$\begin{aligned} \textstyle\sum_{t=1}^{T} (X_{t-1}^d - Y_{t-1}) \tilde{e}_t &= (X_T^d - Y_T)(\textstyle\sum_{s=1}^{T} \tilde{e}_s) \\ &\quad - \textstyle\sum_{t=1}^{T} [(\Delta X_t^d - \Delta Y_t) \sum_{s=1}^{t} \tilde{e}_s] \end{aligned}$$

by partial summation. Therefore, $| \sum_{t=1}^{T} (X_{t-1}^d - Y_{t-1}) \tilde{e}_t |$ is bounded by

$$\big(\max_{t\leq T} |X_t^d - Y_t| + \textstyle\sum_{t=1}^{T} |\Delta X_t^d - \Delta Y_t| \big) \max_{t\leq T} | \textstyle\sum_{s=1}^{t} \tilde{e}_s | \leq O_P(T^{1/2})$$

on $\mathbb{G}_T$, since $\sup_{\mathbb{G}_T;t\leq T}|X_t^d - Y_t|$ and $\sup_{\mathbb{G}_T}\sum_{t=1}^T|\Delta X_t^d - \Delta Y_t|$ are $O_P(1)$ by Lemma A.6(a,b), and

$$\begin{aligned}\max_{t\leq T}|\textstyle\sum_{s=1}^t \tilde{e}_s| &\leq \max_{t\leq T}|\textstyle\sum_{s=1}^t \varepsilon_s| \\ &\quad +(1+k\|\gamma_*\|)(c_*/T)(\max_{t\leq T}|\textstyle\sum_{s=1}^t Y_s| + U) \\ &= O_P(T^{1/2}),\end{aligned}$$

with $U = O_P(1)$ denoting a contribution from initial values. Still by Lemma A.6(b), and since

$$\max_{1\leq t\leq T}|\tilde{e}_t| \leq \max_{t\leq T}|\varepsilon_t| + (1+k\|\gamma_*\|)(c_*/T)\max_{t\leq T}|Y_t| = O_P(T^\tau)$$

we find that

$$\begin{aligned}\|\textstyle\sum_{t=1}^T(\nabla\mathbf{X}_{t-1}^d - \nabla\mathbf{Y}_{t-1})\tilde{e}_t\| &\leq (\textstyle\sum_{t=1}^T\|\nabla\mathbf{X}_{t-1}^d - \nabla\mathbf{Y}_{t-1}\|)\max_{1\leq t\leq T}|\tilde{e}_t| \\ &\leq O_P(1)O_P(T^\tau)\end{aligned}$$

uniformly on $\mathbb{G}_T$. Combining the above evaluations gives $\sup_{\mathbb{G}_T}\|R_{1e}^T\| = O_P(T^{\tau-1/2})$.

An evaluation of $\sup_{\mathbb{G}_T}\|R_{1\nabla}^T\|$ can be obtained similarly:

$$\begin{aligned}\|T^h R_{1\nabla}^T - (\bar{\alpha}_* - 1)\sum_{t=1}^{T-1}\bar{D}_T\mathbf{X}_t^d(Y_t - X_t^d)\| &\leq \max_{0\leq t\leq T}\|\bar{D}_T\mathbf{X}_t^d\|(1+k\|\bar{\gamma}_*\|) \\ &\quad \times\sum_{t=1}^T|\Delta X_t^d - \Delta Y_t|,\end{aligned}$$

which is $O_P(T^\tau)$ by Lemma A.6. Under H_0 and H_c this shows that $\sup_{\mathbb{G}_T}\|R_{1\nabla}^T\| = O_{P_c}(T^{\tau-1/2})$. Under H_f, it gives together with Lemma A.6(d) that $\sup_{\mathbb{G}_T}\|R_{1\nabla}^T + (\bar{\alpha}_* - 1)\mathbf{i}\Upsilon_T\| = O_{P_f}(T^{\tau-1/2})$.

We proceed with the proof of (a). From (A.14), (A.15) and the identity $(S_{11}^T + R_{11}^T)^{-1} = (S_{11}^T)^{-1} - (S_{11}^T + R_{11}^T)^{-1}R_{11}^T(S_{11}^T)^{-1}$ it follows that $(T(\hat{\alpha}_d - \hat{\alpha}), T^{1/2}(\hat{\gamma}_d - \hat{\gamma})')'$ equals

$$(S_{11}^T)^{-1}(R_{1e}^T + R_{1\nabla}^T) - (S_{11}^T + R_{11}^T)^{-1}R_{11}^T(S_{11}^T)^{-1}(S_{1e}^T + R_{1e}^T + R_{1\nabla}^T).$$

To evaluate this matrix, define the event $\mathcal{F}_T := \{r_{11}^T \|(S_{11}^T)^{-1}\| < 1\}$. As $r_{11}^T = O_{P_c}(T^{\tau-1/2})$ and S_{11}^T converges weakly to an a.s. positive definite matrix under P_c, it holds that $P_c(\mathcal{F}_T) \to 1$, and it suffices to study outcomes in $\mathcal{F}_T$. For such outcomes $\|(S_{11}^T + R_{11}^T)^{-1}\| \leq \|(S_{11}^T)^{-1}\|(1 - \|R_{11}^T\|\|(S_{11}^T)^{-1}\|)^{-1}$,[25] so that $\sup_{\mathbb{G}_T} \|(T(\hat{\alpha}_d - \hat{\alpha}), T^{1/2}(\hat{\gamma}_d - \hat{\gamma})')\|$ is dominated by

$$\begin{aligned}
&\|(S_{11}^T)^{-1}\|(r_{1e}^T + r_{1\nabla}^T) + \|(S_{11}^T)^{-1}\|^2 r_{11}^T \frac{\|S_{1e}^T\| + r_{1e}^T + r_{1\nabla}^T}{1 - r_{11}^T\|(S_{11}^T)^{-1}\|} \\
&\quad = O_{P_c}(T^{\tau-1/2})
\end{aligned}$$

by the first part of the proof, and since $(S_{11}^T)^{-1}$ and S_{1e}^T are $O_{P_c}(1)$. As to residual variances,

$$\begin{aligned}
\hat{\sigma}_{\varepsilon,d}^2 - \hat{\sigma}_{\varepsilon}^2 \quad = \quad & T^{-1}\sum_{t=1}^T [(\Delta X_t^d)^2 - (\Delta Y_t)^2] \\
& -(T^{1/2}(\hat{\alpha} - 1), \hat{\gamma}')R_{11}^T(T^{1/2}(\hat{\alpha} - 1), \hat{\gamma}')' \\
& -(T^{1/2}(\hat{\alpha}_d - \hat{\alpha}), (\hat{\gamma}_d - \hat{\gamma})')(S_{11}^T + R_{11}^T) \\
& \times(T^{1/2}(\hat{\alpha}_d - \hat{\alpha}), (\hat{\gamma}_d - \hat{\gamma})')' \\
& -2(T^{1/2}(\hat{\alpha}_d - \hat{\alpha}), (\hat{\gamma}_d - \hat{\gamma})')(S_{11}^T + R_{11}^T)(T^{1/2}(\hat{\alpha} - 1), \hat{\gamma}')'.
\end{aligned}$$

Since

$$\begin{aligned}
\|\sum_{t=1}^T [(\Delta X_t^d)^2 - (\Delta Y_t)^2]\| \quad & \leq \quad (\max_{t\leq T} |\Delta X_t^d| + \max_{t\leq T} |\Delta Y_t|) \sum_{t=1}^T |\Delta X_t^d - \Delta Y_t| \\
& \leq \quad O_{P_c}(T^{\tau})
\end{aligned}$$

uniformly on $\mathbb{G}_T$, see Lemmas A.1 and A.6(b), and since $(T^{1/2}(\hat{\alpha}-1), \hat{\gamma}') = O_{P_c}(1)$ is a well-known fact, from the previous argument it follows that $\sup_{\mathbb{G}_T} |\hat{\sigma}_{\varepsilon,d}^2 - \hat{\sigma}_{\varepsilon}^2|$ is dominated by an $O_{P_c}(T^{\tau-1/2})$-sequence. This completes the proof of (a).

[25] It follows, e.g., from $\|(S_{11}^T + R_{11}^T)^{-1}\| = \|(S_{11}^T)^{-1} - (S_{11}^T + R_{11}^T)^{-1} R_{11}^T (S_{11}^T)^{-1}\| \leq \|(S_{11}^T)^{-1}\| + \|(S_{11}^T + R_{11}^T)^{-1}\|\|R_{11}^T\|\|(S_{11}^T)^{-1}\|$ by solving the inequality for $\|(S_{11}^T + R_{11}^T)^{-1}\|$.

Next, (A.15) and the identity $(S_{11}^T+R_{11}^T)^{-1} = [I_{k+1}-(S_{11}^T+R_{11}^T)^{-1}(R_{11}^T-\mathbf{ii}'\Upsilon_T)](S_{11}^T+\mathbf{ii}'\Upsilon_T)^{-1}$ yield $(\hat{\alpha}_d-\bar{\alpha}_*, (\hat{\gamma}_d-\bar{\gamma}_*)')' + (\bar{\alpha}_*-1)(S_{11}^T\Upsilon_T^{-1}+\mathbf{ii}')^{-1}\mathbf{i} = \mathbf{Q}_1+\mathbf{Q}_2$, where

$$\begin{aligned}\mathbf{Q}_1 &:= (S_{11}^T+\mathbf{ii}'\Upsilon_T)^{-1}(S_{1e}^T+R_{1e}^T+R_{1\nabla}^T+(\bar{\alpha}_*-1)\Upsilon_T\mathbf{i}),\\ \mathbf{Q}_2 &:= (S_{11}^T+R_{11}^T)^{-1}(R_{11}^T-\mathbf{ii}'\Upsilon_T)(S_{11}^T+\mathbf{ii}'\Upsilon_T)^{-1}(S_{1e}^T+R_{1e}^T+R_{1\nabla}^T).\end{aligned}$$

Using that $\|(S_{11}^T+\mathbf{ii}'\Upsilon_T)^{-1}\| \le \|(S_{11}^T)^{-1}\|$, we find that $\sup_{\mathbb{G}_T}\|\mathbf{Q}_1\|$ does not exceed

$$\begin{aligned}&\|(S_{11}^T)^{-1}\|(\|S_{1e}^T\|+\sup_{\mathbb{G}_T}\|R_{1e}^T\|+\sup_{\mathbb{G}_T}\|R_{1\nabla}^T+(\bar{\alpha}_*-1)\Upsilon_T\mathbf{i}\|)\\ &=O_{P_f}(T^{\tau-1/2}),\end{aligned}$$

as S_{11}^T converges to a positive definite matrix in P_f-probability, $S_{1e}^T = O_{P_f}(T^{-1/2})$ by a CLT, and the suprema were found to be $O_{P_f}(T^{\tau-1/2})$. Further, for outcomes in the set $\mathcal{G}_T := \{(\sup_{\mathbb{G}_T}\|R_{11}^T-\mathbf{ii}'\Upsilon_T\|)\|(S_{11}^T+\mathbf{ii}'\Upsilon_T)^{-1}\| < 1\}$, with $P_f(\mathcal{G}_T)\to 1$, we have

$$\begin{aligned}\sup_{\mathbb{G}_T}\|(S_{11}^T+R_{11}^T)^{-1}\| &\le \|(S_{11}^T)^{-1}\|(1-\sup_{\mathbb{G}_T}\|R_{11}^T-\mathbf{ii}'\Upsilon_T\|\|(S_{11}^T)^{-1}\|)^{-1}\\ &= O_{P_f}(1),\end{aligned}$$

so that $\sup_{\mathbb{G}_T}\|\mathbf{Q}_2\|$ is bounded by $O_{P_f}(1)$ times

$$\begin{aligned}&\sup_{\mathbb{G}_T}\|R_{11}^T-\mathbf{ii}'\Upsilon_T\|\|(S_{11}^T)^{-1}\|(\|S_{1e}^T\|+\sup_{\mathbb{G}_T}\|R_{1e}^T\|+\sup_{\mathbb{G}_T}\|R_{1\nabla}^T\|)\\ =\;&O_{P_f}(T^{\tau-1/2})\end{aligned}$$

similarly to $\sup_{\mathbb{G}_T}\|\mathbf{Q}_1\|$. This proves the first part of (b). Since $\hat{\sigma}_{\varepsilon,d}^2$ can be written as

$$\hat{\sigma}_{\varepsilon,d}^2 = T^{-1}\sum_{t=1}^T(\Delta X_t^d)^2 - (\hat{\alpha}_d-1,\hat{\gamma}_d')(S_{11}^T+R_{11}^T)(\hat{\alpha}_d-1,\hat{\gamma}_d')',$$

it is not hard to see that $\sup_{\mathbb{G}_T}|\hat{\sigma}_{\varepsilon,d}^2-(T^{-1}\sum_{t=1}^T(\Delta Y_t)^2-\Lambda_T)| = o_{P_f}(1)$, where, with $(\hat{\alpha}_d^\infty-1,(\hat{\gamma}_d^\infty)') := (\bar{\alpha}_*-1,\bar{\gamma}_*')-(\bar{\alpha}_*-1)\mathbf{i}'(S_{11}^T\Upsilon_T^{-1}+\mathbf{ii}')^{-1}$,

$$\begin{aligned}\Lambda_T &:= (\hat{\alpha}_d^\infty-1,(\hat{\gamma}_d^\infty)')(S_{11}^T+\mathbf{ii}'\Upsilon_T)(\hat{\alpha}_d^\infty-1,(\hat{\gamma}_d^\infty)')'\\ &= (\bar{\alpha}_*-1,\bar{\gamma}_*')S_{11}^T(\bar{\alpha}_*-1,\bar{\gamma}_*')'-(\bar{\alpha}_*-1)^2/(\Upsilon_T^{-1}+\mathbf{i}'(S_{11}^T)^{-1}\mathbf{i}).\end{aligned}$$

As $T^{-1}\sum_{t=1}^{T}(\Delta Y_t)^2 - (\bar{\alpha}_* - 1, \bar{\gamma}_*')S_{11}^T(\bar{\alpha}_* - 1, \bar{\gamma}_*')' \xrightarrow{P_f} \sigma_\varepsilon^2$ and $(\Upsilon_T^{-1} + \mathbf{i}'(S_{11}^T)^{-1}\mathbf{i})^{-1} \leq \Upsilon_T = O_{P_f}(1)$, we conslude that $\sup_{\mathbb{G}_T} \hat{\sigma}_{\varepsilon,d}^2 = O_{P_f}(1)$. ■

A.3.2 Proof of Lemma 1

We verify that, for $3 \leq \nu \leq r-1$, $\hat{e}_t = \Delta X_t$ satisfies the hypotheses of Lemmas A.4 and A.5. First, $\max_{t\leq T}|\varepsilon_t| = O_P(T^\tau)$ with $\tau = r^{-1} \leq 1/4$ in view of Assumption $\mathcal{M}$(b), and $\max_{t\leq T}|\Delta Y_t| = O_P(T^\tau)$ by Lemma A.1(b). Hence, $\max_{1\leq t\leq T}|\Delta X_t - \varepsilon_t - \delta_t\theta_t| = \max_{1\leq t\leq T}|\Delta Y_t - \varepsilon_t| = o_P(T^{1/2})$ as required in Lemma A.4 for $\hat{e}_t = \Delta X_t$ (from here and Lemma A.2, it is seen that $\max_{1\leq t\leq T}|\Delta X_t| = O_P(T^{1/2})$, which will be used below). Second, $(1-\delta_t)|\Delta X_t|^{\nu+1} = (1-\delta_t)|\Delta Y_t|^{\nu+1}$ and $\sum_{t=1}^T(1-\delta_t)|\Delta X_t|^{\nu+1} \leq \sum_{t=1}^T|\Delta Y_t|^{\nu+1} = O_P(T)$ by Lemma A.1(c), since $\nu+1 \leq r$, and hence, also Lemma A.5 is applicable to $\hat{e}_t = \Delta X_t$. Lemma A.5 yields item (a) of Lemma 1 (recall that $\tilde{\delta}_t(\zeta) := \tilde{\delta}(\Delta X_t;\zeta)$), from where the relation for Φ_T^λ in (b).

The relations for Φ_T^η and Φ_T^σ in (b) also obtain using (a):

$$
\begin{aligned}
&\|(\Phi_T^\eta, \Phi_T^\sigma) - T^{-1}\sum_{t=1}^T(\delta_t, 1-\delta_t)(\Delta X_t)^2\| \\
&\qquad = \sqrt{2}T^{-1}|\sum_{t=1}^T(\delta_t - \tilde{\delta}_t(\zeta))(\Delta X_t)^2| \\
&\qquad \leq \sqrt{2}T^{-1}\max_{1\leq t\leq T}(\Delta X_t)^2\sum_{t=1}^T|\delta_t - \tilde{\delta}_t(\zeta)| \\
&\qquad\qquad \leq O_P(T^{-1/2})
\end{aligned}
$$

uniformly on $\mathbb{A}_T$ since $\max_{1\leq t\leq T}|\Delta X_t| = O_P(T^{1/2})$. As $(1-\delta_t)(\Delta X_t)^2 = (1-\delta_t)(\Delta Y_t)^2$ and $\delta_t\Delta X_t = \delta_t(\Delta Y_t + T^{1/2}\eta_t)$, we can insert

$$
T^{-1}\sum_{t=1}^T(\delta_t, 1-\delta_t)(\Delta X_t)^2 = (H_T, T^{-1}\sum_{t=1}^T(\Delta Y_t)^2) + o_P(1)
$$

in the above display; hence, the relations.

Item (a) and Lemma A.7 with $\mathbb{G}_T = \mathbb{A}_T$ and $d_t(\cdot) = \tilde{\delta}(\Delta X_t;\zeta(\cdot))$, $t = 1,...,T$, imply (c).

Consider now the Jacobian $\partial\Phi_T(\varsigma)/\partial\varsigma = T^{-1}\sum_{t=1}^{T}\tilde{\delta}_t(\varsigma)(1-\tilde{\delta}_t(\varsigma))w_t v_t'$, where

$$v_t = \begin{pmatrix} \dfrac{1}{\varsigma^\lambda(1-\varsigma^\lambda/T)} \\ \dfrac{\partial\ln\phi_\nu(\Delta X_t;\varsigma^\sigma+T\varsigma^\eta)}{\partial\varsigma^\eta} \\ \dfrac{\partial\ln[\phi_\nu(\Delta X_t;\varsigma^\sigma+T\varsigma^\eta)/\phi_\nu(\Delta X_t;\varsigma^\sigma)]}{\partial\varsigma^\sigma} \end{pmatrix}$$

and $w_t := (T,(\Delta X_t)^2,-(\Delta X_t)^2)'$. It can be evaluated as follows:

$$\|\frac{\partial\Phi_T(\varsigma)}{\partial\varsigma}\| \leq \max_{1\leq t\leq T}\|v_t\| \max_{1\leq t\leq T}\|T^{-1}w_t\|\sum_{t=1}^{T}\tilde{\delta}_t(\varsigma)(1-\tilde{\delta}_t(\varsigma)). \quad \text{(A.16)}$$

We discuss the factors on the right side separately, starting from the components of v_t. First,

$$\frac{\partial\ln\phi_\nu(\Delta X_t;\varsigma^\sigma+T\varsigma^\eta)}{\partial(\varsigma^\eta,\varsigma^\sigma)} = -\frac{1}{2}\frac{1}{\varsigma^\sigma+T\varsigma^\eta}\left[1-(\nu+1)\frac{\hat{E}_t^2}{1+\hat{E}_t^2}\right]\binom{T}{1},$$

where $\hat{E}_t^2 = (\Delta X_t)^2[\nu(\varsigma^\sigma+T\varsigma^\eta)]^{-1}$. A similar expression (with a different $\hat{E}_t^2$) obtains for $\partial[\ln\phi_\nu(\Delta X_t;\varsigma^\sigma)]/\partial\varsigma^\sigma$. Using that $\hat{E}_t^2/(1+\hat{E}_t^2) < 1$, it follows that

$$\begin{aligned} \left|\frac{\partial\ln\phi_\nu(\Delta X_t;\varsigma^\sigma+T\varsigma^\eta)}{\partial\varsigma^\eta}\right| &< \frac{\nu+2}{2\varsigma^\eta}, \left|\frac{\partial\ln\phi_\nu(\Delta X_t;\varsigma^\sigma+T\varsigma^\eta)}{\partial\varsigma^\sigma}\right| < \frac{\nu+2}{2\varsigma^\sigma} \quad \text{(A.17)} \\ \left|\frac{\partial\ln\phi_\nu(\Delta X_t;\varsigma^\sigma)}{\partial\varsigma^\sigma}\right| &< \frac{\nu+2}{2\varsigma^\sigma}. \end{aligned}$$

Upon evaluation at $(\varsigma^\eta,\varsigma^\sigma) = (\zeta^\eta(z),\zeta^\sigma(z))$, all right sides in (A.17) are $O_P(1)$ uniformly on $\mathbb{A}_T$. Also $\sup_{\mathbb{A}_T}\{(\zeta^\lambda)^{-1}(1-T^{-1}\zeta^\lambda)^{-1}\} = O_P(1)$. Hence, $\sup_{z\in\mathbb{A}_T;1\leq t\leq T}\|v_t|_{\varsigma=\zeta(z)}\| = O_P(1)$. The other terms on the right side of (A.16) are $\max_{1\leq t\leq T}\|T^{-1}w_t\| \leq 1+2T^{-1}\max_{1\leq t\leq T}(\Delta X_t)^2 = O_P(1)$ (see paragraph one of the proof) and $\sum_{t=1}^{T}\tilde{\delta}_t(\varsigma)(1-\tilde{\delta}_t(\varsigma)) \leq \sum_{t=1}^{T}|\delta_t-\tilde{\delta}_t(\varsigma)|$, which upon evaluation at $\zeta(z)$ is $O_P(T^{-1/2})$ uniformly on $\mathbb{A}_T$ (see item (a)). Item (d) follows by inserting the obtained orders of magnitude into (A.16). ■

A.3.3 Proof of Theorem 2

EXISTENCE. To be able to use a standard fixed-point theorem, we set up an auxiliary fixed-point problem for a self-map of a compact. For the auxiliary map, defined by centering, normalizing and truncating Φ_T, existence of a random fixed point is classical, and this point turns to be a fixed point of Φ_T, with P-probability tending to one.

Let $\mathbb{A}_T^\lambda := [-1/2, 1/2]$, $\mathbb{A}_T^\eta := [1/2, 3/2]$, $\mathbb{A}_T^\sigma := [\sigma_Y^2/2, 2\sigma_Y^2]$ and $\mathbb{A}_T := \mathbb{A}_T^\lambda \times \mathbb{A}_T^\eta \times \mathbb{A}_T^\sigma$. With $\zeta : \mathbb{A}_T \to \mathbb{R}^3$ acting on $z = (z^\lambda, z^\eta, z^\sigma)'$ according to $\zeta(z) = (z^\lambda + N_T, H_T z^\eta, z^\sigma)'$, let $\Theta_T := \zeta^{-1} \circ \Phi_T \circ \zeta$. As $H_T \neq 0$ P-a.s., Θ_T is well-defined P-a.s. Further, let $\Theta_T^{tr} : \mathbb{A}_T \to \mathbb{A}_T$ be defined component-wise: for $v \in \{\lambda, \eta, \sigma\}$, $\Theta_T^{tr,v}(z) := \Theta_T^v(z)$ if $\Theta_T^v(z) \in \mathbb{A}_T^v$, $\Theta_T^{tr,v}(z) := \min \mathbb{A}_T^v$ if $\Theta_T^v(z) < \min \mathbb{A}_T^v$ and $\Theta_T^{tr,v}(z) := \max \mathbb{A}_T^v$ if $\Theta_T^v(z) > \max \mathbb{A}_T^v$. Since Θ_T^{tr} is continuous and $\mathbb{A}_T$ is a convex compact, Θ_T^{tr} admits a random fixed point $z_T \in \mathbb{A}_T$. This is guaranteed, e.g., by Theorem 10 in Bharucha-Reid (1976).

For outcomes such that $\Theta_T(z_T) \in \mathbb{A}_T$, we have $z_T = \Theta_T^{tr}(z_T) = \Theta_T(z_T)$, i.e., $\zeta_T = \zeta(z_T)$ is a fixed point of Φ_T. Since, in view of Lemma 1(b), $\Theta_T(z_T) = \zeta^{-1}((N_T, H_T, \sigma_Y^2)' + o_P(1)) = (0, 1, \sigma_Y^2)' + o_P(1)$, and since $(0, 1, \sigma_Y^2)'$ is an interior point of $\mathbb{A}_T$, it follows that $P(\Theta_T(z_T) \in \mathbb{A}_T) \to 1$. Hence, ζ_T is a random fixed point of Φ_T with P-probability approaching one, and by the choice of $\mathbb{A}_T$, ζ_T is bounded and bounded away from zero in P-probability.

UNIQUENESS. We use a contraction argument. Let ς_T have the properties of ζ_T from the existence part of the theorem. Fix $\epsilon > 0$. Since N_T, H_T (by Lemma A.4(a) for η_t and η_t^{-1}), ζ_T and ς_T are bounded and bounded away from 0 in P-probability, there exist constants $q_h > 2q_l > 0$ and a set $\mathcal{A}_1$ with $P(\mathcal{A}_1) > 1 - \epsilon/5$ such that, for outcomes in $\mathcal{A}_1$, $N_T \leq q_h$, $2q_l \leq H_T \leq q_h$, $q_l \mathbf{1}_3 \leq \zeta_T \leq q_h \mathbf{1}_3$ and $q_l \mathbf{1}_3 \leq \varsigma_T \leq q_h \mathbf{1}_3$, where the inequalities for ζ_T and ς_T are component-wise. Define the compact $\mathbb{K} := [\min\{q_l, 1/2\}, 2q_h] \times [q_l, 2q_h] \times [\min\{\sigma_Y^2/2, q_l\}, \max\{2\sigma_Y^2, q_h\}]$. Then, for outcomes in $\mathcal{A}_1$, first, $\zeta_T, \varsigma_T \in \mathbb{K}$, and second, the point $(N_T, H_T, \sigma_Y^2)'$ is interior for $\mathbb{K}$, and at distance from the boundary of $\mathbb{K}$ bounded from below by a positive constant. From the latter and Lemma 1(b), with $\mathbb{K}$ in place of $\mathbb{A}_T$ and with $\zeta(\cdot)$ equal to the identity function, it follows that on some

$\mathcal{A}_2 \subset \mathcal{A}_1$, with $P(\mathcal{A}_2) > 1 - \epsilon/4$, and for $T > T_2$ (say), $\Phi_T(\mathbb{K}) \subset \mathbb{K}$. By Lemma 1(d), there exist $T_3 > T_2$ and $\mathcal{A}_3 \subset \mathcal{A}_2$, with $P(\mathcal{A}_3) > 1 - \epsilon/2$, such that for $T > T_3$ and on $\mathcal{A}_3$, $\sup_{z\in\mathbb{K}} \|(\Phi_T)'_z\| < 1/2$. Since $(\Phi_T)'_z$ is of dimension 3×3, under the same conditions, $\sup_{z_{1,2,3}\in\mathbb{K}} \|(\Phi_T)'_z|_{z_{1,2,3}}\| < \sqrt{3}/2$, where the ith row of $(\Phi_T)'_z$ is evaluated at z_i ($i = 1, 2, 3$). By the mean-value theorem for Φ_T on $\mathbb{K}$, we see that for outcomes in $\mathcal{A}_3$ and for large T, Φ_T is a contraction on $\mathbb{K}$, and has a unique (per outcome) fixed point on $\mathbb{K}$ by Banach's fixed point theorem.

Consider finally T_4 and $\mathcal{A}_4$, with $P(\mathcal{A}_4) > 1 - \epsilon/2$, such that for $T > T_4$ and outcomes in $\mathcal{A}_4$, $\Phi_T(\zeta_T) = \zeta_T$ and $\Phi_T(\varsigma_T) = \varsigma_T$. This is possible by the choice of ζ_T and ς_T. Then, for outcomes in $\mathcal{A}_3 \cap \mathcal{A}_4$ and $T > \max(T_3, T_4)$, $\zeta_T \in \mathbb{K}$ and $\varsigma_T \in \mathbb{K}$ must be equal to the unique fixed point of Φ_T on $\mathbb{K}$. Since $P(\mathcal{A}_3 \cap \mathcal{A}_4) > 1 - \epsilon$ and ϵ is arbitrary, $P(\zeta_T = \varsigma_T) \to 1$.

COMPUTABILITY. A contraction argument is used again. Let $\tilde{\mathbb{K}} := \mathbb{K} \cup \{\zeta_{T0}\}$. As in the proof of uniqueness, on some event $\mathcal{B}_1 \subset \mathcal{A}_1$, with $P(\mathcal{B}_1) > 1 - \epsilon/2$, and for $T > T_1$ (say), Φ_T is a contraction of $\tilde{\mathbb{K}}$ onto $\mathbb{K} \subset \tilde{\mathbb{K}}$, and $\zeta_T \in \tilde{\mathbb{K}}$. Further, from the proof of existence, $\Phi_T(\zeta_T) = \zeta_T$ on some $\mathcal{B}_2$ with $P(\mathcal{B}_2) > 1 - \epsilon/2$, and for $T > T_5$ (say). By Banach's fixed point theorem, for $T > \max(T_1, T_5)$ and outcomes in $\mathcal{B}_1 \cap \mathcal{B}_2$, with $P(\mathcal{B}_1 \cap \mathcal{B}_2) > 1 - \epsilon$, the sequence of iterates of Φ_T converges to the unique (per outcome) fixed point ζ_T of Φ_T on $\tilde{\mathbb{K}}$.

CONSISTENCY. This part of Theorem 2 follows from Lemma 1(a,b) by evaluating the functions there at $z_T \in \mathbb{A}_T$, with z_T, $\zeta(\cdot)$ and $\mathbb{A}_T$ as in the proof of existence. ■

A.3.4 Proof of Theorem 3

Under H_0 and H_c, the result for $ADF^{\delta}_{\hat{\alpha}}$ is direct from (2) and Lemma 1(c), with the functions there evaluated at $z_T \in \mathbb{A}_T$ and with $\zeta(\cdot)$ as in the proof of existence in Theorem 2. For ADF^{δ}_{t}, in addition to these results, also the conclusion that $S^T_{11} + R^T_{11}(\zeta(z_T)) = S^T_{11} + o_{P_c}(1)$ is invoked to ensure asymptotic equivalence of the standard errors based on $\tilde{X}^{\delta}_t$ and on Y_t; see the proof and the notation of Lemma A.7.

Let the OLS estimators of α, γ and σ^2_{ε} from the ADF regression for $\tilde{X}^{\delta}_t$

be $\tilde{\alpha}_\delta$, $\tilde{\gamma}_\delta$ and $\tilde{\sigma}^2_{\varepsilon,\delta}$, and let $\vartheta_T := 1/(1+\Upsilon_T \mathbf{i}'(S_{11}^T)^{-1}\mathbf{i}) \in (0,1]$. After some algebra, it obtains from Lemma A.7(b) (upon evaluation at z_T with $\zeta(\cdot)$ as above) that, under the hypothesis H_f,

$$(\tilde{\alpha}_\delta, \tilde{\gamma}'_\delta)' = \vartheta_T(\bar{\alpha}_*, \bar{\gamma}'_*)' + (1-\vartheta_T)(1, S^\infty_{0\nabla}(S^\infty_{\nabla\nabla})^{-1})' + o_{P_f}(1),$$

where $S^\infty_{0\nabla}$ and $S^\infty_{\nabla\nabla}$ are the P_f-probability limits respectively of $T^{-1}\sum_{t=1}^T \Delta Y_t(\nabla \mathbf{Y}_{t-1})'$ and $T^{-1}\sum_{t=1}^T \nabla \mathbf{Y}_{t-1}(\nabla \mathbf{Y}_{t-1})'$. Since, under H_f, $\bar{\alpha}_* = (\alpha_* - 1)\Gamma_*(1) + 1 < 1$ and $\Upsilon_T \mathbf{i}'(S_{11}^T)^{-1}\mathbf{i} = O_{P_f}(1)$, we find that $P_f(\tilde{\alpha}_\delta < 1) \to 1$, and $T(\tilde{\alpha}_\delta - 1) \overset{P_f}{\to} -\infty$. Since $\hat{\sigma}^2_{\varepsilon,\delta} = O_{P_f}(1)$ and $(S_{11}^T + R_{11}^T(\zeta(z_T)))^{-1} = O_{P_f}(1)$ by Lemma A.7(b) and its proof, it follows that $ADF_t^\delta \overset{P_f}{\to} -\infty$. Further, note that $S^\infty_{0\nabla}(S^\infty_{\nabla\nabla})^{-1}$ is the P_f-probability limit of the OLS estimator of γ from the regression $\Delta Y_t = \gamma' \nabla \mathbf{Y}_{t-1}+\text{error}_t$. As ΔY_t is stationary, $1 - S^\infty_{0\nabla}(S^\infty_{\nabla\nabla})^{-1}\mathbf{1}_k > 0$. Since also $1 - \bar{\gamma}'_* \mathbf{1}_k = \Gamma_*(0) > 0$ as a consequence of Assumption $\mathcal{M}$, we find that $P_f(1 - \tilde{\gamma}'_\delta \mathbf{1}_k > 0) \to 1$. As $\tilde{\gamma}_\delta = O_P(1)$, it follows that $ADF^\delta_{\alpha} = T(\tilde{\alpha}_\delta - 1)/(1 - \tilde{\gamma}'_\delta \mathbf{1}_k) \overset{P_f}{\to} -\infty$. ■

A.4 Proof of Theorem 4 and related results

Let $\rho > 0$ be a real number that we will choose as small as convenient. The compact $\mathbb{A}_T$ from the proof of Theorem 2 is replaced here by $\mathbb{D}_T \subset \mathbb{R}^{T+k+4}$, defined as $\mathbb{D}_T := \mathbb{Y}_T \times \mathbb{A}_T$, with

$$\begin{array}{ll}
\mathbb{Y}_T := \mathbb{D}_T^\delta \times \mathbb{D}_T^a \times \mathbb{D}_T^\gamma, & \mathbb{A}_T := \mathbb{D}_T^\lambda \times \mathbb{D}_T^\eta \times \mathbb{D}_T^\sigma, \\
\mathbb{D}_T^\delta := \{x^\delta \in \mathbb{R}^T : \sum_{t=1}^T |x_t^\delta| \le T^{\rho-1/2}\}, & \mathbb{D}_T^a := \{x^a \in [-T^{\rho-1/2}, T^{\rho-1/2}]\}, \\
\mathbb{D}_T^\gamma := \{x^\gamma \in \mathbb{R}^k : \|x^\gamma - \gamma_*\| \le 1\}, & \mathbb{D}_T^\lambda := \{x^\lambda \in [-1/2, 1/2]\}, \\
\mathbb{D}_T^\eta := \{x^\eta \in [1/2, 3/2]\}, & \mathbb{D}_T^\sigma := \{x^\sigma \in [\sigma^2_{\varepsilon*}/2, 2\sigma^2_{\varepsilon*}]\}.
\end{array}$$

A.4.1 More lemmas

The point at which Ψ_T is evaluated is specified as $\xi(x)$,

$$x = (x^{\delta\prime}, x^a, x^{\gamma\prime}, x^\lambda, x^\eta, x^\sigma)' \in \mathbb{D}_T,$$

where $\xi : \mathbb{D}_T \to \mathbb{R}^{T+k+4}$ is the random function with components $\xi^\delta(x^\delta) = x^\delta + (\delta_1, ..., \delta_T)'$, $\xi^\lambda(x^\lambda) = x^\lambda + N_T$, $\xi^\eta(x^\eta) = H_T x^\eta$, and $\xi^a, \xi^\gamma, \xi^\sigma$ are the identity functions respectively on $\mathbb{D}_T^a$, $\mathbb{D}_T^\gamma$ and $\mathbb{D}_T^\sigma$. It will be shown that there exists a random sequence $x_T \in \mathbb{D}_T$ such that $P_c((\Psi_T \circ \xi)(x_T) = \xi(x_T)) \to 1$. The proof uses properties in the spirit of Lemma 1 that are worked out next. The argument of ξ and its components is subsumed; the notation $\sup_{\mathbb{D}_T} f(\xi)$ is employed for $\sup_{x \in \mathbb{D}_T} f(\xi(x))$, and similarly for the components.

Upon substitution of $d = (d_1, ..., d_T)'$ by ξ^δ in the definition of X_t^d ($t = 1, ..., T$), the de-jumped series $X_t^\xi := X_t - \sum_{s=1}^t \xi_s^\delta \Delta X_s$ obtains; we treat it, as well as the associated residual series $e_t^\xi := \Delta X_t - (\xi^a, (\xi^\gamma)')D_T \mathbf{X}_{t-1}^\xi$, as a function on $\mathbb{Y}_T$. Similarly, the updated de-jumped series $X_t^\Psi := X_t - \sum_{s=1}^t \Psi_{Ts}^\delta(\xi)\Delta X_s$ and residuals $e_t^\Psi := \Delta X_t - \Psi_T^{a,\gamma}(\xi) D_T \mathbf{X}_{t-1}^\Psi$ are functions on $\mathbb{D}_T$. Then $\Psi_{Tt}^\delta(\xi) = \tilde{\delta}(e_t^\xi; \xi^\lambda, \xi^\eta, \xi^\sigma)$.

Lemma A.8 *If* $3 \leq \nu \leq r - 1$, $2\rho(\nu + 1) < 1$ *and* $\max_{t \leq T} |\varepsilon_t| = O_{P_c}(T^\tau)$*, then:*

a. $\sup_{\mathbb{D}_T} \sum_{t=1}^T |\delta_t - \Psi_{Tt}^\delta(\xi)| = O_{P_c}(T^{-1/2})$, $\sup_{\mathbb{D}_T} \sum_{t=1}^T \delta_t(1 - \Psi_{Tt}^\delta(\xi)) = O_{P_c}(T^{-(\nu-2)/2})$*;*

b. $\sup_{\mathbb{D}_T} \|\Psi_T^{a,\gamma}(\xi) - (T^{1/2}(\hat{\alpha} - 1), \hat{\gamma}')'\| = O_{P_c}(T^{\tau-1})$*, with* $\hat{\alpha}, \hat{\gamma}$ *as in Lemma A.7;*

c. $\sup_{\mathbb{D}_T; t:\delta_t = 0} |e_t^\Psi| = O_{P_c}(T^\tau)$ *and* $\sup_{\mathbb{D}_T; 1 \leq t \leq T} |e_t^\Psi| = O_{P_c}(T^{1/2})$*;*

d. $\sup_{\mathbb{D}_T} \|(\Psi_T^\eta(\xi) - H_T, \Psi_T^\sigma(\xi) - \sigma_{\varepsilon *}^2)\| = O_{P_c}(T^{\tau - 1/2})$.

PROOF. Without loss of generality, we take $\tau \leq 1/4$, since $E|\varepsilon_1|^4 < \infty$ under P_c. Note also that $\rho < 1/8$ under the hypothesis of the lemma (this represents no restriction on the model).

For item (a), we check that Lemma A.5 (as stated in Remark A.2) is applicable to $\hat{e}_t(\cdot) = e_t^\xi$. First, we evaluate $e_t^\xi - \varepsilon_t - \delta_t\theta_t - (\gamma_* - \xi^\gamma)'\nabla\mathbf{Y}_{t-1}$ on $\mathbb{Y}_T$. Under H_0 and H_c, they equal

$$(-T^{-1}c_* - T^{-1/2}\xi^a)Y_{t-1} - (\xi^a, \xi^{\gamma\prime})D_T(\mathbf{X}_{t-1}^\xi - \mathbf{Y}_{t-1}),$$

where $\sup_{\mathbb{Y}_T} |\xi^a| = O(T^{\rho - 1/2})$, $\sup_{\mathbb{Y}_T} \|\xi^\gamma\| = O(1)$, $\max_{t \leq T} |Y_t| = O_{P_c}(T^{1/2})$, and Lemma A.6(c) with $d(\cdot) = \xi^\delta(\cdot), \mathbb{G}_T = \mathbb{Y}_T$ and $\beta = \rho$

is applicable to $\sup_{\mathbb{Y}_T} \|D_T(\mathbf{X}^{\xi}_{t-1} - \mathbf{Y}_{t-1})\|$. By collecting the magnitude orders,

$$\begin{aligned} \sup_{\mathbb{Y}_T} |e^{\xi}_t - \varepsilon_t - \delta_t\theta_t - (\gamma_* - \xi^{\gamma})'\nabla\mathbf{Y}_{t-1}| &= O_{P_c}(T^{\rho})\,\boldsymbol{\delta}_{t-1} \\ &\quad + O_{P_c}(T^{\rho+\tau-1/2}). \end{aligned} \tag{A.18}$$

Thus, $\sup_{\mathbb{Y}_T;1\le t\le T} |e^{\xi}_t - \varepsilon_t - \delta_t\theta_t| = o_{P_c}(T^{1/2})$ as required in Remark A.1, since $\sup_{\mathbb{Y}_T} \|\gamma_* - \xi^{\gamma}\| = 1$, $\max_{1\le t\le T} \|\nabla\mathbf{Y}_{t-1}\| = O_{P_c}(T^{\tau})$ by Lemma A.1(c), and $\rho + \tau < 1/2$.

Second, we verify that $\sup_{\mathbb{Y}_T} \sum_{t=1}^{T} (1 - \delta_t) |e^{\xi}_t|^{\nu+1} = O_{P_c}(T)$. Indeed, in consequence of (A.18) and the inequalities $\rho(\nu + 1) < 1/2$, $\rho + \tau < 1/2$, it holds that $(1 - \delta_t) |e^{\xi}_t|^{\nu+1} \le$

$$\begin{aligned} (1 - \delta_t)\, 4^{\nu}[|\varepsilon_t|^{\nu+1} + \|\gamma_* - \xi^{\gamma}\|^{\nu+1}\|\nabla\mathbf{Y}_{t-1}\|^{\nu+1} \\ + (\boldsymbol{\delta}_{t-1})^{\nu+1} o_{P_c}(T^{1/2}) + o_{P_c}(1)] \end{aligned} \tag{A.19}$$

uniformly on $\mathbb{Y}_T$. Further, for $\nu + 1 \le r$, we have $E|\varepsilon_t|^{\nu+1} < \infty$, so that $\sum_{t=1}^{T} |\varepsilon_t|^{\nu+1} = O_{P_c}(T)$ and $\sum_{t=1}^{T} \|\nabla\mathbf{Y}_{t-1}\|^{\nu+1} = O_{P_c}(T)$ by Lemma A.1(c), while $\sup_{\mathbb{Y}_T} \|\gamma_* - \xi^{\gamma}\| = 1$ and $\sum_{t=1}^{T}(\boldsymbol{\delta}_{t-1})^{\nu+1} \le (kN_T)^{\nu+1} = O_{P_c}(1)$. These imply the asserted uniform order of $\sum_{t=1}^{T} (1 - \delta_t) |e^{\xi}_t|^{\nu+1}$. Having checked the conditions of Remark A.2, we obtain item (a).

Item (b) follows from (a) and Lemma A.7 with $d(\cdot) = \Psi^{\delta}_T(\xi(\cdot))$, $\mathbb{G}_T = \mathbb{D}_T$. Further, (b) together with the standard properties $\hat{\alpha} - 1 = O_{P_c}(T^{-1})$ and $\gamma_* - \hat{\gamma} = O_{P_c}(T^{-1/2})$ imply that $\sup_{\mathbb{D}_T} |\Psi^{a}_T| = O_{P_c}(T^{-1/2})$, $\sup_{\mathbb{D}_T} \|\gamma_* - \Psi^{\gamma}_T\| = O_{P_c}(T^{-1/2})$ and $\sup_{\mathbb{D}_T} \|\Psi^{a,\gamma}_T\| = O_{P_c}(1)$. These are useful in a derivation similar to that of (A.18), but this time invoking for $\sup_{\mathbb{D}_T} \|D_T(\mathbf{X}^{\Psi}_{t-1} - \mathbf{Y}_{t-1})\|$ Lemma A.6 with $d(\cdot) = \Psi^{\delta}_T(\xi(\cdot))$, $\mathbb{G}_T = \mathbb{D}_T$ and $\beta = 0$, the latter by (a):

$$\begin{aligned} \sup_{\mathbb{D}_T} |e^{\Psi}_t - \varepsilon_t - T^{1/2}\delta_t\eta_t| &\le O_{P_c}(1)\,\boldsymbol{\delta}_{t-1} + O_{P_c}(T^{\tau-1/2}) \\ &\quad + \sup_{\mathbb{D}_T} \|\gamma_* - \Psi^{\gamma}_T\| \max_{1\le t\le T} \|\nabla\mathbf{Y}_{t-1}\| \\ &= O_{P_c}(1)\,\boldsymbol{\delta}_{t-1} + O_{P_c}(T^{\tau-1/2}), \end{aligned} \tag{A.20}$$

since $\max_{1\le t\le T} \|\nabla\mathbf{Y}_{t-1}\| = O_{P_c}(T^{\tau})$ by Lemma A.1(b). From here and Lemma A.2, (c) follows.

Finally, item (d) can be derived starting from

$$\sup_{\mathbb{D}_T} \|(\Psi_T^\eta(\xi), \Psi_T^\sigma(\xi)) - T^{-1} \sum_{t=1}^T (\delta_t, 1-\delta_t)(e_t^\Psi)^2\|$$
$$= \sqrt{2} T^{-1} \sup_{\mathbb{D}_T} |\sum_{t=1}^T (\Psi_{Tt}^\delta(\xi) - \delta_t)(e_t^\Psi)^2|$$
$$\leq \sqrt{2} T^{-1} \sup_{\mathbb{D}_T, 1\leq t\leq T} (e_t^\Psi)^2 \sup_{\mathbb{D}_T} \sum_{t=1}^T |\Psi_{Tt}^\delta(\xi) - \delta_t|$$
$$= O_{P_c}(T^{-1/2})$$

by (a) and (c). Next, by applying to $u = e_t^\Psi$ and $v = \varepsilon_t + T^{1/2}\delta_t\eta_t$ the inequality $|u^2 - v^2| \leq |u-v|^2 + 2|u-v||v|$, and then (A.20) to evaluate $|u-v|$, we see that

$$\sup_{\mathbb{D}_T} \|T^{-1} \sum_{t=1}^T (\delta_t, 1-\delta_t)[(e_t^\Psi)^2 - (\varepsilon_t + T^{1/2}\delta_t\eta_t)^2]\| = O_{P_c}(T^{\tau-1/2}),$$

$\tau \leq 1/4$. Inserting $T^{-1} \sum_{t=1}^T (\delta_t, 1-\delta_t)(\varepsilon_t + T^{1/2}\delta_t\eta_t)^2 = (H_T, T^{-1}\sum_{t=1}^T \varepsilon_t^2) + O_{P_c}(T^{-1/2})$ and combining with the previous display completes the proof of item (d). ■

Next, for use in contraction arguments, we evaluate the Jacobian matrix of Ψ_T. For a matrix $A = (a_{ij})$, introduce $\|A\|_1 := \sum_{i,j} |a_{ij}|$ and $\|A\|_h := \max_j \sum_i |a_{ij}|$.

Lemma A.9 *Let each row of the Jacobian matrix $(\Psi_T)'_\varsigma = (\Psi_T(\varsigma))'_\varsigma$ be evaluated at some point of the form $\xi(x)$, with x possibly varying across rows, and let $(\Psi_T)'_*$ denote the resulting matrix as a function on $\mathbb{D}_T^{T+k+4}$. Suppose that Assumption $\mathcal{M}$(b) is satisfied for some $r \geq 5$, and that $4 \leq \nu \leq r-1$. Then there exists a $\rho > 0$ such that $\sup_{\mathbb{D}_T^{T+k+4}} \|(\Psi_T)'_*\|_h = o_{P_c}(1)$.*

PROOF. In the proof of Theorem 2 we used that, if $\sup_{z\in\mathbb{K}} \|(\Phi_T)'_z\| = o_P(1)$ for some set $\mathbb{K}$, then uniform infinitesimality holds also when each row of $(\Phi_T)'_z$ is evaluated at a different point in $\mathbb{K}$. The fixed dimension of Φ_T is crucial for this implication. In the case of $(\Psi_T)'_\varsigma$, we can partition it into $(\Psi_T^\delta)'_\varsigma$, with T rows, and into $k+4$ rows containing the partial derivatives of components of Ψ_T other than Ψ_T^δ. Since the dimension of $(\Psi_T^\delta)'_\varsigma$ depends on T, to obtain uniform infinitesimality upon evaluation of its rows at possibly different points of $\mathbb{D}_T$, we cannot use the same indirect approach as for $(\Phi_T)'_z$; instead, we introduce different points of $\mathbb{D}_T$ directly. As to the remaining rows of $(\Psi_T)'_\varsigma$, since their number is fixed, it suffices to evaluate them at one and the same point of $\mathbb{D}_T$, and as for $(\Phi_T)'_z$, infinitesimality carries over to points varying across rows. Thus, to prove Lemma A.9, it is enough to find a $\rho > 0$ such that:

(a) $\sup_{(x_1,...,x_T)\in\mathbb{D}_T^T} \max_{w\in W} \sum_{t=1}^T \|(\Psi_{Tt}^\delta)'_w|_{\varsigma=\xi(x_t)}\| = o_{P_c}(1)$, where $W := \{d_t, a, \gamma, \lambda, \eta^2, \sigma^2 : t = 1, ..., T\}$ collects the components of ς, and

(b) $\sup_{x\in\mathbb{D}_T} \|(\Psi_T)'_\varsigma|_{\varsigma=\xi(x)}\|_h = o_{P_c}(1)$.

For part (a) we need some notation, and some strengthening of earlier evaluations from $\mathbb{D}_T$ to $\mathbb{D}_T^T$. For $y_s \in \mathbb{Y}_T$, $z_s \in \mathbb{A}_T$ and $x_s := (y_s', z_s')'$ ($s = 1, ..., T$), we write ξ^s, $\xi^{\delta,s}$ etc. for $\xi(x_s)$, $\xi^\delta(x_s^\delta)$ etc., and define $X_t^{\xi,s} := X_t - \sum_{u=1}^t \xi_u^{\delta,s} \Delta X_u$ and $e_t^{\xi,s} := \Delta X_t - (\xi^{a,s}, (\xi^{\gamma,s})')D_T \mathbf{X}_{t-1}^{\xi,s}$ as X_t^ξ and e_t^ξ evaluated at y_s. Then $\Psi_{Ts}^\delta(\xi^s) = \tilde{\delta}(e_s^{\xi,s}; \zeta(z_s))$, with $\zeta(z) = (z^\lambda + N_T, H_T z^\eta, z^\sigma)'$.

Now, under the conditions and with the notation of Lemma A.8, we argue for the following relations in place of Lemma A.8(a):

$$\sup_{(x_1,...,x_T)\in\mathbb{D}_T^T} \sum_{t=1}^T |\delta_t - \Psi_{Tt}^\delta(\xi^t)| = O_{P_c}(T^{-1/2}), \tag{A.21}$$

$$\sup_{(x_1,...,x_T)\in\mathbb{D}_T^T} \sum_{t=1}^T \delta_t(1 - \Psi_{Tt}^\delta(\xi^t)) = O_{P_c}(T^{-(\nu-2)/2}); \tag{A.22}$$

$$\sup_{(x_1,...,x_T)\in\mathbb{D}_T^T} \sum_{t=1}^T \boldsymbol{\delta}_{t-1} \Psi_{Tt}^\delta(\xi^t)(1 - \Psi_{Tt}^\delta(\xi^t)) = O_{P_c}(T^{-\min\{1,(\nu-2)/2\}}), \tag{A.23}$$

where $\boldsymbol{\delta}_{t-1} := \sum_{i=1}^k \delta_{t-i}$. They rely on a strengthening of Remark A.2:

REMARK A.1. If $\hat{e}_t(\cdot)$ are random functions satisfying the condition of Remark A.1, and if also $\sup_{(y_1,...,y_T)\in\mathbb{Y}_T^T} \sum_{t=1}^T (1 - \delta_t) |\hat{e}_t(y_t)|^{\nu+1} = O_P(T)$, then the conclusions of Lemma A.5 hold for

$$\sup_{\{(y_t,z_t)\in\mathbb{Y}_T\times\mathbb{A}_T:1\le t\le T\}} \sum_{t=1}^T \delta_t(1 - \tilde{\delta}(\hat{e}_t(y_t); \zeta(z_t)))$$

and

$$\sup_{\{(y_t,z_t)\in\mathbb{Y}_T\times\mathbb{A}_T:1\le t\le T\}}\sum_{t=1}^{T}|\delta_t-\tilde{\delta}\left(\hat{e}_t(y_t);\zeta(z_t)\right)|.$$

The proof of Lemma A.5 requires only notational changes. □

Given that in the proof of Lemma A.8 we have checked the condition of Remark A.1, to obtain (A.21) and (A.22) it remains to verify that $\sup_{(y_1,\dots,y_T)\in\mathbb{Y}_T^T}\sum_{t=1}^{T}(1-\delta_t)\,|e_t^{\xi,t}|^{\nu+1}=O_{P_c}(T)$. Indeed, in consequence of (A.18) and the inequalities $\rho\,(\nu+1)<1/2$, $\rho+\tau<1/2$ (see the conditions of Lemma A.8), it holds that $(1-\delta_t)\,|e_t^{\xi,t}|^{\nu+1}\le$

$$(1-\delta_t)\,4^{\nu}[|\varepsilon_t|^{\nu+1}+\|\gamma_*-\xi^{\gamma,t}\|^{\nu+1}\|\nabla\mathbf{Y}_{t-1}\|^{\nu+1} \qquad \text{(A.24)}$$
$$+(\boldsymbol{\delta}_{t-1})^{\nu+1}o_{P_c}(T^{1/2})+o_{P_c}(1)]$$

uniformly on $\mathbb{Y}_T^T$. Further, for $\nu+1\le r$, we have $E|\varepsilon_t|^{\nu+1}<\infty$, so that $\sum_{t=1}^{T}|\varepsilon_t|^{\nu+1}=O_{P_c}(T)$ and $\sum_{t=1}^{T}\|\nabla\mathbf{Y}_{t-1}\|^{\nu+1}=O_{P_c}(T)$ by Lemma A.1(c), while $\sup_{\mathbb{Y}_T^T}\|\gamma_*-\xi^{\gamma,t}\|=1$ and $\sum_{t=1}^{T}(\boldsymbol{\delta}_{t-1})^{\nu+1}\le(kN_T)^{\nu+1}=O_{P_c}(1)$. These imply the asserted uniform order of $\sum_{t=1}^{T}(1-\delta_t)\,|e_t^{\xi,t}|^{\nu+1}$, and hence, (A.21) and (A.22).

As to (A.23), $0\le\sum_{t=1}^{T}\delta_t\boldsymbol{\delta}_{t-1}\Psi_{Tt}^{\delta}(\xi^t)(1-\Psi_{Tt}^{\delta}(\xi^t))\le k\sum_{t=1}^{T}\delta_t(1-\Psi_{Tt}^{\delta}(\xi^t))\le O_{P_c}(T^{-(\nu-2)/2})$ on $\mathbb{D}_T^T$ according to (A.22), and it remains to evaluate $\sum_{t=1}^{T}(1-\delta_t)\boldsymbol{\delta}_{t-1}\Psi_{Tt}^{\delta}(\xi^t)(1-\Psi_{Tt}^{\delta}(\xi^t))\le\sum_{t=1}^{T}(1-\delta_t)\boldsymbol{\delta}_{t-1}\Psi_{Tt}^{\delta}$. Equation (A.7) holds for $\Psi_{Tt}^{\delta}(\xi^t)$ and $e_t^{\xi,t}$ in place of $\tilde{\delta}(\hat{e}_t;\zeta)$ and $\hat{e}_t$, with magnitude orders uniform on $\mathbb{D}_T^T$. By multiplying this equation with $\boldsymbol{\delta}_{t-1}$ and summing over t, $\sum_{t=1}^{T}(1-\delta_t)\boldsymbol{\delta}_{t-1}\Psi_{Tt}^{\delta}(\xi^t)$ is seen not to exceed

$$O_{P_c}(T^{-3/2})\textstyle\sum_{t=1}^{T}(1-\delta_t)\,\boldsymbol{\delta}_{t-1}+O_{P_c}(T^{-3/2})\sum_{t=1}^{T}(1-\delta_t)\,\boldsymbol{\delta}_{t-1}|e_t^{\xi,t}|^{\nu+1},$$

the evaluation being uniform on $\mathbb{D}_T^T$. Above $\sum_{t=1}^{T}(1-\delta_t)\,\boldsymbol{\delta}_{t-1}\le kN_T=O_{P_c}(1)$ and, in view of (A.24), $\sum_{t=1}^{T}(1-\delta_t)\,\boldsymbol{\delta}_{t-1}|e_t^{\xi,t}|^{\nu+1}$ is dominated on $\mathbb{D}_T^T$ by $4^{\nu}kN_T$ times

$$\max_{t:\boldsymbol{\delta}_{t-1}>0}|\varepsilon_t|^{\nu+1}+\sup_{\mathbb{D}_T^{\gamma}}\|\gamma_*-\xi^{\gamma}\|^{\nu+1}\max_{t:\boldsymbol{\delta}_{t-1}>0}\|\nabla\mathbf{Y}_{t-1}\|^{\nu+1}$$
$$+(kN_T)^{\nu+1}o_{P_c}(T^{1/2})+o_{P_c}(1)\,,$$

which is $o_{P_c}(T^{1/2})$; see Lemma A.2 with $\mathbb{I}_{\{\boldsymbol{\delta}_{t-1}>0\}}$ in place of δ_t. Combining the above orders of magnitude gives (A.23).

PROOF OF LEMMA A.9. We use the symbols $(\cdot)'_w$ and $\partial(\cdot)/\partial w$ interchangeably for Jacobian matrices. Assumption $\mathcal{M}$(b) with $r \geq 5$ implies that $\max_{t\leq T}|\varepsilon_t| = O_{P_c}(T^{\tau})$ with $\tau \leq 1/5$. To apply Lemma A.8, we take $2\rho < 1/(\nu+1)$.

For the proof of

(a) $\sup_{(x_1,\dots,x_T)\in\mathbb{D}_T^T}\max_{w\in W}\sum_{t=1}^T \|(\Psi_{Tt}^\delta)'_w|_{\varsigma=\xi(x_t)}\| = o_{P_c}(1)$,

define $\pi_{Tt}^\delta := \Psi_{Tt}^\delta(1-\Psi_{Tt}^\delta)$; by (A.23), $\Pi_T := \sum_{t=1}^T \pi_{Tt}^\delta(\xi^t) = O_{P_c}(T^{-1/2})$ on $\mathbb{D}_T^T$ (this and all convergence orders in what follows are uniform).

As $(\Psi_{Tt}^\delta)'_\lambda = \pi_{Tt}^\delta T/(T\lambda - \lambda^2)$, we have

$$\begin{aligned}\sum_{t=1}^T |(\Psi_{Tt}^\delta)'_\lambda|_{\varsigma=\xi^t}| &\leq \Pi_T \sup_{\mathbb{D}_T}[T/(T\xi^\lambda - (\xi^\lambda)^2)] \\ &= O_{P_c}(T^{-1/2})\end{aligned}$$

on $\mathbb{D}_T^T$. Next, for an argument w different from λ,

$$(\Psi_{Tt}^\delta)'_w = (\ln f_\nu)'_w(\pi_{Tt}^\delta), \tag{A.25}$$

where $f_\nu := \phi_\nu(e_t^d;\sigma^2+T\eta^2)/\phi_\nu(e_t^d;\sigma^2)$. For $w \in \{\sigma^2,\eta^2\}$, $(\ln f_\nu)'_w = O_{P_c}(1)$ on $\mathbb{D}_T^T$ as in (A.17), and thus, $\sum_{t=1}^T |(\Psi_{Tt}^\delta)'_w|_{\varsigma=\xi^t}| \leq O_{P_c}(1)\,\Pi_T = O_{P_c}(T^{-1/2})$ on $\mathbb{D}_T^T$. For $w=(a,\gamma',d')$,

$$(\ln f_\nu)'_{(a,\gamma',d')} = (\nu+1)\,[h(e_t^d,\nu\sigma^2) - h(e_t^d,\nu(\sigma^2+T\eta^2))](e_t^d)'_{(a,\gamma',d')},$$

where $h(x,y) := x/(y+x^2)^{-1}$ satisfies $|h(x,y)| \leq (4y)^{-1/2}$, $y>0$. Hence, upon evaluation at ξ^t, the factor in front of $(e_t^d)'_{(a,\gamma',d')}$ is bounded by $(\nu+1)\{(\nu\xi^{\sigma,t})^{-1/2} + [\nu(\xi^{\sigma,t}+T\xi^{\eta,t})]^{-1/2}\}/2$ in absolute value, and on $\mathbb{D}_T^T$ does not exceed the constant $K := 2(\nu+1)/(\sqrt{\nu}\sigma_{\varepsilon*})$, whereas

$$(e_t^d)'_{(a,\gamma')} = D_T\mathbf{X}_{t-1}^d \qquad \text{and} \qquad (e_t^d)'_{d_s} = (T^{-1/2}a,\gamma')\mathbb{I}_{ts}\Delta X_s, \tag{A.26}$$

$s = 1, ..., T$, with $\mathbb{I}_{ts}$ the vector of indicators $(\mathbb{I}_{\{s \leq t-1\}}, \mathbb{I}_{\{s=t-1\}}, ..., \mathbb{I}_{\{s=t-k\}})'$. Recalling (A.25), we find

$$\sum_{t=1}^{T} \|(\Psi_{Tt}^{\delta})'_{(a,\gamma')}|_{\varsigma=\xi^t}\|_1 \leq K\Pi_T \sup_{\mathbb{Y}_T; 1\leq t\leq T} \|D_T \mathbf{X}_{t-1}^{\xi}\|_1 = O_{P_c}(T^{\tau-1/2})$$

on $\mathbb{D}_T^T$, since, by Lemma A.6(c) with $\beta = \rho$, the supremum is $O_{P_c}(T^{\tau})$. Still by (A.25) and (A.26),

$$\begin{aligned}
\max_{1\leq s\leq T} \sum_{t=1}^{T} |(\Psi_{Tt}^{\delta})'_{d_s}|_{\varsigma=\xi^t}| &\leq K \max_{1\leq s\leq T} \{\sum_{t=1}^{T} \pi_{Tt}^{\delta}(\xi^t)\big[T^{-1/2}|\xi^{a,t}|\mathbb{I}_{\{s\leq t-1\}} \\
&\quad + \|\xi^{\gamma,t}\| \sum_{i=1}^{k} \mathbb{I}_{\{s=t-i\}}\big]|\Delta X_s|\} \\
&\leq K\big[\Pi_T|\xi^{a,t}| \max_{1\leq s\leq T} |T^{-1/2}\Delta X_s| \qquad \text{(A.27)} \\
&\quad + \|\xi^{\gamma,t}\| \sum_{t=1}^{T} \pi_{Tt}^{\delta}(\xi^t) \sum_{s=t-k}^{t-1} |\Delta X_s|\big].
\end{aligned}$$

From the proof of Lemma 1, $\max_{1\leq s\leq T} |\Delta X_s| = O_{P_c}(T^{1/2})$. Further, $|\xi^{a,t}| \leq T^{\rho-1/2}$ and $\Pi_T = O_{P_c}(T^{-1/2})$ on $\mathbb{D}_T^T$, so that in (A.27) the first term in brackets is $O_{P_c}(T^{\rho-1})$. As $\Delta X_s = \Delta Y_s + T^{1/2}\delta_s \eta_s$, we also have

$$\begin{aligned}
\sum_{s=t-k}^{t-1} |\Delta X_s| &\leq k \max_{t\leq T} |\Delta Y_t| + T^{1/2}\boldsymbol{\delta}_{t-1} \max_{t:\delta_t=1} |\eta_t| \\
&= O_{P_c}(T^{\tau}) + \boldsymbol{\delta}_{t-1} O_{P_c}(T^{1/2})
\end{aligned}$$

by Lemmas A.1 and A.2, so that

$$\begin{aligned}
\sum_{t=1}^{T} \pi_{Tt}^{\delta}(\xi^t) \sum_{s=t-k}^{t-1} |\Delta X_s| &\leq O_{P_c}(T^{\tau})\Pi_T \\
&\quad + O_{P_c}(T^{1/2}) \sum_{t=1}^{T} \boldsymbol{\delta}_{t-1}\pi_{Tt}^{\delta}(\xi^t) \\
&= O_{P_c}(T^{\tau-1/2})
\end{aligned}$$

by (A.23) and since $\nu \geq 4$. Therefore, $\max_{1\leq s\leq T}\sum_{t=1}^{T}|(\Psi_{Tt}^{\delta})'_{d_s}|_{\varsigma=\xi^t}| = O_{P_c}(T^{\tau-1/2})$ on $\mathbb{D}_T^T$.

Summarizing, the largest magnitude order on $\mathbb{D}_T^T$ of $\sum_{t=1}^{T}|(\Psi_{Tt}^{\delta})'_{w}|_{\varsigma=\xi^t}|$, for $w \in W$ and small $\rho > 0$, is $O_{P_c}(T^{\tau-1/2})$; hence, statement (a).

Later we use also that $\max_{w\in W}\sum_{t=1}^{T}\delta_t\|(\Psi_{Tt}^{\delta})'_{w}|_{\varsigma=\xi^t}\|_1 = O_{P_c}(T^{-1/2})$ on $\mathbb{D}_T^T$; this follows by replacing $\Pi_T = \sum_{t=1}^{T}\pi_{Tt}^{\delta}(\xi^t) = O_{P_c}(T^{-1/2})$ with $\sum_{t=1}^{T}\delta_t\pi_{Tt}^{\delta}(\xi^t) \leq \sum_{t=1}^{T}\delta_t(1-\Psi_{Tt}^{\delta}(\xi^t)) = O_{P_c}(T^{-(\nu-2)/2})$, $\nu \geq 4$, in the evaluations above; see (A.22).

(b) $\sup_{x\in\mathbb{D}_T}\|(\Psi_T)'_{\varsigma}|_{\varsigma=\xi(x)}\|_h = o_{P_c}(1)$.

We partition $(\Psi_T)'_{\varsigma}|_{\varsigma=\xi(x)}$ into blocks whose number does not depend on T, and show that, by choosing a small $\rho > 0$, the h-norm of each block can be made $o_{P_c}(1)$ on $\mathbb{D}_T$.

1. *Derivatives of* Ψ_T^{δ} : $\|(\Psi_T^{\delta})'_{\varsigma}|_{\varsigma=\xi(x)}\|_h = O_{P_c}(T^{\tau-1/2})$ by setting $x_1 = ... = x_T$ in (a).

2. *Derivatives of* $\Psi_T^{\lambda} = \sum_{t=1}^{T}\Psi_{Tt}^{\delta}$:

$$\begin{aligned}\|(\Psi_T^{\lambda})'_{\varsigma}|_{\varsigma=\xi(x)}\|_h &\leq \max_{w\in W}\sum_{t=1}^{T}\|(\Psi_{Tt}^{\delta})'_{w}|_{\varsigma=\xi(x)}\|_1 \\ &= \|(\Psi_T^{\delta})'_{\varsigma}|_{\varsigma=\xi(x)}\|_h = O_{P_c}(T^{\tau-1/2})\end{aligned}$$

as just shown.

3. *Derivatives of* $\Psi_T^{a,\gamma}$. Note that

$$\Psi_T^{a,\gamma} = [\sum_{t=1}^{T}D_T\mathbf{X}_{t-1}^{\Psi}(D_T\mathbf{X}_{t-1}^{\Psi})']^{-1}\sum_{t=1}^{T}D_T\mathbf{X}_{t-1}^{\Psi}\Delta X_t^{\Psi}$$

depends on ς only through Ψ_T^{δ}. With $S_{\Psi}^T := \sum_{t=1}^{T}D_T\mathbf{X}_{t-1}^{\Psi}(D_T\mathbf{X}_{t-1}^{\Psi})'$ and with derivatives $\partial(.)/\partial d_s$ that here happen not to depend on the point where they are evaluated, it is checked that $(\Psi_T^{a,\gamma})'_w$ equals $[S_{\Psi}^T]^{-1}$ postmultiplied by

$$D_T \sum_{t=1}^{T} \sum_{s=1}^{T} \{ \frac{\partial \mathbf{X}_{t-1}^d}{\partial d_s} \left[\Delta X_t^{\Psi} - (D_T \mathbf{X}_{t-1}^{\Psi})' \Psi_T^{a,\gamma} \right]$$

$$+ \mathbf{X}_{t-1}^{\Psi} [\frac{\partial \Delta X_t^d}{\partial d_s} - \frac{\partial (D_T \mathbf{X}_{t-1}^d)'}{\partial d_s} \Psi_T^{a,\gamma}] \} \frac{\partial \Psi_{Ts}^{\delta}}{\partial w}$$

$$= D_T \sum_{s=1}^{T} \{ \sum_{t=s+1}^{T} [\mathbb{I}_{ts} \left[(D_T \mathbf{X}_{t-1}^{\Psi})' \Psi_T^{a,\gamma} - \Delta X_t^{\Psi} \right]$$

$$+ \mathbf{X}_{t-1}^{\Psi} \mathbb{I}_{ts}' D_T \Psi_T^{a,\gamma}] - \mathbf{X}_{s-1}^{\Psi} \} \Delta X_s \frac{\partial \Psi_{Ts}^{\delta}}{\partial w}$$

(a prime denotes transposition). Since $T^{-1} S_{\Psi}^T$ (like $S_{11}^T + R_{11}^T$ in the proof of Lemma A.7) converges uniformly to the same non-singular limit as S_{11}^T, the magnitude order of $(\Psi_T^{a,\gamma})_w'$ is determined by T^{-1} times the expression above. If Z_s is the term in braces, it follows that

$$\|(\Psi_T^{a,\gamma})_w'\|_1 \leq O_{P_c}(T^{-1}) \sum_{s=1}^{T} \|Z_s\|_1 \, |\Delta X_s| \, \|(\Psi_{Ts}^{\delta})_w'\|_1, \qquad w \in W, \tag{A.28}$$

where

$$\|Z_s\|_1 \leq 2\| \sum_{t=s+1}^{T} D_T \mathbb{I}_{ts} (D_T \mathbf{X}_{t-1}^{\Psi})' \|_1 \|\Psi_T^{a,\gamma}\|_1 + \| \sum_{t=s+1}^{T} D_T \mathbb{I}_{ts} \Delta X_t^{\Psi} \|_1 + \|D_T \mathbf{X}_{s-1}^{\Psi}\|_1,$$

$$\sum_{t=s+1}^{T} D_T \mathbb{I}_{ts} (D_T \mathbf{X}_{t-1}^{\Psi})' = (T^{-1/2} \sum_{t=s+1}^{T} \mathbf{X}_{t-1}^{\Psi}, \mathbf{X}_s^{\Psi}, \mathbf{X}_{s+1}^{\Psi} \mathbb{I}_{\{s+1 \leq T\}}, \ldots , \mathbf{X}_{s+k-1}^{\Psi} \mathbb{I}_{\{s+k-1 \leq T\}})' D_T$$

has 1-norm bounded by $(T^{1/2} + k) \max_{1 \leq s \leq T} \|D_T \mathbf{X}_s^{\Psi}\|_1$, and the 1-norm of

$$\sum_{t=s+1}^{T} D_T \mathbb{I}_{ts} \Delta X_t^{\Psi} = (T^{-1/2} (X_T^{\Psi} - X_s^{\Psi}), \Delta X_{s+1}^{\Psi} \mathbb{I}_{\{s+1 \leq T\}}, \ldots, \Delta X_{s+k}^{\Psi} \mathbb{I}_{\{s+k \leq T\}})'$$

is $\leq (2+k)\max_{1\leq s\leq T}\|D_T\mathbf{X}_s^{\Psi}\|_1$. Using that $\|\Psi_T^{a,\gamma}\| = O_{P_c}(1)$ on $\mathbb{D}_T$ by Lemma A.8(b), we find that

$$\max_{1\leq s\leq T}\|Z_s\|_1 \leq O_{P_c}(T^{1/2})\max_{0\leq s\leq T}\|D_T\mathbf{X}_s^{\Psi}\|_1.$$

Inserting this into (A.28) together with $|\Delta X_s| \leq |\Delta Y_s| + T^{1/2}\delta_s|\eta_s|$ gives that $\|(\Psi_T^{a,\gamma})'_w|_{\varsigma=\xi(x)}\|_1$ does not exceed $O_{P_c}(T^{-1/2})\max_{0\leq s\leq T}\|D_T\mathbf{X}_s^{\Psi}\|_1$ times

$$\max_{s\leq T}|\Delta Y_s|\sum_{s=1}^{T}\|(\Psi_{Ts}^{\delta})'_w|_{\varsigma=\xi(x)}\|_1 + T^{1/2}\max_{s:\delta_s=1}|\eta_s|\sum_{s=1}^{T}\delta_s\|(\Psi_{Ts}^{\delta})'_w|_{\varsigma=\xi(x)}\|_1.$$

As $\max_{0\leq s\leq T}\|D_T\mathbf{X}_s^{\Psi}\|_1 = O_{P_c}(T^{\tau})$ on $\mathbb{D}_T$ (by Lemma A.6(c) with $\beta=0$), $\max_{s\leq T}|\Delta Y_s| = O_{P_c}(T^{\tau})$ (by Lemma A.1) and $\max_{s:\delta_s=1}|\eta_s| = O_{P_c}(1)$ (by Lemma A.2), while $\|(\Psi_T^{\delta})'_\varsigma|_{\varsigma=\xi(x)}\|_h = O_{P_c}(T^{\tau-1/2})$ and $\max_{w\in W}\sum_{s=1}^{T}\delta_s\|(\Psi_{Ts}^{\delta})'_w|_{\varsigma=\xi(x)}\|_1 = O_{P_c}(T^{-1/2})$ (see the proof of (a)), it obtains that $\|(\Psi_T^{a,\gamma})'_\varsigma|_{\varsigma=\xi(x)}\|_h = O_{P_c}(T^{\tau-1/2})$.

4. *Derivatives of* Ψ_T^{η}:

$$(\Psi_T^{\eta})'_w = T^{-1}\sum_{t=1}^{T}(\Psi_{Tt}^{\delta})'_w(e_t^{\Psi})^2 + 2T^{-1}\sum_{t=1}^{T}\Psi_{Tt}^{\delta}e_t^{\Psi}(e_t^{\Psi})'_w,$$

where

$$\max_{w\in W}\|\sum_{t=1}^{T}(\Psi_{Tt}^{\delta})'_w(e_t^{\Psi})^2\| \leq \max_{1\leq t\leq T}(e_t^{\Psi})^2\max_{w\in W}\sum_{t=1}^{T}\|(\Psi_{Tt}^{\delta})'_w\|_1, \text{ and}$$
$$\max_{w\in W}\|\sum_{t=1}^{T}\Psi_{Tt}^{\delta}e_t^{\Psi}(e_t^{\Psi})'_w\| \leq \max_{1\leq t\leq T}|e_t^{\Psi}|(\sum_{t=1}^{T}\Psi_{Tt}^{\delta})\max_{w\in W;1\leq t\leq T}\|(e_t^{\Psi})'_w\|_1.$$

Using the relations $\|(\Psi_T^{\delta})'_\varsigma|_{\varsigma=\xi(x)}\|_h = O_{P_c}(T^{\tau-1/2})$ and $\sum_{t=1}^{T}\Psi_{Tt}^{\delta}(\xi) = N_T + \sum_{t=1}^{T}(\Psi_{Tt}^{\delta}(\xi)-\delta_t)$, together with Lemma A.8(a,c), we find that on $\mathbb{D}_T$

$$\|(\Psi_T^{\eta})'_\varsigma|_{\varsigma=\xi(x)}\|_h \leq O_{P_c}(T^{\tau-1/2}) + O_{P_c}(T^{-1/2})\max_{w\in W;1\leq t\leq T}\|(e_t^{\Psi})'_w|_{\varsigma=\xi(x)}\|_1. \tag{A.29}$$

Thus, it remains to study $(e_t^{\Psi})'_w = \sum_{v\in W'}(e_t^d)'_v(\Psi_T^v)'_w$, where $W' := \{d_t, a, \gamma' : t = 1, ..., T\}$, and $(e_t^d)'_v$ are the derivatives (A.26) evaluated at $d = (\Psi_{T1}^{\delta}, ..., \Psi_{TT}^{\delta})'$ and $(a, \gamma')' = \Psi_T^{a,\gamma}$. For any $w \in W$, with $(\Psi_T^{\gamma})_i$ denoting the ith component of Ψ_T^{γ} ($i = 1, ..., k$), we find

$$\begin{aligned}\|\sum_{u=1}^{T}(e_t^d)'_{d_u}(\Psi_{Tu}^{\delta})'_w\|_1 &= \|T^{-1/2}\Psi_T^a\sum_{u=1}^{t-1}\Delta X_u(\Psi_{Tu}^{\delta})'_w \\ &\quad + \sum_{u=\max(t-k,1)}^{t-1}(\Psi_T^{\gamma})_{t-u}\Delta X_u(\Psi_{Tu}^{\delta})'_w\|_1 \\ &\leq \max_{1\leq u\leq T}|\Delta X_u|\big[T^{-1/2}|\Psi_T^a|\sum_{u=1}^{T}\|(\Psi_{Tu}^{\delta})'_w\|_1 \\ &\qquad + \|\Psi_T^{\gamma}\|_1\sum_{u=\max(t-k,1)}^{t-1}\|(\Psi_{Tu}^{\delta})'_w\|_1\big].\end{aligned}$$

As $\max_{u\leq T}|\Delta X_u| = O_{P_c}(T^{1/2})$ and, by Lemma A.8(b), $\sup_{\mathbb{D}_T}(|\Psi_T^a(\xi)| + \|\Psi_T^{\gamma}(\xi)\|_1) = O_{P_c}(1)$,

$$\begin{aligned}\max_{w\in W;1\leq t\leq T}\|\sum_{u=1}^{T}(e_t^d)'_{d_u}(\Psi_{Tu}^{\delta})'_w|_{\varsigma=\xi(x)}\|_1 &\leq O_{P_c}(T^{1/2})\sup_{\mathbb{D}_T}(|\Psi_T^a(\xi)| \\ &\quad +\|\Psi_T^{\gamma}(\xi)\|_1)\times\|(\Psi_T^{\delta})'_{\varsigma}|_{\varsigma=\xi(x)}\|_h \\ &\leq O_{P_c}(T^{\tau}). \qquad \text{(A.30)}\end{aligned}$$

In view of (A.29) and the expression for $(e_t^{\Psi})'_w$ given after (A.29), we conclude that on $\mathbb{D}_T$

$$\begin{aligned}\|(\Psi_T^{\eta})'_{\varsigma}|_{\varsigma=\xi(x)}\|_h &\leq O_{P_c}(T^{-1/2})\max_{w\in W;1\leq t\leq T}\|(e_t^d)'_{(a,\gamma')}(\Psi_T^{a,\gamma})'_w|_{\varsigma=\xi(x)}\|_1 \qquad \text{(A.31)} \\ &\quad +O_{P_c}(T^{\tau-1/2}),\end{aligned}$$

where $(e_t^d)'_{(a,\gamma')}$ should be read as a row vector. The maximum is $O_{P_c}(T^{2\tau-1/2})$ on $\mathbb{D}_T$ since it does not exceed $\|(\Psi_T^{a,\gamma})'_{\varsigma}|_{\varsigma=\xi(x)}\|_h \max_{1\leq t\leq T}\|D_T\mathbf{X}_{t-1}^{\Psi}\|_1$, see (A.26), and the two factors are respectively $O_{P_c}(T^{\tau-1/2})$ and $O_{P_c}(T^{\tau})$ on $\mathbb{D}_T$, the latter by Lemma A.6(c). Hence, $\|(\Psi_T^{\eta})'_{\varsigma}|_{\varsigma=\xi(x)}\|_h = O_{P_c}(T^{2\tau-1} + T^{\tau-1/2}) = O_{P_c}(T^{\tau-1/2})$.

5. *Derivatives of* Ψ_T^σ. Now $(\Psi_T^\sigma)'_w = -T^{-1}\sum_{t=1}^T(\Psi_{Tt}^\delta)'_w(e_t^\Psi)^2 + 2T^{-1}\sum_{t=1}^T(1-\Psi_{Tt}^\delta)e_t^\Psi(e_t^\Psi)'_w$, and the first normalized summation has already been shown, upon evaluation at $\xi(x)$, to be $O_{P_c}(T^{\tau-1/2})$ on $\mathbb{D}_T$ in h-norm. For the second one, it can be used that

$$\|\sum_{t=1}^T(1-\Psi_{Tt}^\delta)e_t^\Psi(e_t^\Psi)'_w\|_1 \leq \max_{1\leq t\leq T}|(1-\Psi_{Tt}^\delta)e_t^\Psi|(\sum_{t=1}^T\|(e_t^\Psi)'_w\|_1).$$

Instead of (A.30), now $\max_{w\in W}\sum_{t=1}^T\|\sum_{u=1}^T(e_t^d)'_{d_u}(\Psi_{Tu}^\delta)'_w|_{\varsigma=\xi(x)}\|_1$ is bounded on $\mathbb{D}_T$ by

$$O_{P_c}(T^{1/2})\sup\nolimits_{\mathbb{D}_T}\left[T^{1/2}|\Psi_T^a(\xi)|+k\|\Psi_T^\gamma(\xi)\|_1\right]\|(\Psi_T^\delta)'_\varsigma|_{\varsigma=\xi(x)}\|_h = O_{P_c}(T^\tau).$$

Next, $\max_{1\leq t\leq T}|(1-\Psi_{Tt}^\delta(\xi))e_t^\Psi| \leq \max_{t:\delta_t=0}|e_t^\Psi|+(\max_{1\leq t\leq T}|e_t^\Psi|)\sum_{t=1}^T|\delta_t-\Psi_{Tt}^\delta(\xi)| \leq O_{P_c}(T^\tau)$ on $\mathbb{D}_T$ by Lemma A.8(a,c). Thus, instead of (A.31),

$$\begin{aligned}\|(\Psi_T^\sigma)'_\varsigma|_{\varsigma=\xi(x)}\|_h &\leq O_{P_c}(T^{\tau-1})\max_{w\in W}\sum_{t=1}^T\|(e_t^d)'_{(a,\gamma')}(\Psi_T^{a,\gamma})'_w|_{\varsigma=\xi(x)}\|_1\\ &\quad +O_{P_c}(T^{2\tau-1})\\ &\leq T^\tau\|(\Psi_T^{a,\gamma})'_\varsigma|_{\varsigma=\xi(x)}\|_h(T^{-1}\sum_{t=1}^T\|D_T\mathbf{X}_{t-1}^\Psi\|_1)\\ &\quad +O_{P_c}(T^{2\tau-1}) = O_{P_c}(T^{2\tau-1/2}),\end{aligned}$$

since $\sum_{t=1}^T\|D_T\mathbf{X}_{t-1}^\Psi\|_1 \leq \sum_{t=1}^T\|D_T\mathbf{Y}_{t-1}\|_1+\sum_{t=1}^T|X_{t-1}^\Psi-Y_{t-1}|+k\sum_{t=1}^T|\Delta Y_t-\Delta X_t^\Psi| \leq O_{P_c}(T)$ by the standard properties of Y_t and Lemma A.6(a,b) with $\beta=0$.

Summarizing, among the considered finitely many blocks of $(\Psi_T)'_\varsigma$, for small $\rho>0$ the largest magnitude order is $O_{P_c}(T^{2\tau-1/2})$, which for $\tau=1/5$ is $o_{P_c}(1)$. ■

A.4.2 Proof of Theorem 4

The argument is similar to that for Theorem 2. For $\xi(\cdot)$ defined in the introduction to section A.4, let $\Xi_T := \xi^{-1}\circ\Psi_T\circ\xi$; it is well-defined P_c-a.s.

Let $\rho > 0$ be chosen such that the conclusions of Lemmas A.8 are applicable; this choice is possible under the conditions of Theorem 2. Further, let $\Xi_T^{tr} : \mathbb{D}_T \to \mathbb{D}_T$ be obtained from Ξ_T by truncating it as follows: $(\Xi_T^{tr})^\delta(x) := \Xi_T^\delta(x)$ if $\Xi_T^\delta(x) \in \mathbb{D}_T^\delta$ and $(\Xi_T^{tr})^\delta(x) := T^{\rho-1/2}\Xi_T^\delta(x)/\Sigma_{t=1}^T|\Xi_{Tt}^\delta(x)|$ otherwise, $(\Xi_T^{tr})^\gamma(x) := \Xi_T^\gamma(x)$ if $\Xi_T^\gamma(x) \in \mathbb{D}_T^\gamma$ and $(\Xi_T^{tr})^\gamma(x) := \gamma_* + (\Xi_T^\gamma(x) - \gamma_*)/\|\Xi_T^\gamma(x) - \gamma_*\|$ otherwise, and $(\Xi_T^{tr})^\lambda$, $(\Xi_T^{tr})^a$, $(\Xi_T^{tr})^\eta$, $(\Xi_T^{tr})^\sigma$ defined similarly to the components of Θ_T^{tr} in the proof of Theorem 2. Then Ξ_T^{tr} has a random fixed point $\varkappa_T$ on $\mathbb{D}_T$ by Theorem 10 in Bharucha-Reid (1976). Similarly to the proof of Theorem 2, but invoking Lemma A.8(a,b,d) instead of Lemma 1, we conclude that $P_c(\Psi_T(\xi(\varkappa_T)) = \xi(\varkappa_T)) \to 1$. The sequence ξ_T whose existence is asserted in (i) can, accordingly, be defined as $\xi_T := \xi(\varkappa_T)$.

Item (ii) follows from Lemma A.8 since $\varkappa_T \in \mathbb{D}_T$. Item (iii) follows once noticed that the ADF statistics based on the de-jumped series are

$$ADF_{\hat{\alpha}}^\Psi = T^{1/2}\Psi_T^a(\xi_T)(1 - \mathbf{1}_k'\Psi_T^\gamma(\xi_T))^{-1}$$

and

$$ADF_t^\Psi = T^{-1/2}\Psi_T^a(\xi_T)(\sigma_\Psi^2(\xi_T)v_{11})^{-1/2},$$

where $\sigma_\Psi^2(\xi_T)$ is the residual variance from (13), and v_{11} is the first entry of the matrix $[\sum_{t=1}^T \mathbf{X}_{t-1}^\Psi(\mathbf{X}_{t-1}^\Psi)']^{-1}$, both evaluated at ξ_T. As $\varkappa_T \in \mathbb{D}_T$, Lemma A.8 allows us to invoke Lemma A.7(a), and to obtain the limiting distributions of the statistics as in the proof of Theorem 3.

In the rest of the proof we assume that $r \geq 5$, so that we can apply Lemma A.9.

Uniqueness is established next. Fix an $\epsilon \in (0, 1/2)$. Since $\mathbb{D}_T$ is convex, for every $x_1, x_2 \in \mathbb{D}_T$ by the mean-value theorem $\Xi_T(x_1) - \Xi_T(x_2) = (\Xi_T)_*'(x_1 - x_2)$, where $(\Xi_T)_*'$ is the Jacobian matrix $(\Xi_T)_x'$ with each row evaluated at some point in $\mathbb{D}_T$, possibly varying across rows and also with x_1, x_2. We have

$$\begin{aligned}\sup_{\mathbb{D}_T^{T+k+4}} \|(\Xi_T)_*'\|_h &\leq \max(H_T, H_T^{-1}) \sup_{\mathbb{D}_T^{T+k+4}} \|(\Psi_T)_*'\|_h \\ &= o_{P_c}(1)\end{aligned}$$

in view of Assumption $\mathcal{S}$(b) and Lemmas A.2 and A.9, where $(\Psi_T)'_*$ is as in Lemma A.9. Therefore, we can find a set $\mathcal{B}_1$, with $P_c(\mathcal{B}_1) > 1 - \epsilon/4$, such that for $T > T_1$ (say) and outcomes in $\mathcal{B}_1$, $\|(\Xi_T)'_*\|_h \leq 1/2$ on $\mathbb{D}_T^{T+k+4}$. Further, $\|\Xi_T(x_1) - \Xi_T(x_2)\|_1 \leq \|x_1 - x_2\|_1 \sup_{\mathbb{D}_T^{T+k+4}} \|(\Xi_T)'_*\|_h \leq (1/2)\|x_1 - x_2\|_1$, i.e. Ξ_T is Lipschitz on $\mathbb{D}_T$ with modulus 1/2. On the other hand, in view of Lemma A.8(a,b,d), there exist a set $\mathcal{B}_2 \subset \mathcal{B}_1$, with $P_c(\mathcal{B}_2) > 1-\epsilon/2$, and an integer $T_2 \geq T_1$ such that, on $\mathcal{B}_2$ and for $T > T_2$, $\Xi_T(\mathbb{D}_T) \subset \mathbb{D}_T$. Thus, for $T > T_2$ and outcomes in $\mathcal{B}_2 \subset \mathcal{B}_1$, Ξ_T is a contraction on $\mathbb{D}_T$, and by Banach's fixed point theorem, Ξ_T has a unique (per outcome) fixed point on $\mathbb{D}_T$.

Further, if ξ_T and ς_T satisfy item (i), there exists a set $\mathcal{B}_3$, with $P_c(\mathcal{B}_3) > 1-\epsilon/4$, such that on $\mathcal{B}_3$ and for $T > T_3$ (say), $\Psi_T(\xi_T) = \xi_T$ and $\Psi_T(\varsigma_T) = \varsigma_T$. If they also satisfy items (ii) and (iii), so that $\Psi_T(\xi_T) = \xi(\varkappa_\infty) + o_{P_c}(1)$ and $\Psi_T(\varsigma_T) = \xi(\varkappa_\infty) + o_{P_c}(1)$, with $\varkappa_\infty := (0_{T+1}, \gamma'_*, 0, 1, \sigma^2_{\varepsilon *})'$ and with convergence at rate faster than the shrinkage rates of $\mathbb{D}_T^\delta$ and $\mathbb{D}_T^a$ for the respective components $\Psi_T^\delta(\xi_T), \Psi_T^\delta(\varsigma_T)$ and $\Psi_T^a(\xi_T), \Psi_T^a(\varsigma_T)$, then on some set $\mathcal{B}_4 \subset \mathcal{B}_3$ with $P_c(\mathcal{B}_4) > 1 - \epsilon/2$, and for $T > T_4 \geq T_3$ (say), $\xi^{-1}(\xi_T) \in \mathbb{D}_T$ and $\xi^{-1}(\varsigma_T) \in \mathbb{D}_T$. On $\mathcal{B}_4 \subset \mathcal{B}_3$ and for $T > T_4$, $\xi^{-1}(\xi_T)$ and $\xi^{-1}(\varsigma_T)$ are fixed points of Ξ_T on $\mathbb{D}_T$, while on $\mathcal{B}_2 \cap \mathcal{B}_4$ with $P_c(\mathcal{B}_2 \cap \mathcal{B}_4) > 1 - \epsilon$ and for $T > \max(T_2, T_4)$, we must have $\xi^{-1}(\xi_T) = \xi^{-1}(\varsigma_T)$. As ϵ is arbitrary, uniqueness of ξ_T with P_c-probability tending to 1 follows.

Finally, consider computation. From the proofs of Theorems 2 and 3, $x_{T0} := \xi^{-1}(\xi_{T0}) \xrightarrow{P_c} \varkappa_\infty$ at a faster rate than the shrinkage rates of $\mathbb{D}_T^\delta$ and $\mathbb{D}_T^a$ for x_{T0}^δ and x_{T0}^a respectively. Thus, $x_{T0} \in \mathbb{D}_T$ on some $\mathcal{B}_5$ with $P_c(\mathcal{B}_5) > 1 - \epsilon/2$, and for $T > T_5$ (say). By Banach's fixed point theorem, on $\mathcal{B}_2 \cap \mathcal{B}_5$ and for $T > \max(T_2, T_5)$, the sequence defined by $x_{Ti} := \Xi(x_{T,i-1})$, $i \geq 1$, converges as $i \to \infty$ to the unique (per outcome) fixed point of Ξ_T on $\mathbb{D}_T$, which on $\mathcal{B}_2 \cap \mathcal{B}_4 \cap \mathcal{B}_5$ equals $\xi^{-1}(\xi_T)$. Since $P_c(\mathcal{B}_2 \cap \mathcal{B}_4 \cap \mathcal{B}_5) > 1 - 2\epsilon$, we obtain the convergence $\xi_{Ti} = \xi(x_{Ti}) \to \xi_T$ as $i \to \infty$ with P_c-probability tending to 1 as $T \to \infty$. ■

S Appendix: Supplementary Monte Carlo results

S.1 Comparisons with tests based on traditional level shift estimation method

Here we compare our tests with two further proposals: using our de-jumped-based tests with the level shift dates estimated using traditional level shift detection methods. Specifically:

– *Chen-Tiao* [CT] $ADF^{\delta}_{\hat{\alpha}}, ADF^{\delta}_{t}$: unit root tests based on Chen and Tiao's (1990) additive shift detection procedure and on de-jumping under the null. We have used 2.8 as critical value for shift detection, which Chen and Tiao suggest as the most liberal choice; choosing a more conservative value doesn't improve the size/power properties;

– *Chen–Liu* [CL] $ADF^{\delta}_{\hat{\alpha}}, ADF^{\delta}_{t}$: unit root tests based on shifts detected as above and on de-jumping under the general model (i.e., without imposing a unit root). Specifically, de-jumping is performed according to equation (20) in Chen and Liu (1993, p.287), as they suggest at step II.1 on p.287.

We show in Table S.1 the results obtained through Monte Carlo simulation in the four level shift model (cf. Section 5). Overall, our method (in particular, the 'finer' de-jumping method, see Section 4) delivers the most accurate sizes for all the sample sizes considered. Note that the largest size distortions occur with the use of CL, which confirms the importance of de-jumping under the null hypothesis for controlling test size. Further, our 'finer' de-jumping method appears to be the best method also in terms of power. Tests based on CL also have, in general, good power properties but recall their failure to control size.

Notice that a further drawback, or rather, inconvenience of standard methods is that they require the practitioner to choose a threshold value for the sequence of test statistics associated with the presence of outliers. Suggested thresholds are typically simulated and depend on the DGP. In contrast, in our proposal the only parameter that has to be chosen is the degrees of freedom ν of the smoothing t-densities, and we suggest to choose it arbitrarily large; see section 5 for some Monte Carlo comparisons.

S.1 Comparisons with tests based on a 'binary' level shift date estimator

As we discuss in the new Remarks 3.1-3.2, in our framework a 0-1 estimator with the interpretation of a standard outlier detection estimator can be constructed as follows:

$$\check{\delta}_s = \begin{cases} 1 & \text{if } \tilde{\delta}_s > \kappa \\ 0 & \text{otherwise} \end{cases},$$

where $\tilde{\delta}_s$ is our smooth estimator and $\kappa \in (0, 1)$. We have carried out Monte Carlo experiments over a grid of values of κ, and found no improvements either under the null hypothesis or under (local and non-local) alternatives. See Table S.2 for $\kappa = 0.5$, where a small power deterioration can be observed when the binary version of the level shift estimator is used.

Notice, finally, that tests based on the binary estimators typically perform better than tests based on CT and CL, see section S.1.

S.1 Student t innovations: extended tables

In Tables S.3 and S.4 we report some extended Monte Carlo results obtained for all the 3 models considered in the paper (see section 5). Results are qualitatively similar to those reported in Section 5.

S.1 Power against fixed alternatives: further results

In this section we briefly investigate the power properties of the tests for finite T and alternatives further away from the null hypothesis. In Figure 1 we report the size-adjusted power of the tests (both the basic version, long dashed line in the figure, and the finer version, thick line in the figure), for the DGP S_4 (four level shifts at fixed dates), under $T = 100, 200, 400$, $\gamma = -0.5, 0, 0.5$ and α ranging from 1 to 0. When simulating the data, the errors u_t have been normalized in order to have $Var\,(\Delta Y_t)$ constant across different values of α and γ; this is done in order to make the size of the level shifts, relative to the variance of ΔY_t, independent of α and γ. The tests are also compared with a test based on de-jumping combined with level shift dates estimated through Bai and Perron (1998) break date estimator (short dashed line in the figure).

The following conclusions can be drawn from the inspection of the figure:

1. The finer version of our test dominates the basic version. Hence, for alternatives far from the null it is preferable to jointly estimate the level shift indicators and the autoregressive parameters, as suggested in section 4.

2. The size-adjusted power is not monotonic in α. Nevertheless, and especially for the finer de-jumping test, power tends to stabilize as α declines. Fortunately, there is no evidence that, for values of α far from the null, the power becomes very small. The important message that comes from this Monte Carlo exercise is that the finer version of the tests should be preferable to the tests based on the basic de-jumping method.

3. As T increases, power increases for all considered values of α, hence confirming the consistency result provided in Theorem 3.

4. The implementation of the Bai-Perron (1998) break date estimator algorithm leads to higher power for some configurations of γ and T if α is sufficiently far from the null hypothesis. This test seems to be an interesting complement to our tests, especially for moderate sample sizes and alternatives far from the null.

References

Aitchison J. and S.D. Silvey, Maximum likelihood estimation of parameters subject to restraints, *Annals of Statistics* 29, 813-828.

Amsler C. and J. Lee (1995), An LM test for a unit root in the presence of a structural change, *Econometric Theory* 11, 359–368.

Bai J. and P. Perron (1998), Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 47–78.

Balke N.S. and T.B. Fomby (1991a), Shifting trends, segmented trends, and infrequent permanent shocks, *Journal of Monetary Economics* 28, 61–85.

Balke N.S. and T.B. Fomby (1991b), Infrequent permanent shocks and the finite-sample performance of unit root tests, *Economics Letters* 36, 269–73.

Banerjee A., R.L. Lumsdaine and J.H. Stock (1992), Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence, *Journal of Business and Economic Statistics* 10, 271-287.

Bharucha-Reid A.T. (1976), Fixed point theorems in probabilistic analysis, *Bulletin of the American Mathematical Society* 82, 641-657

Busetti F. and A.C. Harvey (2001), Testing for the presence of a random walk in series with structural breaks, *Journal of Time Series Analysis* 22, 127–150.

Busetti F. and A.M.R. Taylor (2003), Testing against stochastic trends and seasonality in the presence of unattended breaks and unit roots, *Journal of Econometrics* 117, 21–53.

Cavaliere G. (2004), Unit root tests under time-varying variances, *Econometric Reviews* 23, 259–292.

Chang Y. and J.Y. Park (2002), On the asymptotics of ADF tests for unit roots, *Econometric Reviews* 21, 431–447.

Chen C. and G.C.Tiao (1990), Random level shift time series models, ARIMA approximations and level shift detection, *Journal of Business and Economic Statistics* 8, 170–186.

Chen C. and L.-M. Liu (1993) Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association* 88, 284–297.

Clemente J., A. Montañés and M. Reyes (1998), Testing for a unit root in variables with a double change in the mean. *Economics Letters* 59, 175–182.

Doornik J. (2001), *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultatnts, London.

Elliott G., T.J. Rothemberg and J.H. Stock (1996), Efficient tests for an autoregressive unit root, *Econometrica* 64, 813–836.

Franses P.H. and N. Haldrup (1994), The effects of additive outliers on tests for unit roots and cointegration, *Journal of Business and Economic Statistics* 12, 471–78.

Franses P.H. and A. Lucas (1998), Outlier detection in cointegration analysis, *Journal of Business and Economic Statistics* 16, 459–468.

Fuller W. (1976), *Introduction to Statistical Time Series*, Wiley, New York.

Georgiev I. (2005), Functional Poisson convergence in asymptotics for linear processes, Universidade Nova de Lisboa, manuscript

Hansen B.E. (1992), Convergence to stochastic integrals for dependent heterogeneous processes, *Econometric Theory* 8, 489–500.

Jacod J. and A. Shiryaev (2003), *Limit Theorems for Stochastic Processes*, Springer Verlag, Berlin.

Kapetanios G. (2005), Unit-root testing against the alternative hypothesis of up to m structural breaks, *Journal of Time Series Analysis* 26, 123-133.

Kwiatkowski D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic series have a unit root?, *Journal of Econometrics* 54, 159–178.

Kurtz T. and P. Protter (1991), Weak limit theorems for stochastic integrals and stochastic differential equations, *Annals of Probability* 19, 1035-1070.

Lanne M., H. Lütkepohl and P. Saikkonen (2002), Comparison of unit root tests for time series with level shifts, *Journal of Time Series Analysis* 23, 667–685.

Leipus R. and M.-C. Viano (2003), Long memory and stochastic trend, *Statistics and Probability Letters* 61, 177–190.

Lütkepohl H., P. Saikkonen and C. Trenkler (2004), Testing for the cointegration rank of a VAR process with level shift at unknown time, *Econometrica* 72, 647–662.

Leybourne S., P. Newbold and D. Vougas (1998), Unit roots and smooth transitions, *Journal of Time Series Analysis* 19, 83–97.

Leybourne S. and P. Newbold (2000a), Behaviour of the standard and symmetric Dickey-Fuller type tests when there is a break under the null hypothesis, *Econometrics Journal* 3, 1-15.

Leybourne S. and P. Newbold (2000b), Behavior of Dickey-Fuller *t*-tests when there is a break under the alternative hypothesis, *Econometric Theory* 16, 779-789

Lumsdaine R.L. and D.H. Papell (1997), Multiple trend breaks and the unit root hypothesis, *Review of Economics and Statistics* 79, 212–218.

Müller U.K. and G. Elliott (2003), Tests for unit roots and the initial condition, *Econometrica* 71, 1269–1286.

Nelson C.R., Piger, J. and Zivot, E. (2001), Markov regime-switching and unit root tests, *Journal of Business and Economics Statistics* 19, 404–415.

Ng S. and P. Perron (1995),Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag, *Journal of the Americal Statistical Association* 90, 268–281.

Ng S. and P. Perron (2001), Lag length selection and the construction of unit root tests with good size and power, *Econometrica* 69, 1519–1554.

Papell D.H., C.J. Murray, H. Ghiblawi (2000), The structure of unemployment, *Review of Economics and Statistics* 82, 309–315.

Perron P. (1989), The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* 57, 1361–401.

Perron P (1990), Testing for a unit root in a time series with a changing mean, *Journal of Business and Economic Statistics* 8, 153–162.

Perron P. (1992), Trend, unit root and structural change: a multi-country study with historical data, *Proceedings of the Business and Economic Statistics Section, Americal Statistical Association*, 144–149.

Perron P. (2005), Dealing with structural breaks, *Palgrave Dictionary of Econometrics* 1, forthcoming.

Perron P. and Z. Qu (2004), An analytical evaluation of the log-periodogram estimate in the presence of level shifts and its implications for stock returns volatility, Manuscript, Department of Economics, Boston University.

Perron P. and Z. Qu (2006), A simple modification to improve the finite sample properties of Ng and Perron's unit root tests, *Economics Letters*, forthcoming.

Perron P. and G. Rodriguez (2003), GLS detrending, efficient unit root tests and structural change, *Journal of Econometrics* 115, 1–27.

Perron P. and T.J. Vogelsang (1992), Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business and Economic Statistics* 10, 301–320.

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

Phillips P.C.B. (1988), Regression theory for near-integrated time series, *Econometrica* 56, 1021–1043

Saikkonen P. and H. Lütkepohl (2001) Testing for unit roots in time series with level shifts, *Allgemeines Statistisches Archiv* 85, 1–25.

Saikkonen P. and H. Lütkepohl (2002), Testing for a unit root in a time series with a level shift at unknown time, *Econometric Theory* 18, 313-348.

Vogelsang T.J. and P.Perron (1998) Additional tests for a unit root allowing for a break in the trend function at an unknown time, *International Economic Review* 39, 1073–1100.

White H. (2001), *Asymptotic theory for econometricians*, Academic Press, San Diego.

Zivot E. and D. Andrews (1992), Further evidence on the Great crash, the oil price shock and the unit root hypothesis, *Journal of Business and Economic Statistics* 10, 251-270.

Table 1. Empirical size of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}). Raw data.

		Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}
100	−0.5	5.1	5.1	5.0	5.0	5.8	5.9	3.0	3.2	3.9	4.0	6.5	6.9	3.6	3.6	4.3	4.4	6.3	6.5
	0	5.2	5.1	5.2	5.2	5.7	5.7	4.9	4.9	5.4	5.3	6.9	6.7	4.9	4.8	5.2	5.3	6.8	7.0
	0.5	5.5	5.4	5.5	5.4	6.0	5.7	3.6	3.7	4.0	3.8	5.5	5.4	3.5	3.5	4.0	3.9	5.4	5.4
200	−0.5	5.3	5.4	5.2	5.2	5.7	5.7	3.0	3.1	4.9	5.1	6.2	6.3	3.9	3.9	4.5	4.6	5.6	5.8
	0	4.6	4.6	4.8	4.7	5.0	4.9	4.8	4.8	5.2	5.2	5.8	5.8	4.8	4.7	5.2	5.2	5.7	5.6
	0.5	5.0	4.8	4.9	4.8	5.2	5.0	3.3	3.3	4.3	4.3	4.9	4.9	3.7	3.6	4.7	4.7	5.4	5.3
400	−0.5	4.9	5.1	4.9	5.0	5.1	5.2	2.6	2.8	4.7	4.8	5.1	5.2	3.5	3.6	4.7	4.6	5.2	5.0
	0	4.8	4.8	4.8	4.8	4.9	4.9	4.3	4.4	4.8	4.9	5.0	5.1	4.8	4.7	5.1	5.0	5.3	5.2
	0.5	5.2	5.2	5.0	5.1	5.2	5.2	3.5	3.4	5.0	5.0	5.3	5.3	3.5	3.5	5.1	4.9	5.2	5.1

Notes: ADF$^{\delta}$ and ADF$^{\Psi}$ denote the ADF tests under rough de-jumping and finer de-jumping. In both cases for estimation of δ_t a Gaussian distribution is used as a proxy for Student-t with large ν. Asymptotic critical values at the 5% level as reported in Fuller (1976) are employed.

Table 2. Size-adjusted local power of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}). Raw data.

		Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	-0.5	50.7	51.5	48.1	48.2	45.3	47.5	0.0	0.0	10.2	9.9	12.5	12.3	10.2	10.2	12.0	12.0	14.9	14.8
	0	49.3	50.1	46.1	46.1	45.9	46.7	0.2	0.2	19.3	19.4	19.8	20.1	13.3	13.5	19.9	20.1	20.0	19.8
	0.5	43.0	43.8	42.5	42.7	42.0	42.6	10.1	10.1	27.3	27.9	30.0	30.0	23.7	23.4	29.1	28.9	30.4	30.7
200	-0.5	47.9	48.9	45.8	47.0	45.2	45.9	0.0	0.0	20.0	19.7	25.0	25.9	9.1	9.2	23.5	23.2	29.2	29.0
	0	51.8	52.0	49.3	50.1	49.3	50.0	0.2	0.2	34.1	34.9	34.0	35.1	13.9	14.1	35.6	35.6	36.7	36.4
	0.5	47.1	47.7	46.9	47.6	47.0	47.5	10.8	10.4	40.9	41.0	43.2	42.6	24.3	24.5	38.6	38.8	41.1	41.7
400	-0.5	49.2	48.8	47.7	47.5	47.7	48.0	0.0	0.0	38.0	37.6	38.3	38.4	9.7	9.8	38.8	38.8	39.4	40.2
	0	50.3	50.4	50.8	50.5	50.1	50.3	0.2	0.2	44.7	44.4	44.9	44.1	12.9	13.2	42.5	43.5	42.6	43.2
	0.5	47.0	47.1	47.1	47.3	46.9	46.9	11.0	11.2	45.4	46.3	45.4	45.7	24.3	24.7	44.3	45.6	44.9	45.9

Notes: Local power is evaluated at α=1−c/T with c=7. See also Table 1.

Table 3. Size-adjusted power against fixed alternatives of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}). Raw data.

		Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	-0.5	76.3	76.8	70.8	71.1	68.3	70.5	0.0	0.0	11.2	10.7	15.6	15.3	11.2	11.4	14.3	14.0	19.4	19.2
	0.0	72.7	73.4	68.4	68.0	68.4	68.8	0.0	0.0	24.6	24.7	25.7	26.2	15.8	15.7	25.7	25.6	26.8	26.4
	0.5	63.8	64.4	61.9	62.3	61.8	62.6	7.9	7.8	35.0	34.8	39.2	39.3	30.0	29.2	38.8	38.2	41.6	41.4
200	-0.5	99.9	99.8	96.2	95.9	96.7	96.6	0.0	0.0	37.5	38.0	55.3	55.5	12.1	12.3	42.2	41.1	59.9	59.5
	0.0	99.7	99.6	97.4	97.3	97.7	97.5	0.0	0.0	72.9	71.9	75.4	74.9	20.7	20.8	73.7	72.5	76.8	76.0
	0.5	98.8	98.6	97.9	97.7	98.2	98.0	4.1	3.8	82.4	81.8	88.0	87.2	43.8	42.6	82.8	82.1	88.8	88.3
400	-0.5	100.0	100.0	98.6	98.5	99.1	99.1	0.0	0.0	83.8	82.8	89.8	89.2	13.5	13.2	84.9	84.0	89.4	89.1
	0.0	100.0	100.0	99.8	99.8	99.8	99.8	0.0	0.0	98.3	98.0	98.7	98.3	23.8	23.7	97.2	97.0	97.6	97.5
	0.5	100.0	100.0	100.0	100.0	100.0	100.0	0.6	0.5	99.8	99.8	99.9	99.9	51.1	50.2	99.0	99.2	99.3	99.4

Notes: Power is evaluated at $\alpha = 0.9$. See also Table 1.

Table 4. Size and size-adjusted power of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}).
Four level shifts at fixed fractions and of fixed size. Raw data.

		Size ($\alpha = 1$)						Size-Adjusted Power ($\alpha = 1 - c/T$)						Size-Adjusted Power ($\alpha = 0.9$)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	2.6	2.6	4.0	4.1	6.7	6.7	0.0	0.0	10.4	10.0	12.8	12.4	0.0	0.0	12.0	11.9	15.4	15.6
	0.0	4.5	4.6	5.3	5.5	6.9	7.0	0.2	0.1	19.3	19.2	19.5	20.0	0.0	0.0	24.3	23.9	25.2	25.2
	0.5	3.8	3.7	4.1	4.0	5.4	5.4	10.1	10.1	27.3	27.9	29.8	29.8	8.1	7.7	35.0	34.8	40.3	40.1
200	−0.5	3.6	3.8	4.5	4.7	6.2	6.2	0.1	0.1	13.5	13.4	17.6	17.3	0.0	0.0	21.5	20.9	31.5	32.2
	0.0	4.7	4.8	4.8	4.8	5.7	5.8	2.8	2.7	27.1	26.9	27.2	27.3	0.0	0.0	51.0	50.4	54.3	53.8
	0.5	3.9	3.9	4.3	4.2	5.1	5.0	21.4	21.8	34.1	34.0	35.9	36.2	31.2	30.7	68.8	69.0	77.3	77.3
400	−0.5	4.2	4.3	4.3	4.2	5.0	5.0	1.7	1.6	16.7	16.6	22.0	22.0	0.0	0.0	35.1	34.1	53.9	53.3
	0.0	4.9	4.9	4.9	4.9	5.4	5.2	12.6	12.1	32.4	32.2	32.8	32.0	1.7	1.6	70.5	69.7	73.5	73.1
	0.5	4.5	4.3	4.3	4.3	4.8	4.7	32.3	32.1	37.5	38.3	40.1	40.7	93.9	92.2	95.4	94.8	97.1	96.6

Notes: Shifts are generated as $\mu_t := [4.0 I_{\{t \geq \lfloor 0.2T \rfloor\}} + 3.5 I_{\{t \geq \lfloor 0.35T \rfloor\}} - 3.5 I_{\{t \geq \lfloor 0.6T \rfloor\}} + 4.0 I_{\{t \geq \lfloor 0.8T \rfloor\}}]$. Local power is evaluated at α=1-c/T with c=7. See also Table 1.

Table 5. Size-adjusted power of standard ADF tests (ADF) and of the modi□ed ADF tests (ADF^{δ} and ADF^{Ψ}) with $\nu = 8$ and $\nu = \infty$. Model S_4 with standardized $t\,(10)$ innovations

				$\nu = 8$				$\nu = \infty$			
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	0.0	0.0	0.2	0.2	0.4	0.4	9.4	9.8	11.5	11.8
	0	0.2	0.2	1.3	1.3	1.5	1.4	18.7	19.2	18.7	18.6
	0.5	10.2	9.7	12.2	11.9	13.2	12.7	25.5	26.2	29.1	30.2
200	−0.5	0.0	0.0	1.7	1.7	3.8	4.1	20.3	20.6	23.3	23.4
	0	0.2	0.2	9.7	9.2	9.6	9.7	30.9	32.0	31.3	31.4
	0.5	10.1	9.7	17.5	18.3	20.8	21.3	38.5	38.5	39.7	39.5
400	−0.5	0.0	0.0	15.1	14.9	25.9	25.9	31.2	31.1	29.8	29.0
	0	0.1	0.1	33.3	33.7	33.2	33.5	38.1	37.7	38.5	38.7
	0.5	11.5	11.1	36.4	36.5	40.3	41.0	45.6	45.5	45.8	45.7

Notes: Power is evaluated at α=1-c/T with c=7. See also Table 1.

Table 6. Size and size-adjusted power of various unit root test under a single level shift at $\lfloor \tau T \rfloor$, $\tau = 0.5$. Raw data

						Size ($\alpha = 1$)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	ADF_t^{PV}	ADF_t^{PR}	$ADF_{t.\tau}^{P}$	$ADF_{t.\tau}^{PR}$
100	-0.5	4.1	4.2	4.9	5.1	6.4	6.6	9.0	14.6	5.6	6.6
	0.0	4.9	4.9	5.2	5.3	6.3	6.3	5.8	12.0	5.5	6.7
	0.5	4.4	4.4	4.7	4.8	5.6	5.5	4.2	7.3	5.5	6.5
200	-0.5	4.3	4.3	5.0	5.0	5.5	5.4	9.8	11.6	5.3	5.6
	0.0	5.1	5.0	5.2	5.1	5.5	5.4	6.5	9.8	5.4	5.7
	0.5	4.4	4.3	4.8	4.7	5.2	5.1	4.1	5.7	5.6	5.8
400	-0.5	4.1	4.3	4.9	4.9	5.1	5.0	10.9	10.0	5.0	5.3
	0.0	5.3	5.3	5.0	5.0	5.1	5.1	7.0	8.4	5.0	5.4
	0.5	4.7	4.8	5.1	5.0	5.3	5.1	5.0	5.0	5.1	5.6
						Size-Adjusted Power ($\alpha = 1 - c/T$)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	ADF_t^{PV}	ADF_t^{PR}	$ADF_{t.\tau}^{P}$	$ADF_{t.\tau}^{PR}$
100	-0.5	2.6	2.4	18.7	18.7	24.8	25.1	20.0	18.6	25.5	41.8
	0.0	12.2	12.1	32.6	32.3	30.6	30.9	17.0	24.7	25.1	45.0
	0.5	31.3	30.8	37.2	36.8	38.3	38.4	12.0	34.3	23.3	43.6
200	-0.5	2.3	2.2	35.9	35.6	40.4	40.3	18.9	18.2	24.2	45.7
	0.0	12.2	12.2	43.2	43.5	43.5	43.3	15.4	26.1	24.5	47.7
	0.5	32.2	32.0	44.2	45.2	44.6	45.9	11.0	37.6	22.5	44.9
400	-0.5	2.1	2.0	44.6	45.2	45.1	45.9	18.4	18.1	25.6	48.0
	0.0	11.0	10.6	47.9	48.5	48.0	48.6	15.0	26.6	25.8	49.5
	0.5	31.5	31.5	46.8	47.9	46.5	47.5	8.7	39.6	25.1	48.3
						Size-Adjusted Power ($\alpha = 0.9$)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	ADF_t^{PV}	ADF_t^{PR}	$ADF_{t.\tau}^{P}$	$ADF_{t.\tau}^{PR}$
100	-0.5	0.9	0.9	24.4	23.9	35.3	35.3	33.3	28.3	43.0	61.0
	0.0	11.4	11.2	43.8	43.4	43.6	44.1	27.5	38.4	40.8	66.8
	0.5	40.9	40.6	53.1	52.6	55.0	55.2	19.2	50.5	35.5	64.2
200	-0.5	0.0	0.0	71.7	70.9	85.5	85.1	79.1	74.2	91.7	95.2
	0.0	6.6	6.4	92.1	91.5	93.1	92.6	71.3	86.7	89.6	98.6
	0.5	68.2	65.9	94.8	94.6	96.7	96.4	53.0	93.9	81.8	98.1
400	-0.5	0.0	0.0	97.3	97.1	98.9	98.8	99.9	98.0	100.0	99.8
	0.0	1.7	1.5	99.7	99.7	99.8	99.8	99.8	100.0	100.0	100.0
	0.5	93.4	90.7	100.0	100.0	100.0	100.0	97.5	100.0	100.0	100.0
						Size-Adjusted Power ($\alpha = 0.9$), fixed level shift size					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	ADF_t^{PV}	ADF_t^{PR}	$ADF_{t.\tau}^{P}$	$ADF_{t.\tau}^{PR}$
200	-0.5	4.1	3.6	40.8	40.5	57.9	57.5	72.9	82.3	91.7	95.2
	0.0	47.2	46.4	78.8	77.7	79.9	79.3	63.8	91.2	89.6	98.6
	0.5	87.1	86.8	92.1	91.7	93.4	93.0	50.6	95.0	81.8	98.1
400	-0.5	68.5	63.5	81.1	78.3	88.1	86.6	99.8	99.9	100.0	99.8
	0.0	99.2	98.0	98.6	98.1	98.9	98.4	99.2	100.0	100.0	100.0
	0.5	100.0	99.9	99.8	99.8	99.9	99.9	97.1	100.0	100.0	100.0

Notes: ADF$_{t.\tau}^{P}$ denotes Perron (1990) unit root test (known level shift date). ADF$_{t.\tau}^{P}$ denotes the same test but with GLS detrending instead of OLS detrending. ADF$_t^{PV}$ denotes Perron-Vogelsang (1992) unit root test (unknown level shift date). ADF$_t^{PR}$ denotes Perron-Rodriguez (2003) GLS-based unit root test (unknown level shift date). Local power is evaluated at α=1−c/T with c=7. See also Table 1.

Table 7. Empirical size of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}). GLS de-trended data

		Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}
100	−0.5	6.9	7.4	6.9	7.4	7.1	7.6	2.8	3.2	4.7	5.1	5.3	5.8	4.7	4.9	5.0	5.5	5.6	6.1
	0	7.0	7.5	7.1	7.6	7.3	7.8	6.0	6.3	7.1	7.4	6.9	7.6	7.0	7.3	7.3	7.9	7.4	8.0
	0.5	7.0	7.0	7.0	7.0	7.3	7.2	4.0	4.2	4.6	4.8	5.2	5.4	4.4	4.5	5.1	5.3	5.8	5.9
200	−0.5	5.8	6.2	5.6	6.1	6.0	6.3	2.0	2.1	4.7	5.0	5.7	6.1	3.7	4.1	4.6	4.9	5.2	5.6
	0	5.3	5.6	5.5	5.7	5.4	5.7	5.2	5.5	5.5	5.6	5.6	5.8	5.3	5.5	5.8	6.1	6.0	6.2
	0.5	5.8	5.8	5.7	5.8	5.9	5.9	3.3	3.3	5.1	5.3	5.8	5.8	3.2	3.4	5.0	5.3	5.6	5.8
400	−0.5	4.9	5.2	4.8	5.0	4.9	5.2	1.8	1.9	4.3	4.6	4.6	4.9	2.8	3.0	4.6	4.9	4.8	5.0
	0	5.0	5.1	5.0	5.2	5.0	5.2	4.1	4.4	4.8	5.1	5.0	5.2	4.7	5.0	5.0	5.3	5.1	5.4
	0.5	4.9	5.0	4.9	5.1	4.9	5.1	2.6	2.7	4.7	4.8	4.9	5.0	2.7	2.8	4.7	4.8	4.9	5.0

Notes: Asymptotic critical values at the 5% level as reported in Ng and Perron (2001) are employed. See also Table 1.

Table 8. Size-adjusted power of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}). GLS de-trended data.

		Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	49.6	50.5	47.2	47.7	47.4	48.3	4.9	5.3	12.0	12.5	13.3	13.6	11.9	12.0	16.1	15.7	16.8	16.3
	0	47.4	47.9	44.9	46.0	44.8	45.5	12.5	16.6	24.6	24.6	25.1	24.9	15.1	15.5	25.1	25.3	24.4	24.4
	0.5	38.2	38.9	37.4	38.0	37.5	38.2	27.7	28.2	32.5	33.4	32.5	33.1	26.9	27.4	31.4	31.7	31.2	31.7
200	−0.5	49.7	50.4	48.7	49.6	49.0	49.5	5.3	5.4	27.8	28.1	32.0	32.1	10.3	10.5	29.4	29.4	34.4	34.4
	0	51.2	51.2	50.3	50.3	50.2	50.6	12.1	12.5	42.6	42.9	42.2	42.3	17.6	17.9	40.5	41.3	40.2	40.9
	0.5	43, 4	43.7	42.9	43.4	42.8	43.4	30.5	30.9	39.0	39.0	39.4	39.8	29.9	29.6	40.6	40.6	41.0	41.0
400	−0.5	50.7	50.7	50.1	50.5	50.2	50.4	4.6	5.0	44.5	45.3	45.8	45.7	11.0	11.0	44.0	44.1	45.5	45.1
	0	49.9	50.3	49.6	49.7	49.5	49.6	14.0	13.9	48.9	48.5	48.2	48.2	17.2	17.0	47.0	46.6	46.9	47.0
	0.5	47.3	48.0	47.0	47.5	47.2	47.6	33.0	34.2	46.9	47.2	46.8	47.5	31.8	32.5	46.7	47.8	46.7	47.9

Notes: Power is evaluated at α=1-c/T with c=-13.5. See also Tables 1 and 8.

Table 9. Size and size-adjusted power of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}) under unknown lag order. Raw data.

		Size ($\alpha = 1$)								Size-adjusted power ($\alpha = 1 - c/T$)							
		t-significance				BIC				t-significance				BIC			
T	γ	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_{t}	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_{t}	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_{t}	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_{t}	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_{t}	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_{t}	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_{t}	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_{t}
100	−0.5	5.5	4.8	9.2	8.0	4.9	5.1	7.9	8.3	9.5	9.4	12.6	12.2	9.8	9.6	12.1	12.1
	0.0	5.9	5.0	7.7	6.7	5.0	5.3	6.5	6.8	17.9	18.1	18.9	19.3	19.5	19.8	20.4	20.4
	0.5	5.6	4.7	6.9	5.7	4.3	4.0	6.0	5.8	24.7	25.3	27.7	29.2	27.1	27.6	29.4	29.5
200	−0.5	5.7	5.3	6.7	6.5	4.5	4.9	5.9	5.9	19.3	19.6	25.7	25.8	20.8	20.2	27.2	27.3
	0.0	5.3	4.9	5.9	5.4	4.8	5.1	5.4	5.6	33.4	33.4	34.2	34.6	35.5	35.4	36.1	35.9
	0.5	5.3	4.6	5.7	5.3	4.8	4.7	5.6	5.4	36.6	38.1	38.0	38.6	39.0	39.4	40.7	41.4
400	−0.5	5.3	5.1	5.6	5.4	4.8	4.7	5.1	5.2	34.6	34.3	36.2	36.0	36.5	36.7	37.9	38.1
	0.0	5.4	5.2	5.6	5.3	5.0	5.1	5.1	5.3	41.4	41.8	42.1	42.0	43.9	44.0	44.2	44.2
	0.5	5.5	5.1	5.5	5.3	4.9	5.0	5.1	5.1	43.7	43.9	43.9	44.0	44.9	45.4	44.9	45.2

Notes: Power is evaluated at α=1-c/T with c=7. See also Table 1.

Table S.1. Comparison betweet test based on our methods and tests based on traditional level shift detection methods (in italics). Size and size-adjusted power against local and non-local alternatives

								traditional shift estimation methods			
Size ($\alpha = 1$)								Chen-Tiao		Chen–Liu	
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	3.0	3.2	3.9	4.0	6.5	6.9	*7.4*	*7.5*	*10.9*	*11.1*
100	0.0	4.9	4.9	5.5	5.3	6.9	6.7	*9.3*	*9.2*	*12.1*	*12.0*
100	0.5	3.6	3.7	4.0	3.8	5.5	5.4	*7.1*	*6.7*	*9.3*	*9.3*
200	−0.5	2.6	2.7	4.9	5.0	5.8	5.9	*5.6*	*5.7*	*7.5*	*7.4*
200	0.0	4.8	4.9	5.0	4.8	5.4	5.3	*6.4*	*6.4*	*10.2*	*10.1*
200	0.5	3.4	3.4	4.5	4.4	5.3	5.3	*6.5*	*6.3*	*8.3*	*8.2*
400	−0.5	2.5	2.7	4.8	4.8	5.1	5.1	*5.2*	*5.1*	*6.1*	*6.1*
400	0.0	4.6	4.7	5.3	5.3	5.5	5.6	*6.5*	*6.5*	*9.4*	*9.3*
400	0.5	3.5	3.5	5.2	5.3	5.4	5.4	*6.8*	*6.8*	*8.2*	*8.4*
								traditional shift estimation methods			
Size-adjusted power ($\alpha = 1 - c/T$)								Chen-Tiao		Chen–Liu	
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	0.0	0.0	10.1	9.6	12.8	12.9	*11.4*	*11.2*	*10.1*	*10.0*
100	0.0	0.1	0.2	19.1	19.1	19.4	20.1	*14.8*	*14.9*	*11.5*	*11.4*
100	0.5	11.0	10.9	27.5	27.9	30.0	30.6	*29.9*	*30.6*	*27.3*	*28.5*
200	−0.5	0.0	0.0	20.3	20.7	27.0	27.9	*24.1*	*23.9*	*22.8*	*23.5*
200	0.0	0.2	0.2	35.6	36.6	36.2	37.4	*28.8*	*28.4*	*19.4*	*20.4*
200	0.5	11.7	11.2	39.7	39.8	40.9	42.3	*38.3*	*38.7*	*35.1*	*36.0*
400	−0.5	0.0	0.0	38.1	38.0	39.3	39.7	*27.3*	*27.6*	*27.7*	*28.7*
400	0.0	0.1	0.1	42.4	42.5	42.1	42.5	*29.5*	*29.5*	*21.7*	*22.2*
400	0.5	10.6	10.4	43.6	44.0	44.1	45.0	*38.8*	*38.8*	*36.7*	*37.1*
								traditional shift estimation methods			
Size-adjusted power ($\alpha = 0.9$)								Chen-Tiao		Chen–Liu	
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	0.0	0.0	11.9	11.3	15.4	15.2	*14.2*	*14.1*	*14.1*	*14.1*
100	0.0	0.0	0.0	24.6	24.4	25.1	25.8	*19.8*	*20.1*	*16.0*	*16.0*
100	0.5	7.5	7.8	34.9	35.4	39.7	39.8	*40.2*	*41.0*	*36.9*	*36.9*
200	−0.5	0.0	0.0	38.7	38.4	56.3	56.0	*50.0*	*49.5*	*53.9*	*53.9*
200	0.0	0.0	0.0	72.2	71.8	74.7	74.5	*60.9*	*60.4*	*52.4*	*52.4*
200	0.5	3.5	3.4	83.4	83.0	89.1	88.9	*85.5*	*85.4*	*85.0*	*85.0*
400	−0.5	0.0	0.0	84.3	83.4	90.2	89.9	*63.7*	*63.3*	*70.4*	*70.4*
400	0.0	0.0	0.0	98.1	97.8	98.4	98.1	*74.8*	*73.7*	*70.3*	*70.3*
400	0.5	0.7	0.6	99.7	99.6	99.9	99.9	*97.0*	*96.8*	*98.3*	*98.3*

Table S.2. Comparison between tests based on continuous $\hat{\delta}$ and tests based on the binary estimator $\check{\delta}\,(0.5)$ =round($\hat{\delta}$)

Size ($\alpha = 1$)									
		continuous $\hat{\delta}$				$\check{\delta}$ =round($\hat{\delta}$)			
T	γ	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}
100	-0.5	3.9	4.0	6.5	6.9	3.6	3.8	5.3	5.3
100	0.0	5.5	5.3	6.9	6.7	5.2	5.3	6.2	6.3
100	0.5	4.0	3.8	5.5	5.4	4.0	3.9	5.1	5.0
200	-0.5	4.9	5.0	5.8	5.9	4.9	4.9	5.5	5.6
200	0.0	5.0	4.8	5.4	5.3	5.0	4.8	5.2	5.1
200	0.5	4.5	4.4	5.3	5.3	4.6	4.5	5.0	5.0
400	-0.5	4.8	4.8	5.1	5.1	5.0	4.9	5.0	4.9
400	0.0	5.3	5.3	5.5	5.6	5.1	5.0	5.3	5.2
400	0.5	5.2	5.3	5.4	5.4	4.8	4.8	4.9	4.9
Size-adjusted power ($\alpha = -7/T$)									
		continuous $\hat{\delta}$				$\check{\delta}$ =round($\hat{\delta}$)			
T	γ	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}
100	-0.5	10.1	9.6	12.8	12.9	10.1	10.0	11.6	12.0
100	0.0	19.1	19.1	19.4	20.1	19.2	19.3	19.5	19.4
100	0.5	27.5	27.9	30.0	30.6	25.7	25.5	27.9	28.4
200	-0.5	20.3	20.7	27.0	27.9	20.2	20.5	26.1	26.0
200	0.0	35.6	36.6	36.2	37.4	36.1	37.0	35.9	36.6
200	0.5	39.7	39.8	40.9	42.3	38.0	38.3	40.3	40.6
400	-0.5	38.1	38.0	39.3	39.7	35.9	36.0	37.0	37.3
400	0.0	42.4	42.5	42.1	42.5	42.1	43.2	41.4	42.7
400	0.5	43.6	44.0	44.1	45.0	46.1	46.5	46.4	46.4
Size-adjusted power ($\alpha = 0.9$)									
		continuous $\hat{\delta}$				$\check{\delta}$ =round($\hat{\delta}$)			
T	γ	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}
100	-0.5	11.9	11.3	15.4	15.2	11.6	11.2	11.2	13.8
100	0.0	24.6	24.4	25.1	25.8	23.4	23.2	23.2	23.9
100	0.5	34.9	35.4	39.7	39.8	33.3	33.1	33.1	37.6
200	-0.5	38.7	38.4	56.3	56.0	36.5	36.5	36.5	53.9
200	0.0	72.2	71.8	74.7	74.5	72.8	72.4	72.4	73.8
200	0.5	83.4	83.0	89.1	88.9	81.0	80.5	80.5	86.9
400	-0.5	84.3	83.4	90.2	89.9	85.4	84.5	84.5	88.5
400	0.0	98.1	97.8	98.4	98.1	97.3	97.0	97.0	97.3
400	0.5	99.7	99.6	99.9	99.9	99.6	99.5	99.5	99.7

Table S.3. Size and size-adjusted power of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}) with $\nu = 8$ and $\nu = \infty$. Model S_0 with standardized $t\,(10)$ innovations

Size				$\nu = 8$				$\nu = \infty$			
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	-0.5	5.2	5.2	5.1	5.2	5.3	5.4	5.1	5.0	6.5	6.6
100	0.0	5.0	5.1	5.1	5.0	5.1	5.2	5.3	5.2	6.3	6.3
100	0.5	5.4	5.2	5.4	5.2	5.5	5.4	5.4	5.3	6.2	6.0
200	-0.5	5.1	5.2	5.0	5.2	5.1	5.2	5.0	5.0	6.0	6.0
200	0.0	4.9	4.9	4.8	4.9	4.9	5.0	5.1	5.2	5.7	5.6
200	0.5	5.3	5.3	5.3	5.2	5.3	5.2	5.0	4.9	5.4	5.4
400	-0.5	5.0	5.0	5.1	5.0	5.2	5.1	5.2	5.0	5.9	5.9
400	0.0	5.0	4.8	5.0	4.8	5.0	4.8	5.0	5.1	5.3	5.4
400	0.5	5.1	4.8	5.1	4.8	5.1	4.8	4.9	4.9	5.2	5.1
Local power				$\nu = 8$				$\nu = \infty$			
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	-0.5	49.5	50.6	49.8	49.6	49.7	49.8	40.4	41.6	35.6	35.2
100	0.0	49.1	49.7	48.5	49.1	48.3	49.8	37.8	38.5	37.0	36.9
100	0.5	42.5	44.6	43.0	44.2	42.7	44.4	39.8	40.2	38.9	39.5
200	-0.5	49.0	49.2	49.1	48.9	48.9	49.3	40.9	40.6	36.1	36.6
200	0.0	49.7	49.6	49.5	49.0	49.5	49.1	39.7	39.1	39.3	39.6
200	0.5	45.9	45.7	46.1	46.0	46.1	46.1	44.4	45.2	44.3	44.1
400	-0.5	49.3	50.1	49.0	49.8	48.7	49.4	38.3	39.5	34.9	35.9
400	0.0	49.2	50.4	49.2	50.5	48.9	50.6	41.6	40.9	41.7	42.0
400	0.5	47.4	49.3	47.4	49.8	47.8	49.8	46.5	47.3	45.5	46.5

Table S.4. Size and size-adjusted power of standard ADF tests (ADF) and of the modified ADF tests (ADF^{δ} and ADF^{Ψ}) with $\nu = 8$ and $\nu = \infty$. Model S_4 with standardized $t\,(10)$ innovations

Size		$\nu = 8$						$\nu = \infty$			
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	2.5	2.8	2.7	3.0	3.1	3.4	4.2	4.3	7.2	7.2
100	0.0	4.6	4.8	4.8	4.8	5.0	5.0	5.2	5.2	7.1	7.0
100	0.5	3.6	3.6	3.8	3.9	4.0	4.1	4.1	3.9	5.7	5.3
200	−0.5	2.8	2.9	3.4	3.5	4.3	4.5	4.4	4.4	6.2	6.3
200	0.0	5.0	5.1	4.7	4.9	5.3	5.3	4.9	4.7	5.8	5.6
200	0.5	3.6	3.6	3.8	3.8	4.2	4.2	4.5	4.5	5.4	5.4
400	−0.5	2.4	2.5	4.1	4.2	5.2	5.2	5.1	5.0	6.0	6.0
400	0.0	4.6	4.8	5.4	5.6	5.7	5.8	5.0	5.1	5.5	5.5
400	0.5	3.4	3.4	4.4	4.5	5.0	4.9	4.6	4.8	5.0	5.0
Local power				$\nu = 8$				$\nu = \infty$			
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t	$ADF^{\delta}_{\hat{\alpha}}$	ADF^{δ}_t	$ADF^{\Psi}_{\hat{\alpha}}$	ADF^{Ψ}_t
100	−0.5	0.0	0.0	0.2	0.2	0.4	0.4	9.4	9.8	11.5	11.8
100	0.0	0.2	0.2	1.3	1.3	1.5	1.4	18.8	19.3	18.8	18.6
100	0.5	10.2	9.7	12.2	12.0	13.2	12.7	25.5	26.2	29.1	30.2
200	−0.5	0.0	0.0	1.7	1.8	3.9	4.0	20.3	20.6	23.3	23.5
200	0.0	0.2	0.2	9.7	9.2	9.6	9.8	30.9	32.0	31.3	31.4
200	0.5	10.1	9.7	17.5	18.3	20.8	21.4	38.5	38.5	39.7	39.5
400	−0.5	0.0	0.0	15.1	14.9	26.0	25.9	31.3	31.1	29.8	29.1
400	0.0	0.1	0.1	33.3	33.7	33.2	33.5	38.0	37.7	38.6	38.7
400	0.5	11.5	11.1	36.4	36.6	40.3	41.0	45.6	45.5	45.8	45.7

Table S.5. Size and size-adjusted power of standard ADF tests (ADF) and of the modiﬁed ADF tests (ADF^{δ} and ADF^{Ψ}) with $\nu = 8$ and $\nu = \infty$. Model S_r with standardized $t\,(10)$ innovations

Size				$\nu = 8$				$\nu = \infty$			
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}
100	−0.5	3.9	3.9	3.9	3.9	4.2	4.1	3.9	3.9	6.6	6.8
100	0.0	5.2	5.2	5.4	5.4	5.7	5.7	5.3	5.4	7.1	7.0
100	0.5	3.8	3.7	3.8	3.7	4.0	3.9	4.2	4.2	6.0	5.8
200	−0.5	3.8	3.7	4.0	4.0	4.7	4.7	4.3	4.4	6.1	6.3
200	0.0	5.2	5.3	5.3	5.4	5.7	5.7	5.1	5.1	5.9	6.0
200	0.5	3.9	3.9	3.9	3.8	4.3	4.2	4.9	4.8	5.6	5.6
400	−0.5	4.1	4.0	4.4	4.4	5.5	5.5	5.1	5.2	5.9	5.8
400	0.0	4.8	4.8	4.8	4.7	5.0	5.0	5.0	5.0	5.3	5.3
400	0.5	3.3	3.3	4.3	4.3	4.9	4.8	4.7	4.7	5.1	5.0
Local power				$\nu = 8$				$\nu = \infty$			
T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}	$ADF_{\hat{\alpha}}^{\delta}$	ADF_t^{δ}	$ADF_{\hat{\alpha}}^{\Psi}$	ADF_t^{Ψ}
100	−0.5	9.0	9.1	9.5	9.6	9.7	9.8	12.7	12.6	13.6	13.6
100	0.0	13.3	13.2	13.4	13.4	13.5	13.3	18.9	18.6	20.0	20.1
100	0.5	23.1	23.1	24.3	25.0	24.7	25.4	27.4	27.7	28.4	29.0
200	−0.5	8.8	8.9	10.0	10.4	11.4	11.6	22.5	22.0	24.4	24.2
200	0.0	12.6	12.2	17.5	17.5	17.7	17.7	31.3	31.7	31.5	32.0
200	0.5	22.1	22.8	25.3	25.2	26.5	26.9	35.6	36.3	38.2	38.3
400	−0.5	8.3	8.3	17.7	17.6	24.9	25.1	30.8	30.2	29.4	29.2
400	0.0	13.8	14.0	36.3	36.2	36.0	35.9	38.2	38.4	39.0	39.0
400	0.5	24.1	24.4	36.5	37.0	40.7	41.0	45.3	45.8	45.1	45.7

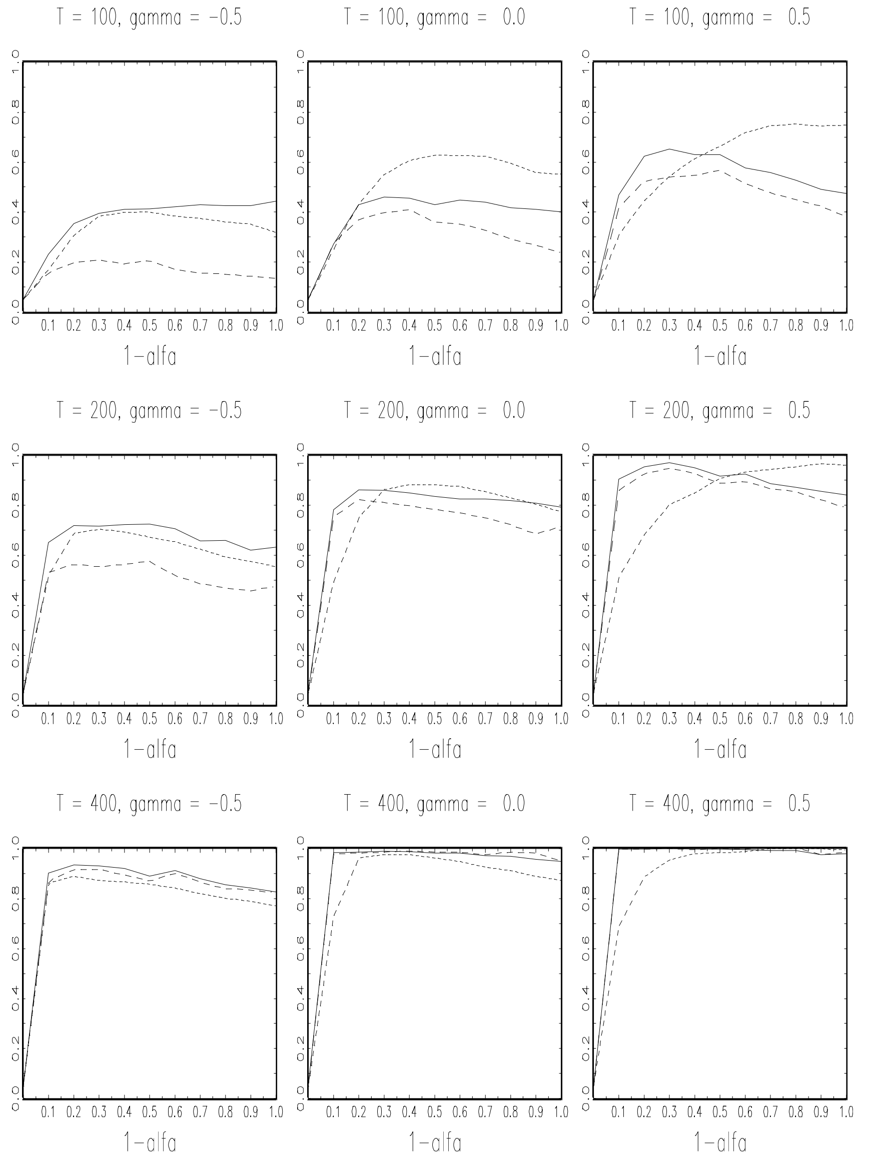

Figure 1: Size adjusted power of the basic de-jumping ADF_α^δ test (long dashed line), of the finer de-jumping ADF_α^Ψ test (thick line) and of the ADF_α test based on de-jumping and Bai-Perron (1998) break date estimator (short dashed line).