

Testing for unit roots in autoregressions with multiple level shifts

Giuseppe Cavaliere*
University of Bologna
Department of Statistical Sciences

Iliyan Georgiev
Universidade Nova de Lisboa
Faculdade de Economia

May 2005

Abstract

The asymptotic distributions of Augmented-Dickey-Fuller (ADF) unit root test statistics for autoregressive processes with a unit or near-unit root are derived in the presence of multiple stochastic level shifts of large size. The distributions depend on a Brownian motion and a Poisson-type jump process. The latter is related to the occurrence of level shifts, and induces for tests based on standard critical values power losses increasing rapidly with the number and the magnitude of the shifts. A new approach to unit root testing is suggested which does not require to know either the location or the number of level shifts. It is proposed to remove possible shifts from a time series by weighting each of its increments according to how likely it is, with respect to an ad hoc postulated distribution, a shift to have occurred in the respective period. Hence, no explicit decision rule on the occurrence of shifts is required. It is proved that if the number of level shifts is bounded in probability, the proposed test statistics have pivotal limiting distributions coinciding with those of the corresponding ADF statistics under standard conditions. A Monte Carlo experiment shows that, despite their generality, the new tests perform well in small samples.

Keywords: Unit roots, level shifts, 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. Many researchers have focused on whether economic time series have unit roots or, conversely, are stationary around a single level shift. 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

*Address for Correspondence: Giuseppe Cavaliere, Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, I-40126 Bologna, Italy. E-mail cavaliere@stat.unibo.it

dates have been initially proposed by Banerjee et al. (1992), Zivot and Andrews (1992) and subsequently by Leybourne et al. (1998), Saikkonen and Lütkepohl (2002), Lütkepohl et al. (2004), among others.

Few attempts have been made to robustify unit root tests in the presence of multiple level shifts. Lumsdaine and Papell (1997) generalize the tests proposed by Banerjee et al. (1992) by allowing for two level shifts at unknown dates. Unfortunately, although there is little justification for fixing the number of shifts to one or two *a priori* (cf. Lumsdaine and Papell, 1997, p.218), their procedure can hardly be generalized to a larger number of shifts. 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 reference critical values depend on the shift dates and the power of the tests decline as the maximum number of allowed shifts 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 number 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) although in the basic setup shifts are exogenous, forms of dependence with the shocks driving the ordinary autoregressive dynamics can be allowed without affecting the results; (v) shift sizes are random and of larger magnitude order than the shocks driving the autoregressive dynamics. Note that assumptions (i)–(v) are quite general and differ in several respects from what has been previously studied in the literature. For instance, they do not require from the investigator 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 to infinity as the number of observations increases (cf. Balke and Fomby, 1991a, 1991b; Franses and Haldrup, 1994; Nelson et al., 2001), but we adopt it since it preserves in the limit the distinction between ordinary shocks and level shifts.

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 standard 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 under 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^δ . In order to remove the shifts, for each observation we suggest to compute a pseudo-probability $\tilde{\delta}_t$ (with respect to an ad hoc probability measure) that a level shift has occurred at time t , given the data. Then, shifts are removed by defining the ‘de-jumped’ time series \tilde{X}_t^δ as $\tilde{X}_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s \Delta X_s$. The unit root hypothesis is tested by running standard ADF tests on \tilde{X}_t^δ , which are found to have the same (pivotal) limiting distributions as standard ADF tests under no level shifts – a property which holds both under the null hypothesis and under local alternatives.

The idea of weighting each observation with a (pseudo) probability is present in the work of Franses and Lucas (1998) on the robustification of likelihood-based 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 decision rules for consistent estimation of shift dates are feasible.

To evaluate the relevance of the asymptotic theory developed here, the agreement of its predictions with well-known finite-sample evidence is worth to be noted. Specifically, we show that if level shifts are neglected in testing, the asymptotic local power function of the ADF tests can be far below the power function under no level shifts. This 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 the same as the object which is actually computed - an obvious requirement that may sometimes be hard to check.¹

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 we 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. A basic version of the tests is reported in Section 3.2, while a refined version with better finite sample properties is discussed in Section 3.3. The small sample properties of the new tests are analyzed through a set of Monte Carlo simulations in Section 4. Section 5 concludes. All proofs are collected in the Appendix. The following notation is used: ‘ \xrightarrow{w} ’ denotes weak convergence and ‘ \xrightarrow{P} ’ convergence in P -probability respectively; $\mathbb{I}(\cdot)$ 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 $[\cdot]$ signifies the largest integer not greater than its argument. With $\mathcal{D}[0,1]$ we denote the space of cadlag functions on $[0,1]$, endowed with the Skorohod topology. Finally, for a scalar sequence Z_t we define the related sequences $\nabla \mathbf{Z}_t := (\Delta Z_t, \dots, \Delta Z_{t-k+1})'$ and $\mathbf{Z}_t := (Z_t, (\nabla \mathbf{Z}_t)')$.

2 Model and preliminary results

We are interested in testing the unit root null hypothesis $H_0 : \alpha = 1$ against local alternatives $H_c : \alpha = 1 - c/T$ in the following additive level-shift model for the observable variable X_t :

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

where Y_t is an unobservable autoregressive process, Z_t is a vector of deterministic terms (e.g. a constant and a linear trend) and μ_t denotes an unobservable level-shift component. The

¹For example, numerical techniques may deliver a local maximum of a criterion function, whereas the asymptotic analysis is carried out for a global one.

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) ε_t are IID($0, \sigma_\varepsilon^2$) and $E|\varepsilon_1|^r < \infty$ for some $r \geq 4$.

The assumption prevents Y_t from being I(2) or seasonally integrated, and ensures that the so-called long-run variance, $\sigma^2 := \sigma_\varepsilon^2 \Gamma(1)^{-2}$ hereafter, is well-defined. The novelty of the paper lies in the way μ_t is specified and dealt with. To focus the exposition on this aspect, we regard the lag order k as known, and in the entire analytical part we consider the special case $\varphi = 0$. De-meaning and de-trending will be addressed in Section 4. Pre-sample values are set to $\mu_{-k}, \dots, \mu_0 = 0$, $Y_{-k} = O_P(1)$, and $(u_{-k+1}, \dots, u_0)'$ is given the stationary distribution induced by the difference equation $u_t = \sum_{i=1}^k \gamma_i u_{t-i} + \varepsilon_t$.²

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 $\gamma := (\gamma_1, \dots, \gamma_k)'$ 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 test statistics $ADF_{\hat{\alpha}} := T(\hat{\alpha} - 1) / \hat{\Gamma}(1)$ and $ADF_t := (\hat{\alpha} - 1) / s(\hat{\alpha})$, where $\hat{\Gamma}(1) := 1 - \mathbf{1}'_k \hat{\gamma}$ and $s(\hat{\alpha})$ is the (OLS) standard error of $\hat{\alpha}$. It is known (see, e.g., Chang and Park, 2002) that under Assumption \mathcal{M} and for $\alpha = 1 - c/T$, $\hat{\alpha}$ and $\hat{\gamma}$ are consistent and, as $T \rightarrow \infty$,

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

where $B_c(s) := \int_0^s e^{-c(s-z)} dB(z)$ is an Ornstein-Uhlenbeck process on $[0, 1]$, 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 so-called univariate Dickey-Fuller distributions.

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_1 at time t^* (cf. Perron, 1989; Amsler and Lee, 1995; Saikkonen and Lütkepohl, 2002, p.316). This model corresponds to μ_t being generated as

$$\mu_t := \mu_0 + \theta^* \mathbb{I}(t \geq t^*), \quad \theta^* := \mu_1 - \mu_0, \quad (3)$$

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

$$\mu_t := \sum_{s=1}^t \delta_s \theta^*, \quad \delta_s := \mathbb{I}(s = t^*), \quad (4)$$

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 a rather general unobservable sequence of binary variables indicating the occurrences of shifts, and $\{\theta_t\}$ is the (random) sequence of shift sizes. 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} above.

²The results of the paper remain valid under the more general specification $u_0, \dots, u_{-k+1} = O_P(1)$.

Assumption \mathcal{S} . (a) N_T is bounded in probability, unconditionally and conditionally on $N_T \geq 1$; (b) $\theta_t = T^{1/2}\eta_t$, where η_t and η_t^{-1} are $O_P(1)$ sequences; (c) δ_t is independent of Y_{-k} , $(u_{1-k}, \dots, u_0)'$ and (ε_s, η_s) for all $t, s \leq T$.

Several points are worth to note.

Remark 2.1. Formally, since (δ_t, θ_t) (as well as the α of (1)) depend on T , we are considering a triangular array of the form $\{X_{T,t} = Y_{T,t} + \mu_{T,t}; t = -k, -k+1, \dots, T, T = 1, 2, \dots\}$. Unless differently specified, in order to keep notation simple, we drop the ‘ T ’ subscript.

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.

Remark 2.3. Our model can equivalently be expressed in the multiple-break format (see e.g. Bai and Perron, 1998)

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

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

Remark 2.4. Assumption $\mathcal{S}(b)$ relates explicitly the magnitude of the level shifts to the sample size; this assumption is not new in the literature, cf. e.g. Leybourne and Newbold (2000a,2000b), Doornik et al. (1998) and Perron (1989, p. 1372). Specifically, the $T^{1/2}$ scale for the size of the level shifts – which is the natural choice as it corresponds to the magnitude order of the unconditional standard deviation of Y_t under the parametrization $\alpha = 1 - c/T$ – 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. Note also that the same rate has been used by Müller and Elliott (2003) to model the initial observation of processes with unit roots near unity.

Remark 2.5. 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 if P_1 is P conditional on the occurrence of at least one level shift, it holds that $\max_{t \leq T} |\delta_t \varepsilon_t| =: \max_{t: \delta_t=1} |\varepsilon_t|$, $\max_{t: \delta_t=1} |\Delta Y_t|$, $\max_{t: \delta_t=1} |\eta_t|$ and $\max_{t: \delta_t=1} |\eta_t^{-1}|$ are bounded in P and P_1 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, all the results given in this paper continue to hold.³

Remark 2.6. Assumption \mathcal{S} imposes no restrictions on the dependence structure of δ_t . In particular, since δ_t does not need to be independent over time, it allows for clusters of level shifts.

³To see that some kinds of dependence between δ_t and (ε_s, η_s) violate boundedness, consider the extreme case $\delta_t := \mathbb{I}\{|\varepsilon_t| = \max_{s \leq T} |\varepsilon_s|\}$, where $N_T = 1$ a.s. but $\max_{t: \delta_t=1} |\varepsilon_t| = \max_{t \leq T} |\varepsilon_t|$ is, in general, not bounded in probability.

In the presence of level shifts satisfying Assumption \mathcal{S} , the asymptotic distributions of the ADF statistics provided in (2) is no longer valid. Under the complete generality of the assumption, the test statistics need not even to have weak limits. To be able to describe their asymptotic behavior in terms of functionals of a stochastic process, we specialize to the case of $\eta_t = \eta$ and δ_t IID with $P(\delta_t = 1) = \lambda/T$, the last in order to bound the number of level shifts N_T as T diverges. Asymptotics for the ADF statistics are given in the next theorem, where $\mathcal{P}_\lambda(\cdot)$ denotes a Poisson process with intensity λ . The process $\mathcal{P}_\lambda(\cdot)$ is defined on $[0,1]$ and is constant apart from finitely many jumps equal to unity.

Theorem 1 *Under Assumptions \mathcal{M} and \mathcal{S} , if $\alpha = 1 - c/T$ with $c \geq 0$, $\eta_t = \eta$ and δ_t are IID with $P(\delta_t = 1) = \lambda/T$, then, as $T \rightarrow \infty$:*

$$\frac{1}{T^{1/2}} X_{[Ts]} \xrightarrow{w} \sigma H_c(s), \quad H_c(s) := B_c(s) + \frac{\eta}{\sigma} \mathcal{P}_\lambda(s), \quad s \in [0,1], \quad (5)$$

where convergence is on $\mathcal{D}[0,1]$, and $B_c(\cdot)$ is an Ornstein-Uhlenbeck process independent of $\mathcal{P}_\lambda(\cdot)$. Moreover,

$$ADF_{\hat{\alpha}} \xrightarrow{w} \frac{\int_0^1 H_c(s) dH_c(s) + \omega_0}{\int_0^1 H_c(s)^2 ds}, \quad ADF_t \xrightarrow{w} \frac{\int_0^1 H_c(s) dH_c(s) + \omega_0}{(\omega_1 \int_0^1 H_c(s)^2 ds)^{1/2}}, \quad (6)$$

where ω_0 and ω_1 depend on $\eta^2 \mathcal{P}_\lambda(1) / \sigma^2$, and equal respectively 0 and 1 when $\mathcal{P}_\lambda(1) = 0$. In the special case $k = 0$, $\omega_0 = 0$ and $\omega_1 = 1 + \eta^2 \mathcal{P}_\lambda(1) / \sigma_\varepsilon^2$.⁴

Remark 2.7. The most prominent feature of Theorem 1 is that, in the presence of level shifts, the limiting distributions of the ADF statistics depend on a discontinuous process. This distinguishes them from the usual Dickey-Fuller distributions (2) arising in the case of no level shifts. The difference in the asymptotic distributions 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 4 illustrate this aspect.

Remark 2.8. Intuitively, distributions (6) obtain because in the ADF regression: (i) X_{t-1} appears as a regressor instead of Y_{t-1} , and hence, in (6) $H_c(\cdot)$ appears as integrand instead of $B_c(\cdot)$; (ii) X_t appears on the left-side of the ADF regression instead of Y_t , and hence, in (6) $H_c(\cdot)$ appears as integrator instead of $B_c(\cdot)$; (iii) the lagged differences of X_t appear as regressors instead of those of Y_t , which leads to inconsistent estimation of γ , and gives rise to the term ω_0 in (6). The term ω_1 is due to the inconsistent estimation of both σ_ε^2 and γ .

Remark 2.9. To elaborate on the terms ω_0 and ω_1 , consider instead of (1) the following ‘innovational’ specification: $X_t = Y_t$, $Y_t = \alpha Y_{t-1} + u_t$ and $u_t = \sum_{i=1}^k \gamma_i u_{t-1} + \varepsilon_t + \delta_t \theta_t$. Here $\delta_t \theta_t$ generate level shifts in X_t under the null hypothesis $\alpha = 1$. A crucial difference from (1) is that ε_t and $\delta_t \theta_t$ feed into X_t via the same dynamics. As a consequence, γ is estimated consistently, and the asymptotic distributions of the ADF statistics could be shown to be given for any k by (6) with $\omega_0 = 0$ and $\omega_1 = 1 + \eta^2 \mathcal{P}_\lambda(1) / \sigma_\varepsilon^2$. In this case the term ω_1 is entirely due to the inconsistent estimation of σ_ε^2 . The difference from the general expressions

⁴The expressions for ω_0 and ω_1 for arbitrary k are given in the Appendix.

(6) sheds light on why additive and innovational shifts have different effects on finite-sample testing (see, e.g., Nielsen, 2004).

Remark 2.10. The importance of the Poisson component in the limiting distributions of Theorem 1 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 occurrence probability of level shifts. When the level shifts are small in size, that is η/σ is close to zero, the asymptotic distributions of Theorem 1 are close to the usual Dickey-Fuller type distributions. Conversely, when the shifts are large, conditionally on $N_T > 0$ (i.e., at least one level shift occurs), the Poisson component becomes the dominant one.

Remark 2.11. In Theorem 1 it is imposed that the possible level shifts all have the same size η . A more general assumption, which still allows to obtain asymptotic distributions in terms of functionals of a stochastic process, is that the shift sizes η_t are IID. In this case the results of the theorem hold provided that $\eta\mathcal{P}_\lambda(\cdot)$ is replaced by the *compound* Poisson process $\mathcal{C}_\lambda(\cdot) := \sum_{i=1}^{\mathcal{P}_\lambda(\cdot)} \eta_i$ (if $\eta_t = \eta$ for all t , $\mathcal{C}_\lambda(\cdot) = \eta\mathcal{P}_\lambda(\cdot)$), and $\eta^2\mathcal{P}_\lambda(1)$ is replaced by $\sum_{i=1}^{\mathcal{P}_\lambda(\cdot)} \eta_i^2$, the quadratic variation of $\mathcal{C}_\lambda(\cdot)$ on $[0, 1]$.

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 test statistics based on the observed time series X_t do not have the usual asymptotic distributions (2), in this section we propose variants of these statistics which *do* have (2) as their asymptotic distributions. As a consequence, two powerful results hold: on the one hand, standard tables of asymptotic critical values (see e.g. Fuller, 1976, for the case of no deterministic or OLS-based detrended data; Ng and Perron, 2001, for pseudo-GLS detrended data⁵) can be used for testing the unit root null hypothesis; on the other hand, whatever the number and the size of the shifts are, the new tests have the same asymptotic power function as the corresponding ADF tests in the standard, no-shift case.

Recall that, under Assumption \mathcal{M} , the asymptotic distributions of the ADF statistics from the regression $Y_t = \alpha Y_{t-1} + \gamma' \nabla \mathbf{Y}_{t-1} + error_t$ are as given in (2), where $Y_t = X_t - \mu_t$ can be thought of as obtained from X_t by removing the level shifts. Since μ_t and Y_t are unobservable, this regression is not feasible empirically. Instead, 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 the deterministic component, including possible (deterministic) level shifts. Since in our case μ_t is a random jump process, our procedure will be referred to as ‘de-jumping’ in what follows.

If the δ_t 's were observable, de-jumping could be based on approximating μ_t by $\hat{\mu}_t := \sum_{s=1}^t \delta_s \Delta X_s = \mu_t + \sum_{s=1}^t \delta_s \Delta Y_s$. The approximation error $\hat{\mu}_t - \mu_t$ is bounded in probability according to Lemma 7 in the Appendix, while μ_t has stochastic order of magnitude $T^{1/2}$. This difference turns out to be sufficient for the ADF statistics based on the de-jumped process $X_t^\delta := X_t - \sum_{s=1}^t \delta_s \Delta X_s$ to have asymptotic distributions (2). In the case of unobservable δ_t we imitate the above de-jumping procedure by estimating δ_t . The starting point is to recognize that the distribution of ΔX_t is a mixture, with mixing variable δ_t and mixture

⁵For the latter case the conclusion is based exclusively on Monte Carlo evidence.

components⁶ the distributions of ΔY_t ($\delta_t = 0$) and $\Delta Y_t + \theta_t$ ($\delta_t = 1$). 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’ (cf. Titterton et al., 1985) and consistent classification is possible.

Our proposed estimator $\tilde{\delta}_t$ of δ_t is the conditional expectation of δ_t with respect to an ad hoc postulated distribution. This approach does not require an explicit decision rule about when level shifts have occurred.⁷ Instead, each observation is weighted according to how likely it is, given the data and the postulated distribution, a level shift to have occurred in the corresponding period.

Two estimators of δ_t are discussed in the following subsections, leading to ADF-type tests with identical asymptotic distributions. Both tests rely on an ADF regression for a de-jumped series $\tilde{X}_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s \Delta X_s$. The first test uses an estimator of δ_t based on the observable increments $\Delta X_t = \Delta Y_t + \theta_t \delta_t$. The second one exploits the fact that ε_t and $\theta_t \delta_t$ are better separated than ΔY_t and $\theta_t \delta_t$ (since ε_t has smaller variance than ΔY_t), and rests on the joint estimation of δ_t and $\varepsilon_t + \theta_t \delta_t$; furthermore, it needs as an input the estimator of δ_t based on ΔX_t . The small sample properties of both tests are discussed in Section 4.

Some notation related to de-jumping is now introduced. For a vector $d := (d_1, \dots, d_T)'$, the process X_t de-jumped with weights d is denoted by $X_t^d := X_t - \sum_{s=1}^t d_s \Delta X_s$, $t \geq 1$, and $X_t^d := X_t$, $t = -k, \dots, 0$. According to the notation introduced in Section 1, we let $\nabla \mathbf{X}_t^d := (\Delta X_t^d, \dots, \Delta X_t^d)'$ and $\mathbf{X}_t^d := (X_t^d, (\nabla \mathbf{X}_t^d)')$. The processes \mathbf{X}_t^d (as well as \mathbf{Y}_t) sometimes appear normalized by the diagonal $(k+1) \times (k+1)$ matrix $A_T = \text{diag}(T^{-1/2}, 1, \dots, 1)$.

3.2 Rough de-jumping

3.2.1 Definitions of estimators and unit root tests

The definitions are inspired by the random level shifts specification from Theorem 1 and Remark 2.11; nevertheless, all subsequent results are obtained in the more general setup of Assumption \mathcal{S} . For the purposes of the definition, pretend that the ΔX_t ’s are independent draws from a mixture of a Student $t(\nu)$ distribution re-scaled with some $\sigma > 0$, and another $t(\nu)$ distribution re-scaled with $(\sigma^2 + T\theta^2)^{1/2}$, with mixture weights $1 - \lambda/T$ and λ/T respectively, in agreement with Assumption \mathcal{S} (a). The quasi-likelihood function for λ , regarding σ^2 and θ^2 as known, is

$$\prod_{t=1}^T \left(\frac{\lambda}{T} \phi_\nu(\Delta X_t; \sigma^2 + T\theta^2) + \left(1 - \frac{\lambda}{T}\right) \phi_\nu(\Delta X_t; \sigma^2) \right), \quad (7)$$

where $\phi_\nu(e; a^2)$ is the density of $a \cdot t(\nu)$. The QML estimator of λ is easily seen to satisfy the equation

$$\lambda = \sum_{t=1}^T \tilde{\delta}_t(\zeta), \quad \zeta := (\lambda, \sigma^2, \theta^2)', \quad (8)$$

⁶Although 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).

⁷Nevertheless, as it will be shown later, our approach allows for consistent estimation of shift dates as a by-product.

with

$$\tilde{\delta}_t(\zeta) := \frac{\lambda}{T} \left[\frac{\lambda}{T} + \left(1 - \frac{\lambda}{T}\right) \frac{\phi_\nu(\Delta X_t; \sigma^2)}{\phi_\nu(\Delta X_t; \sigma^2 + T\theta^2)} \right]^{-1}, \quad (9)$$

the latter corresponding to the expectation of δ_t conditional on the data; that is, $\tilde{\delta}_t(\zeta) = E(\delta_t | \{\Delta X_s\}_{s=1}^T) = E(\delta_t | \Delta X_t)$, with $E(\cdot | \{\Delta X_s\}_{s=1}^T)$ denoting expectation under the adopted pseudo distribution of $\{\Delta X_s\}_{s=1}^T$ and conditional on the data.

From (8) it follows that, for known σ^2 and θ^2 , the QML estimator of λ is a *fixed point* of the mapping $\lambda \rightarrow \sum_{t=1}^T \tilde{\delta}_t(\zeta)$. Such a fixed point can sometimes be computed by iterating the mapping until convergence; this is a simple instance of the Expectation-Maximization algorithm (Dempster et al., 1977), with $\tilde{\delta}_t(\zeta)$ computed at the expectation step, and their sum λ at the maximization step. Upon convergence, the estimate of λ can be inserted into (9), thus yielding the estimates of δ_t needed to eliminate the level shifts from the original time series.

In general $\tilde{\delta}_t(\zeta)$ is not the true conditional expectation of δ_t for any choice of ζ , but λ , θ^2 and σ^2 can still be thought of as related to the number of level shifts N_T and to the squared magnitude of the draws from the mixture components, $T^{-1} \sum_{t=1}^T \delta_t \Delta X_t^2$ and $T^{-1} \sum_{t=1}^T (1 - \delta_t) \Delta X_t^2$. Thus, to obtain an estimator of δ_t based on (9), we insert for ζ an estimator of $(N_T, T^{-1} \sum_{t=1}^T \delta_t \Delta X_t^2, T^{-1} \sum_{t=1}^T (1 - \delta_t) \Delta X_t^2)$. Namely, we define the random mapping $\Phi_T(\zeta) = (\Phi_T^\lambda(\zeta), \Phi_T^\theta(\zeta), \Phi_T^\sigma(\zeta))'$ whose components are given as follows:

$$\Phi_T^\lambda(\zeta) = \sum_{t=1}^T \tilde{\delta}_t(\zeta), \quad \Phi_T^\theta(\zeta) = T^{-1} \sum_{t=1}^T \tilde{\delta}_t(\zeta) \Delta X_t^2, \quad \Phi_T^\sigma(\zeta) = T^{-1} \sum_{t=1}^T (1 - \tilde{\delta}_t(\zeta)) \Delta X_t^2, \quad (10)$$

and then insert into (9) a random fixed point $\tilde{\zeta}_T$ of this mapping.

Once $\tilde{\zeta}_T$ has been found, level shifts are eliminated from X_t by computing the de-jumped time series $\tilde{X}_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s(\tilde{\zeta}_T) \Delta X_s$. Then, the unit root hypothesis can be tested by running standard ADF tests, henceforth ADF_α^δ and ADF_t^δ , based on the auxiliary ADF regression $\tilde{X}_t^\delta = \tilde{\alpha} \tilde{X}_{t-1}^\delta + \tilde{\gamma}' \nabla \tilde{\mathbf{X}}_{t-1}^\delta + error_t$. The logic behind this approach is that if $\tilde{\delta}_t(\tilde{\zeta}_T)$ estimate δ_t consistently at an appropriate rate, then \tilde{X}_t^δ will be sufficiently close to Y_t for the ADF statistics based on \tilde{X}_t^δ to be close to those based on the unobservable Y_t , the latter having the usual Dickey-Fuller type distribution. The asymptotic properties of this procedure are discussed in the next subsection.

Remark 3.1. The definition of Φ_T in (10) is not the only possible one, but among those capturing the meaning of ζ , it is a rather simple one. A natural alternative is to maximize the quasi-likelihood function (7) with respect to ζ , which gives for θ^2 and σ^2 the equations $\theta^2 = \frac{\nu+1}{\nu} \Phi_T^\theta(\zeta) - T^{-1} \sigma^2$ and $\sigma^2 = \frac{\nu+1}{\nu} \Phi_T^\sigma(\zeta)$. Ours obtain by dropping $\frac{\nu+1}{\nu}$ and $T^{-1} \sigma^2$.

3.2.2 Asymptotic results

We start this section with a theorem about existence, uniqueness and asymptotics for a fixed point $\tilde{\zeta}_T$ of the mapping Φ_T , such that $\tilde{\delta}_t(\tilde{\zeta}_T)$ are consistent for δ_t uniformly in t .⁸ The theorem also ensures that $\tilde{\zeta}_T$ can be computed by iterating Φ_T until convergence.

⁸It is immediate that $\Phi_T(\zeta)$ has a 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_1 be the probability measure induced by model (1) with $\alpha = 1 - c/T$, $c \geq 0$, under Assumptions \mathcal{M} and \mathcal{S} , and conditional on the realization of at least one level shift. If ν of (9) satisfies $3 \leq \nu \leq r - 1$, the following facts can be established:

EXISTENCE. There exists a random sequence $\tilde{\zeta}_T$ such that $P_1(\Phi_T(\tilde{\zeta}_T) = \tilde{\zeta}_T) \rightarrow 1$, $\tilde{\zeta}_T = O_{P_1}(1)$ and $\tilde{\zeta}_T$ is component-wise bounded away from zero in P_1 -probability.

UNIQUENESS. If $\tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$ both have the above properties, then $P_1(\tilde{\zeta}_T^1 = \tilde{\zeta}_T^2) \rightarrow 1$.

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 $P_1(\zeta_{Ti} \rightarrow_{i \rightarrow \infty} \tilde{\zeta}_T) \rightarrow 1$ as $T \rightarrow \infty$.

CONSISTENCY. Let $H_T := \sum_{t=1}^T \delta_t \eta_t^2$, and let σ_u^2 denote the variance of u_t . Then $\sum_{t=1}^T |\delta_t - \tilde{\delta}_t(\tilde{\zeta}_T)| = O_P(T^{-1/2})$, $\sum_{t=1}^T \delta_t |1 - \tilde{\delta}_t(\tilde{\zeta}_T)| = O_{P_1}(T^{-(\nu-2)/2})$ and $\tilde{\zeta}_T = (N_T, H_T, \sigma_u^2) + o_{P_1}(1)$.

Thus, $\tilde{\zeta}_T$ is consistent for (N_T, H_T, σ_u^2) , and level shifts are detected consistently in the sense that $\tilde{\delta}_t(\tilde{\zeta}_T)$ are consistent for δ_t at the uniform rate of $T^{1/2}$.

Remark 3.2. According to the consistency part of Theorem 2, it holds that $(\tilde{\zeta}_T^\lambda, \tilde{\zeta}_T^\theta) = (N_T, H_T) + o_{P_1}(1)$. This introduces a small complication into the argument for existence of $\tilde{\zeta}_T$, because in general no compact in \mathbb{R}^3 contains N_T and H_T (and hence, $(\tilde{\zeta}_T^\lambda, \tilde{\zeta}_T^\theta)$) with probability approaching one, while existence theorems for random fixed points are typically formulated for mappings from a compact onto itself. One consequence is that, instead of existence, we only establish existence with probability approaching one. This is not unusual in econometrics, see e.g. Saikkonen (2001).

Remark 3.3. A further consequence is that in order to prove the theorem we resort to the trick of looking for a fixed point of the form $\tilde{\zeta}_T = \zeta(\tilde{z}_T)$, where ζ is the random function $\zeta(z) = (N_T + z^\lambda, H_T z^\theta, z^\sigma)'$ and $(\tilde{z}_T^\lambda, \tilde{z}_T^\theta) = (0, 1) + o_{P_1}(1)$. For $H_T \neq 0$ the inverse function ζ^{-1} is well-defined, and \tilde{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.

Before introducing in Theorem 4 our main result, we present an important lemma which forms the basis of the proofs of both Theorems 2 and 4. The lemma contains several convergence statements uniform on compacts. One regards the rates of consistent detection of level shifts; another one, the distance between the OLS estimators of α and γ from ADF regressions for X_t^δ and for the unobserved Y_t ; yet another statement is an evaluation of the Jacobian of Φ_T that underlies contraction arguments. In the lemma, for compacts \mathbb{A}_T in \mathbb{R}^3 and 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 3 Let P_1 be as in Theorem 2, $\zeta(z)$ be either the identity on \mathbb{R}^3 or the random function $(N_T + z^\lambda, H_T z^\theta, z^\sigma)'$, and the compact $\mathbb{A}_T \subset \mathbb{R}^3$ be such that $\max_{z \in \mathbb{A}_T} \zeta(z) = O_{P_1}(1)$ and the components of $\min_{z \in \mathbb{A}_T} \zeta(z)$ are bounded away from zero in P_1 -probability. Define $X_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s(\zeta) \Delta X_s$, and let \mathbf{X}_t^δ be defined accordingly. If ν of (9) satisfies $3 \leq \nu \leq r - 1$, then:

- $\sup_{\mathbb{A}_T} \sum_{t=1}^T |\delta_t - \tilde{\delta}_t(\zeta)| \leq O_{P_1}(T^{-1/2})$ and $\sup_{\mathbb{A}_T} \sum_{t=1}^T \delta_t |1 - \tilde{\delta}_t(\zeta)| \leq O_{P_1}(T^{-(\nu-2)/2})$;
- $\sup_{\mathbb{A}_T} (|N_T - \Phi_T^\lambda(\zeta)|, |H_T - \Phi_T^\theta(\zeta)|, |\sigma_u^2 - \Phi_T^\sigma(\zeta)|) \leq o_{P_1}(1)$ component-wise;

c. $\sup_{\mathbb{A}_T} \|(T(\hat{\alpha}_Y - \hat{\alpha}_\delta), T^{1/2}(\hat{\gamma}_Y - \hat{\gamma}_\delta)')\| = o_{P_1}(1)$ and $\sup_{\mathbb{A}_T} |\hat{\sigma}_Y^2 - \hat{\sigma}_\delta^2| = o_{P_1}(1)$, where $\hat{\alpha}_Y, \hat{\gamma}_Y$ and $\hat{\sigma}_Y^2$ ($\hat{\alpha}_\delta, \hat{\gamma}_\delta$ and $\hat{\sigma}_\delta^2$) are the OLS estimators of α, γ and σ_ε^2 from the regression $Y_t = \alpha Y_{t-1} + \gamma' \nabla \mathbf{Y}_{t-1} + \text{error}_t$ ($X_t^\delta = \alpha X_{t-1}^\delta + \gamma' \nabla \mathbf{X}_{t-1}^\delta + \text{error}_t$);

d. $\sup_{z \in \mathbb{A}_T} \|(\Phi_T)'_\zeta|_{\zeta=\zeta(z)}\| \leq o_{P_1}(1)$, where $(\Phi_T)'_\zeta$ is the Jacobian matrix of Φ_T as a function of ζ .

Remark 3.4. The maximum and the minimum in the hypothesis of Lemma 3 are well-defined random variables due to choice of $\zeta(z)$. The notation $\sup_{\mathbb{A}_T}(\cdot) \leq o_{P_1}(\cdot)$ corresponds to the statement that there exists an $o_{P_1}(\cdot)$ sequence which dominates $\sup_{\mathbb{A}_T}(\cdot)$, while $\sup_{\mathbb{A}_T}(\cdot) = o_{P_1}(\cdot)$ means that there exist, first, an $o_{P_1}(\cdot)$ sequence, and second, a sequence of sets with P_1 -probability approaching one, such that for outcomes in these sets, $\sup_{\mathbb{A}_T}(\cdot)$ is dominated by the $o_{P_1}(\cdot)$ sequence. For the analysis here the two kinds of uniform convergence have the same implications.

Remark 3.5. According to the lemma, the pseudo probabilities $\tilde{\delta}_t(\zeta)$ are consistent for δ_t uniformly in $z \in \mathbb{A}_T$. The faster rate of $T^{-(\nu-2)/2}$ obtains if attention is restricted to periods with level shifts. This rate implies that better detection of periods with level shifts obtains by choosing higher values of ν ; under the hypothesis of the lemma, the optimal choice is $\nu = r - 1$. A possible explanation is that a larger value of r implies, through the thinner tails of the ε_t -distribution, better separation between ΔY_t and $\Delta Y_t + \theta_t$, while a larger value of ν takes advantage of the separation.

Remark 3.6. The lemma states also that the OLS estimators of α and γ based on the de-jumped data and on the unobservable Y_t are asymptotically equivalent, and their closeness is uniform in $z \in \mathbb{A}_T$. Asymptotic equivalence holds for any z belonging to any \mathbb{A}_T , not only for a fixed point of $\Phi_T(\zeta(\cdot))$. However, since the fixed point is defined to match relevant sample characteristics, it can be expected that a test based on it has better finite sample properties.

Remark 3.7. Lemma 3 is proved in the Appendix also without conditioning on the realization of level shifts; this is the only analytical result we derive unconditionally. It means that if Φ_T is truncated such that its domain and range are a closed cuboid in the positive orthant of \mathbb{R}^3 , then the fixed point of the truncated mapping can be used to construct unit root tests with the same unconditional asymptotics as those stated in Theorem 4 for tests based on $\tilde{\zeta}_T$ under the conditional measure P_1 (for the definition of truncation see the proof of Theorem 2).

We are now ready to state our main result.

Theorem 4 Let $\tilde{\zeta}_T$ be as in Theorem 2, and $\tilde{X}_t^\delta := X_t - \sum_{s=1}^t \tilde{\delta}_s(\tilde{\zeta}_T) \Delta X_s$. Let also ν of (9) satisfy $3 \leq \nu \leq r - 1$. Then, under assumptions \mathcal{M}, \mathcal{S} and the hypothesis $\alpha = 1 - c/T$, the ADF statistics $ADF_{\tilde{\alpha}}^\delta, ADF_t^\delta$ based on the regression $\tilde{X}_t^\delta = \tilde{\alpha} \tilde{X}_{t-1}^\delta + \tilde{\gamma}' \nabla \tilde{\mathbf{X}}_{t-1}^\delta + \text{error}_t$ have asymptotic distributions given by (2), i.e.

$$ADF_{\tilde{\alpha}}^\delta \xrightarrow{w} \frac{B_c(1)^2 - 1}{2 \int_0^1 B_c(s)^2 ds}, \quad ADF_t^\delta \xrightarrow{w} \frac{B_c(1)^2 - 1}{2(\int_0^1 B_c(s)^2 ds)^{1/2}},$$

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

According to Theorem 4, in the presence of level shifts the asymptotic distribution of the ADF statistics based on the de-jumped time series \tilde{X}_t^δ 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), and does not require any explicit decision rule about the set of periods affected by level shifts. Nevertheless, due to the consistency of $\tilde{\delta}_t(\tilde{\zeta}_T)$, consistent explicit decision rules are possible: for instance, for any $a \in (0, 1)$ the set of periods with $\tilde{\delta}_t(\tilde{\zeta}_T) > 1 - a$ coincides, with P_1 -probability approaching one, with the set of periods where level shifts occur.

Remark 3.8. For X_t with non-zero mean or with a linear trend, de-jumping can be combined with de-meaning and de-trending. Consider e.g. GLS de-trending; de-meaning is analogous. For the computation of the statistics ADF^δ , let δ_t be estimated and \tilde{X}_t^δ be defined as in Theorems 2 and 4. Then GLS de-trending can be applied to \tilde{X}_t^δ as described in Elliott et al. (1996), and ADF^δ can be obtained from an ADF regression for the de-trended \tilde{X}_t^δ . We study the properties of this procedure by simulation in Section 4.

3.3 Finer de-jumping

In order to improve the finite-sample properties of both the de-jumping procedure and the related unit root tests, in this subsection we discuss a de-jumping algorithm based on estimates of $\varepsilon_t + \theta_t \delta_t$ instead of ΔX_t . It is implemented by iterating until convergence the following steps: estimation of δ_t , de-jumping of X_t , ADF regression for the de-jumped series, and estimation of $\varepsilon_t + \theta_t \delta_t$ using the regression estimates of α and γ . The fixed points of the iteration are among those of the mapping $\Psi_T(\xi)$ defined next.

For a given $d = (d_1, \dots, d_T)'$ and $(a, \gamma)'$, where a substitutes $T^{1/2}(\alpha - 1) = T^{-1/2}c$, introduce $e_t^d := \Delta X_t - (a, \gamma)' A_T \mathbf{X}_{t-1}^d$. Define further $\Psi_T(\xi)$, with argument $\xi := (d', \lambda, a, \gamma', \sigma^2, \theta^2)'$ and with components $\Psi_T^\delta(\xi)$, $\Psi_T^\lambda(\xi)$, $\Psi_T^{a,\gamma}(\xi)$, $\Psi_T^\sigma(\xi)$ and $\Psi_T^\theta(\xi)$, as follows. First, for $t = 1, \dots, T$, let

$$\Psi_{Tt}^\delta(\xi) := \frac{\lambda}{T} \left[\frac{\lambda}{T} + \left(1 - \frac{\lambda}{T}\right) \frac{\phi_\nu(e_t^d; \sigma^2)}{\phi_\nu(e_t^d; \sigma^2 + T\theta^2)} \right]^{-1} \quad (11)$$

and $\Psi_T^\lambda(\xi) := \sum_{t=1}^T \Psi_{Tt}^\delta(\xi)$. These are respectively the functions in (9) with ΔX_t substituted by e_t^d , and the counterpart of (8). Denote the series X_t de-jumped with $\Psi_T^\delta(\xi)$ by $X_t^\Psi := X_t - \sum_{s=1}^t \Psi_{Ts}^\delta(\xi) \Delta X_s$; for \mathbf{X}_{t-1}^Ψ defined accordingly, let $\Psi_T^{a,\gamma}(\xi)$ be the estimator of $(a, \gamma)'$ from the regression

$$\Delta X_t^\Psi = (a, \gamma)' A_T \mathbf{X}_{t-1}^\Psi + error_t. \quad (12)$$

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

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

Let ξ_T be a fixed point of Ψ_T , and consider the ADF statistics, say ADF_α^Ψ and ADF_t^Ψ , from regression (12) for X_t^Ψ evaluated at $\xi = \xi_T$. The next theorem discusses existence and uniqueness of a fixed point ξ_T with the property that $\Psi_{Tt}^\delta(\xi_T)$ estimate δ_t consistently. It also states conditions ensuring that ADF_α^Ψ and ADF_t^Ψ have the usual asymptotic distributions (2).

Theorem 5 *Under the conditions of Theorem 2 the following facts can be established.*

EXISTENCE. *There exists a sequence ξ_T such that:*

i. $P_1(\Psi_T(\xi_T) = \xi_T) \rightarrow 1$, i.e. ξ_T is a random fixed point of Ψ_T with P_1 -probability approaching one;

ii. CONSISTENCY, 1. $\Psi_{Tt}^\delta(\xi_T)$ estimate δ_t consistently in the sense that $\sum_{t=1}^T |\delta_t - \Psi_{Tt}^\delta(\xi_T)| = O_{P_1}(T^{-1/2})$ and $\sum_{t=1}^T \delta_t(1 - \Psi_{Tt}^\delta(\xi_T)) = O_{P_1}(T^{-(\nu-2)/2})$;

iii. CONSISTENCY, 2. $(\Psi_T^\lambda(\xi_T), \Psi_T^\theta(\xi_T), \Psi_T^\sigma(\xi_T), (\Psi_T^\gamma(\xi_T))') = (N_T, H_T, \sigma_\varepsilon^2, \gamma') + o_{P_1}(1)$;

iv. UNIT ROOT TESTS. *The statistics $ADF_{\tilde{\alpha}}^\Psi$ and ADF_t^Ψ have limit distributions given by*

$$ADF_{\tilde{\alpha}}^\Psi \xrightarrow{w} \frac{B_c(1)^2 - 1}{2 \int_0^1 B_c(s)^2 ds}, \quad ADF_t^\Psi \xrightarrow{w} \frac{B_c(1)^2 - 1}{2(\int_0^1 B_c(s)^2 ds)^{1/2}},$$

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^1 and ξ_T^2 have properties (i) to (iv), then $P_1(\xi_T^1 = \xi_T^2) \rightarrow 1$.*

COMPUTABILITY. ξ_T can be computed by iteration: if $\tilde{\delta}(\tilde{\zeta}_T) := (\delta_1(\tilde{\zeta}_T), \dots, \delta_t(\tilde{\zeta}_T))$, $\xi_{T0} := (\tilde{\delta}(\tilde{\zeta}_T), \tilde{\zeta}_T^\lambda, T^{1/2}(\tilde{\alpha} - 1), \tilde{\gamma}', \tilde{\zeta}_T^\sigma, \tilde{\zeta}_T^\theta)'$ with components defined in Theorems 2 and 4, and $\xi_{Ti} = \Psi_T(\xi_{T,i-1})$ for $i \geq 1$, then $P_1(\xi_{Ti} \rightarrow_{i \rightarrow \infty} \xi_T) \rightarrow 1$ as $T \rightarrow \infty$.

The theorem is similar to Theorems 2 and 4. Uniqueness is established within a smaller class of sequences than in Theorem 2, but this has no practical implications given that the iterative algorithm converges to the relevant fixed point. The choice of initial value for the iteration is important: Theorems 2 and 4 ensure that ξ_{T0} , defined through the fixed point of the rough de-jumping procedure, is sufficiently close to ξ_T for the iterative algorithm to converge.

Remark 3.9. Compared to Φ_T from (10), 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 . In the definition of Φ_T we were able to avoid a sample-dependent dimension because the right sides of (9) upon evaluation at the fixed point $\tilde{\zeta}_T$ depended on $\tilde{\delta}_t(\tilde{\zeta}_T)$ only through $\tilde{\zeta}_T^\lambda = \sum_{t=1}^T \tilde{\delta}_t(\tilde{\zeta}_T)$. Since ΔX_t is now substituted by quantities obtained through de-jumping, it is not possible to reduce their dependence on the estimates of δ_t to dependence on a fixed number of summary variables.

Remark 3.10. As a continuation of Remark 3.8, consider an iterated version of de-jumping and GLS de-trending. The vector ξ and the mapping Ψ_T , corresponding respectively to the initial and the updated value at each iteration, are augmented with components φ_τ and Ψ_T^φ for the slope of the 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 rough de-jumping procedure. In the next section the properties of this procedure are investigated by simulation.

4 Monte Carlo results

In this section the finite-sample size and power properties of standard ADF tests ($ADF_{\hat{\alpha}}$, ADF_t) and of the proposed ADF tests based on level shifts removal are investigated by Monte Carlo simulation, either with or without level shifts. We need to establish three main things: first, that allowing for many possible shifts as assumed in our approach does not result in deteriorated size properties; 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; third, that in the presence of deterministic trends the proposed tests can be successfully applied in conjunction with a proper detrending procedure, despite that all our analytical results are proved under the assumption of no deterministic trends in the DGP. In addition, we wish to assess how the properties of the tests are affected by the choice of the number of degrees of freedom ν of the pseudo- t distribution; in particular, we want to assess whether the bounds on ν appearing in Theorems 2, 4 and 5 are strict. Finally, as a by-product we wish to provide additional evidence of the properties of standard ADF tests under multiple level shifts.

The employed DGPs are as follows. Data are simulated for sample sizes of $T = 100, 200, 400$ according to model (1) with $k = 1$, $\gamma := \gamma_1 \in \{-0.5, 0, 0.5\}$, $Y_{-1} = 0$ and IID innovations. Two distributional choices for the innovations ε_t are considered. In the first case, $\varepsilon_t \sim N(0, 1)$; Assumption $\mathcal{M}(b)$ is valid for any $r > 0$, and hence, behavior as predicted by asymptotic results is expected for arbitrarily large ν . The second considered distribution is $\varepsilon_t \sim t(10)$, which satisfies Assumption $\mathcal{M}(b)$ with $r < 10$. We consider both the unit root case, which obtains by setting α of (1) equal to unity, and the sequence of local alternatives $\alpha = 1 - c/T$, where $c := 7$ unless a linear trend is included in the DGP and in the model; in the latter case, we set $c := 13.5$.

Three specifications of the level shift component are employed. First, the standard 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 , $i = 1, \dots, 4$, 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 corresponding size magnitudes η_i , $i = 1, \dots, 4$, satisfying $\eta_1 = \eta_4 := 0.4$ and $\eta_2 = -\eta_3 := 0.35$; consequently, the level shift component is defined as

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

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

We consider both the tests based on rough de-jumping ($ADF_{\hat{\alpha}}^\delta, ADF_t^\delta$) and tests based on the finer version of Section 3.3 ($ADF_{\hat{\alpha}}^\Psi, ADF_t^\Psi$). Several values of the degrees of freedom ν have been used. In what follows, for the case of Gaussian innovations we report the results obtained by setting $\nu = \infty$ (i.e. Gaussian densities instead of t densities are used in the estimation of δ_t); we have checked that the same results obtain by choosing sufficiently large finite values of ν . For the case of $t(10)$ innovations, in order to investigate the importance of the upper bound on ν in Theorems 2, 4 and 5, we have used both $\nu = 8$ (which satisfies

the bound $\nu \leq r - 1$, see e.g. Theorem 2) and $\nu = \infty$. Initially tests based on raw data are considered; later we will discuss the case of tests run on GLS-detrended data. All tests are based on the 5% (asymptotic) nominal level, with critical values taken from Fuller (1976, Tables 10.A.1 and 10.A.2) for tests without detrending and from Ng and Perron (2001, Table I) for tests based on GLS detrending. All computations are based on 10,000 Monte Carlo replications and are carried out using Ox.

[Table 1 about here]

In Table 1 we report the finite sample size of the six tests considered for the case of no deterministic in the DGP; all tests statistics are based on raw data. The innovations are Gaussian and large values of ν are considered by approximating the pseudo- $t(\nu)$ density used for estimating δ_t with a Gaussian distribution. 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 extremely accurate, with sizes ranging from 4.8 to 5.5 for the ADF^δ tests (rough 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 hypothesis⁹. 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 $\gamma = -0.5$ and $\gamma = 0.5$ ¹⁰. Conversely, despite their generality the ADF^δ and ADF^Ψ tests display extremely 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.

[Table 2 about here]

In Table 2 the size-adjusted power of the six tests is investigated under the local alternative $\alpha = 1 - 7/T$. In the case of no level shifts, it is immediate to see that the new tests, although allowing for multiple level shifts, have roughly the same power properties as the standard ADF tests. That is, allowing for possible level shifts does not deteriorate the power properties of the tests. The picture changes dramatically when level shifts occur. For Model S_4 , standard ADF tests have extremely low power: 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 striking result is that the tests based on de-jumping have excellent properties, with powers growing rapidly toward the corresponding asymptotic power envelope as the sample size increases. For $T = 100$, 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 $\gamma = 0, 0.5$. The unit root tests based on finer de-jumping (ADF^Ψ) seems to be preferable over the tests based on rough de-jumping (ADF^δ) for small

⁹This is an important conclusion given that we did not present analytical results for model S_0 .

¹⁰This result agrees with the asymptotics of Section 2, where it is shown (i) that in the presence of level shifts standard unit root tests does not behave according to the standard asymptotic theory even under the null hypothesis, and (ii) that the ADF test statistics are not longer invariant to the short run parameters γ .

sample sizes. Note also that for low T also the power function of the new tests are affected by γ ; however, as T grows power is less dependent on γ , as expected by the asymptotic arguments of Section 3. Results for the random shift model S_r do not qualitatively differ from those obtained for the S_4 model; in general, the power loss experienced by standard ADF tests is less severe than for S_4 , and our tests have power functions which are closer to the asymptotic power envelope. Overall, the above power results are extremely significant, as they show that we are able to distinguish between unit root processes and processes which are stationary apart from many level shifts even in samples of moderate dimension.

We now turn the attention to the case of tests based on deterministic corrections. As stressed earlier, this point is extremely important since our asymptotic theory covers the case of no deterministic only. The suggested approach is to combine pseudo-GLS detrending with de-jumping in the computation of the ADF^δ and ADF^Ψ statistics. For a time series X_t and for some chosen $\bar{\alpha} := 1 - \bar{c}/T$ ($\bar{c} \geq 0$), the pseudo-GLS detrended series is defined as $\tilde{X}_t^{\bar{\alpha}} := X_t^{\bar{\alpha}} - \hat{\varphi}^{\bar{\alpha}'} Z_t^{\bar{\alpha}}$, where $(X_0^{\bar{\alpha}}, X_t^{\bar{\alpha}}) := (X_0, (1 - \bar{\alpha}L) X_t)$, $(Z_0^{\bar{\alpha}}, Z_t^{\bar{\alpha}}) := (Z_0, (1 - \bar{\alpha}L) Z_t)$ and $\hat{\varphi}^{\bar{\alpha}}$ minimizes $S(\bar{\alpha}, \hat{\varphi}^{\bar{\alpha}}) := \sum_t (X_t^{\bar{\alpha}} - \hat{\varphi}^{\bar{\alpha}'} Z_t^{\bar{\alpha}})^2$. Instead of X_t , we detrend the de-jumped series \tilde{X}_t^δ and X_t^Ψ , see remarks 3.8 and 3.10. Here we consider the case $Z_t = (1, t)'$ and, as is standard (cf. e.g. Elliott et al., 1996; Ng and Perron, 2001), we set $\bar{c} = 13.5$.

[Tables 3–4 about here]

In Tables 3 and 4 the size and size-adjusted power results for the case of GLS detrended data are reported. All the conclusions obtained for the case of no deterministic carry over: first, in the absence of level shifts our tests behave as the standard ADF tests, while in the presence of multiple shifts our tests are not (slightly) undersized as standard tests; second, although having the same power properties of standard tests under S_0 (no shifts), our tests do not experience the serious power loss of standard ADF tests which occur under multiple level shifts. On the whole, GLS detrending of the de-jumped time series does not seem to affect the size and power properties of our tests. The only difference with respect to the results of Tables 1–2 is that the (still severe) power loss of standard ADF tests stemming from the presence of level shifts is partially mitigated by detrending the data.¹¹

[Table 5 about here]

We conclude this Section by examining the properties of the tests when the innovations, instead of being Gaussian, follow a Student $t(10)$ distribution. The ADF^δ and ADF^Ψ statistics are computed both using pseudo- $t(\nu)$ densities with $\nu = 8$ (according to the constraints of Section 3) and using Gaussian densities (denoted by $\nu = \infty$). Only the size-adjusted power under Model S_4 (raw data) is discussed in the following¹²; results are reported in Table 5.

¹¹Further simulation exercises (not reported) has shown that OLS detrending could be used as well instead of pseudo-GLS detrending. Obviously, OLS detrending lead to different (asymptotic) power properties (cf. Elliott et al., 1996).

¹²Extended tables covering both size and power under models S_0 , S_4 and S_r are available from the webpage <http://www2.stat.unibo.it/cavaliere/lshifts>.

First, the power of standard tests is almost as poor as for the case of Gaussian innovations, ranging from 0 ($\gamma = -0.5$) to 15% ($\gamma = 0.5, T = 400$). Regarding the ADF^δ and ADF^Ψ tests, for $\nu = 8$ only $T > 200$ ensures a substantial power gain over standard tests. By converse, we found that the tests employing $\nu = \infty$, although violating the constraint $\nu \leq r - 1$ appearing in the asymptotic analysis, have power properties close to those observed for Gaussian errors, with only a slightly stronger dependence on the short run coefficient γ . Hence, for sample sizes comparable to those considered in the Monte Carlo experiment, it seems convenient to use large values of ν , even under non-Gaussian innovations.

5 Conclusions

In this paper, we have proposed a modification of the well-known augmented Dickey-Fuller (ADF) tests which allows to test for unit roots against stable alternatives in the presence of multiple level shifts. Contrary to previous results, 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. Nevertheless, the proposed test statistics have a limiting null distribution for which critical values are well known in the literature; moreover, they have the same asymptotic power functions as standard ADF tests under no level shifts. A Monte Carlo simulation has allowed to show that the new tests behave well in small samples, and that they can easily account for general deterministic time trends.

The results of this paper can be extended in various directions, which we reserve for further research. For instance, the number of lags in the autoregressive component of the model is often unknown in practice and needs to be estimated from the data; hence, it is certainly of interest to know how the new tests behave in conjunction with automatic, data dependent criteria for determining the number of lags, such as sequential t tests for the significance of the last lag or the MAIC developed by Ng and Perron (2001). A second extension would be to replace the finite autoregression assumption by a general linear process assumption, along the work of Chang and Park (2002) for the Said-Dickey-Fuller tests. A further important extensions is to generalize the univariate framework analyzed in this paper to a multivariate context. As far as we known, apart from the recent work by Lütkepohl et al. (2004) on single level shifts, so far no contributions which deal with co-integrated systems in the presence of level shifts have been made.

A Appendix

The appendix contains proofs of all the results given in the paper. Some lemmas used in the main proofs are demonstrated in A.1. The argument for Theorem 1 is presented in A.2. Theorem 4 and related results of Section 3.2 are proved in A.3. Proofs of Theorem 5 are finally reported in A.4. The following notation is used: as previously, P denotes the sequence of probability measures induced by model (1) with $\alpha = 1 - c/T$, $c \geq 0$, under assumptions \mathcal{M} and \mathcal{S} , and P_1 is P conditional on $N_T \geq 1$; uniform boundedness (convergence) in statements is denoted as in Remark 3.4, but in proofs may be abbreviated when uniformness is clear from the context, i.e. $(\cdot) \leq o_P(1)$ is sometimes used instead of $\sup(\cdot) \leq o_P(1)$, and similarly for $O_P(\cdot)$; τ denotes a number such that $\max_{t \leq T} |\varepsilon_t| = O_P(T^\tau)$, under assumption $\mathcal{M}(b)$ the choice $\tau = r^{-1}$ can be made; asterisk denotes the true values c_* and γ_* in cases of ambiguity; $\|\cdot\|$ denotes the Euclidean norm, but $\|\cdot\|_1$ and $\|\cdot\|_\infty$ norms are also used.

A.1 Preliminary lemmas

Lemma 6 *a. If $\max_{t \leq T} |\varepsilon_t| = O_P(T^\tau)$, then $\max_{t \leq T} |u_t|$ and $\max_{t \leq T} |\Delta Y_t|$ are $O_P(T^\tau)$;*

b. If $E|\varepsilon_1|^{\nu+1} < \infty$, then $\sum_{t=1}^T |\Delta Y_t|^{\nu+1} = O_P(T)$.

Since ε_t , Y_{-k} and (u_{1-k}, \dots, u_0) are independent of δ_s for all t, s , the lemma also holds under P_1 , i.e. conditionally on $N_T \geq 1$.

PROOF. Part (a) is standard from the moving average representation of u_t with exponentially decreasing coefficients, and from $\Delta Y_t = (-c/T)Y_{t-1} + u_t$ with $\max_{t \leq T} |T^{-1/2}Y_{t-1}| \xrightarrow{w} \max_{s \in [0,1]} |B_c(s)|$ (see (2) for the definition of $B_c(s)$). Next, since u_t is the stationary solution of $u_t = \sum_{i=1}^k u_{t-i} + \varepsilon_t$, it follows that $(u_t)^{\nu+1}$ is stationary and ergodic (Th. 3.35 in White (2001)). Then $E|\varepsilon_1|^{\nu+1} < \infty$ implies by the Zygmund-Marcinkiewicz or Burkholder's inequality that $E|u_t|^{\nu+1} < \infty$. By an ergodic LLN (Th. 3.34 in White (2001)), $\sum_{t=1}^T |u_t|^{\nu+1} = O_P(T)$. From

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

it is seen that $\sum_{t=1}^T |\Delta Y_t|^{\nu+1} = O_P(T)$, since $\sum_{t=1}^T |T^{-1/2}Y_{t-1}|^{\nu+1} \leq T(\max_{t \leq T} |T^{-1/2}Y_t|)^{\nu+1} = O_P(T)$. ■

Lemma 7 *If $y_t = O_Q(1)$, $Q \in \{P, P_1\}$, is a sequence of random variables independent of δ_s for all t, s under Q , and if N_T satisfies Assumption $\mathcal{S}(a)$, then $\sum_{t=1}^T \delta_t |y_t| = O_Q(1)$ and $\max_{t: \delta_t=1} |y_t| := \max_{t \leq T} |\delta_t y_t| = O_Q(1)$.*

PROOF. Direct from the definition of boundedness in probability. ■

Lemma 8 *If \hat{e}_t is a sequence such that $\max_{t \leq T} |\hat{e}_t - \varepsilon_t - \delta_t \theta_t| = o_Q(T^{1/2})$, then $\max_{t: \delta_t=1} \hat{e}_t^2 = O_Q(T)$, $Q \in \{P, P_1\}$. Further, $T^{-1} \min_{t: \delta_t=1} \hat{e}_t^2$ is bounded away from zero in P_1 -probability.*

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

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

by applying Lemma 7 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_1 -probability obtains from $T^{-1} \min_{t:\delta_t=1} \hat{e}_t^2 \geq \min_{t:\delta_t=1} \eta_t^2 \min_{t:\delta_t=1} |1 - |\eta_t^{-1}(T^{-1/2}\hat{e}_t - \eta_t)||^2$. The first minimum on the right side equals $[\max_{t:\delta_t=1} \eta_t^{-2}]^{-1} = [O_{P_1}(1)]^{-1}$ by Lemma 7 and the assumption that $\eta_t^{-1} = O_{P_1}(1)$, while the second one is $1 + o_{P_1}(1)$. Indeed, since

$$\max_{t:\delta_t=1} |\eta_t^{-1}(\hat{e}_t - T^{1/2}\eta_t)| \leq \max_{t:\delta_t=1} |\eta_t^{-1}| \left[\max_{t \leq T} |\hat{e}_t - \varepsilon_t - T^{1/2}\delta_t\eta_t| + \max_{t:\delta_t=1} |\varepsilon_t| \right] = o_{P_1}(T^{1/2}),$$

it follows that, with P_1 -probability approaching one,

$$\min_{t:\delta_t=1} \left| 1 - |\eta_t^{-1}(T^{-1/2}\hat{e}_t - \eta_t)| \right| = 1 - T^{-1/2} \max_{\delta_t=1} |\eta_t^{-1}(\hat{e}_t - T^{1/2}\eta_t)| = 1 + o_{P_1}(1).$$

■

A.2 Proof of Theorem 1.

Recall that $X_t = Y_t + \mu_t$ where, using local-to-unity asymptotics, $T^{-1/2}Y_{[T \cdot]} \xrightarrow{w} \sigma B_c(\cdot)$. To the term $T^{-1/2}\mu_{[T \cdot]} = \eta \sum_{t=1}^{\lfloor T \cdot \rfloor} \delta_t$ (under the specification $\eta_t = \eta$) we can apply e.g. Theorem 1 of He and Wang (1995), which shows the weak convergence $\sum_{t=1}^{\lfloor T \cdot \rfloor} \delta_t \xrightarrow{w} \mathcal{P}_\lambda(\cdot)$. Due to the stochastic independence of η , μ_t and Y_t , the joint convergence

$$T^{-1/2}(Y_{[T \cdot]}, \mu_{[T \cdot]}) \xrightarrow{w} (\sigma B_c(\cdot), \eta \mathcal{P}_\lambda(\cdot)) \quad (13)$$

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

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

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

Let $\nabla \mathbf{X}_{t-1} := (\Delta X_{t-1}, \dots, \Delta X_{t-k})'$, and similarly for $\nabla \mathbf{Y}_{t-1}$ and $\nabla \mu_{t-1}$. The numerator of the $ADF_{\hat{\alpha}}$ test statistic based on X_t can be expressed as (Chang and Park, 2002) $T(\hat{\alpha} - 1) = (A_T/T)(B_T/T^2)^{-1}$, where

$$A_T := \sum_{t=1}^T X_{t-1} \tilde{\varepsilon}_t - \left(\sum_{t=1}^T X_{t-1} \nabla \mathbf{X}'_{t-1} \right) \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} \nabla \mathbf{X}'_{t-1} \right)^{-1} \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} \tilde{\varepsilon}_t \right),$$

$$B_T := \sum_{t=1}^T X_{t-1}^2 - \left(\sum_{t=1}^T X_{t-1} \nabla \mathbf{X}'_{t-1} \right) \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} \nabla \mathbf{X}'_{t-1} \right)^{-1} \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} X_{t-1} \right).$$

To write the limits of these and other related quantities, we use representations in terms of stochastic integrals. They follow from Theorem 2.7 of Kurtz and Protter (1991), which requires the convergence

$$T^{-1/2}(Y_{[T \cdot]}, \mu_{[T \cdot]}, \sum_{t=1}^{\lfloor T \cdot \rfloor} \varepsilon_t, T^{1/2} \sum_{t=1}^{\lfloor T \cdot \rfloor} \delta_t) \xrightarrow{w} (\sigma B_c(\cdot), \eta \mathcal{P}_\lambda(\cdot), \sigma_\varepsilon B(\cdot), \mathcal{P}_\lambda(\cdot)) \quad (14)$$

on the space $\mathcal{D}_4[0, 1]$ of 4-vector functions with cadlag components endowed with the Skorohod topology. The convergence holds on the product space $(\mathcal{D}[0, 1])^{\times 4}$ similarly to (13), and can be extended to $\mathcal{D}_4[0, 1]$ by Proposition 2.2(b), Ch.6, in Jacod and Shiryaev (1987). Thus, Kurtz and Protter's theorem delivers

$$T^{-1} \sum_{t=1}^T (Y_{t-1}, \mu_{t-1})' (\varepsilon_t, \Delta\mu_t) \xrightarrow{w} \int_0^1 (\sigma B_c(s), \eta \mathcal{P}_\lambda(s))' d(\sigma_\varepsilon B(s), \eta \mathcal{P}_\lambda(s)) \quad (15)$$

jointly with (14). 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. Since $\sum_{t=1}^T \delta_{t-i} \delta_{t-j} = o_P(1)$ for $i \neq j$, while $\delta_{T-i} = 0$ with probability approaching one for $i = 0, \dots, k$, it follows that

$$\mathbf{M} = \eta^2 N_T \mathbf{I}_{k+1} + o_P(1) \xrightarrow{w} \eta^2 \mathcal{P}_\lambda(1) \mathbf{I}_{k+1}. \quad (16)$$

Convergence is joint with (15) and (14) because $\eta^2 N_T \mathbf{I}_{k+1}$ is a continuous transformation of the left side of (14).

We can now prove the following lemma about convergence of product moments.

Lemma 9 *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 -vector of ones. Introduce also the notation $[\mathcal{P}_\lambda] := \eta^2 \mathcal{P}_\lambda(1)$ for the quadratic variation of $\eta \mathcal{P}_\lambda(\cdot)$ over $[0, 1]$. Then, as $T \rightarrow \infty$, the following converge jointly:*

- (i) $T^{-2} \sum_{t=1}^T X_{t-1}^2 \xrightarrow{w} \sigma^2 \int_0^1 H_c(s)^2 ds$;
- (ii) $T^{-1} \sum_{t=1}^T X_{t-1} \tilde{\varepsilon}_t \xrightarrow{w} \Gamma(1) \sigma^2 \int_0^1 H_c(s) dH_c(s) + (\Gamma(1) - 1) [\mathcal{P}_\lambda]$;
- (iii) $T^{-1} \sum_{t=1}^T \nabla \mathbf{X}_{t-1} \nabla \mathbf{X}'_{t-1} \xrightarrow{w} \Omega_{\nabla\nabla} + [\mathcal{P}_\lambda] \mathbf{I}_k =: \Omega_{\nabla\nabla}^{\mathcal{P}}$;
- (iv) $T^{-1} \sum_{t=1}^T X_{t-1} \nabla \mathbf{X}'_{t-1} \xrightarrow{w} \mathbf{1}'_k \{ \sigma^2 \int_0^1 H_c(s) dH_c(s) + [\mathcal{P}_\lambda] \} + \Omega_{1\nabla}$;
- (v) $T^{-1} \sum_{t=1}^T \nabla \mathbf{X}_{t-1} \tilde{\varepsilon}_t \xrightarrow{w} -\gamma [\mathcal{P}_\lambda]$;
- (vi) $T^{-1} \sum_{t=1}^T \tilde{\varepsilon}_t^2 \xrightarrow{w} \sigma_\varepsilon^2 + [\mathcal{P}_\lambda] (1 + \gamma' \gamma)$.

PROOF. Convergence (i) obtains from (5) and CMT applied to the functional $f \rightarrow \int_0^1 f(s)^2 ds$. Further, the left side of (ii) equals $\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3$, with the following \mathbf{A}_i 's. First, $\mathbf{A}_1 := T^{-1} \sum_{t=1}^T X_{t-1} \varepsilon_t \xrightarrow{w} \sigma^2 \Gamma(1) \int_0^1 H_c(s) dB(s)$ by (15) 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 as

$$\mathbf{A}_2 = - \sum_{i=0}^k \gamma_i \left[T^{-1} \sum_{t=1}^T X_{t-i-1} \Delta\mu_{t-i} + T^{-1} \sum_{t=1}^T \left(\sum_{j=1}^i \Delta\mu_{t-j} \right) \Delta\mu_{t-i} + T^{-1} \sum_{t=1}^T \left(\sum_{j=1}^i \Delta Y_{t-j} \right) \Delta\mu_{t-i} \right]$$

with $\gamma_0 := -1$. The first summation in brackets contributes to \mathbf{A}_2 with $\Gamma(1) T^{-1} \sum_{t=1}^T X_{t-1} \Delta\mu_t + o_P(1) \xrightarrow{w} \Gamma(1) \sigma \eta \int_0^1 H_c(s) d\mathcal{P}_\lambda(s)$. The contribution of the second one is a continuous transformation of \mathbf{M} in (16) and tends to $(\Gamma(1) - 1) [\mathcal{P}_\lambda]$. Since $T^{-1} \sum_{t=1}^T \Delta Y_{t-j} \Delta\mu_{t-i} = o_P(1)$ by Lemma 7, we can conclude that $\mathbf{A}_2 \xrightarrow{w} \Gamma(1) \sigma \eta \int_0^1 H_c(s) d\mathcal{P}_\lambda(s) + (\Gamma(1) - 1) [\mathcal{P}_\lambda]$. Third, $\mathbf{A}_3 := -T^{-2} c \sum_{t=1}^T X_{t-1} \Gamma(L) Y_{t-1} \rightarrow -c \sigma^2 \Gamma(1) \int_0^1 H_c(s) B_c(s) ds$ by (14) and CMT. Combining the limits of the \mathbf{A}_i 's, and recalling that $B_c(s)$ satisfies the Langevin equation $dB_c(s) = dB(s) - cB_c(s) ds$ gives the limit asserted in (ii).

Convergence (iii) follows from the identity

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

where $\mathbf{B} := T^{-1} \sum_{t=1}^T \nabla \mathbf{Y}_{t-1} \nabla \boldsymbol{\mu}'_{t-1} = o_P(1)$ by Lemma 7. The other two terms have limits $\Omega_{\nabla \nabla}$ and $[\mathcal{P}_\lambda] \mathbf{I}_k$ respectively, the latter by (16). For (iv), note that $T^{-1} \sum_{t=1}^T X_{t-1} \nabla \mathbf{X}'_{t-1}$ equals

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

where the limit of the first term is given in the hypothesis of the lemma, the second term equals $T^{-1} \sum_{t=1}^T (\mu_{t-2} \Delta Y_{t-1}, \dots, \mu_{t-k-1} \Delta Y_{t-k}) + o_P(1)$ by Lemma 7, and converges weakly to $\sigma \eta \mathbf{1}'_k \int_0^1 \mathcal{P}_\lambda(s) dB_c(s)$ by (15), while, again by Lemma 7, the third term equals

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

and can be seen to have as weak limit $\mathbf{1}'_k \{\sigma \eta \int_0^1 H_c(s) d\mathcal{P}_\lambda(s) + [\mathcal{P}_\lambda]\}$ by (15), (16) and CMT.

For items (v) and (vi) we have

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

since two kinds of $o_P(1)$ terms appear in these expressions. First, $\mathbf{C} := T^{-1} \sum_{t=1}^T \nabla \mathbf{Y}_{t-1} \Gamma(L) \Delta \mu_t$, $\mathbf{D} := T^{-1} \sum_{t=1}^T \nabla \boldsymbol{\mu}_{t-1} \varepsilon_t$ and $\mathbf{G} := T^{-1} \sum_{t=1}^T \varepsilon_t \Gamma(L) \Delta \mu_t$ are $o_P(1)$ by Lemma 7. Second, $\mathbf{E} := T^{-1} \sum_{t=1}^T \nabla \mathbf{Y}_{t-1} \varepsilon_t \xrightarrow{P} 0$ by an LLN for $T^{-1} \sum_{t=1}^T (u_{t-1}, \dots, u_{t-k})' \varepsilon_t$ and by (15) for $cT^{-2} \sum_{t=1}^T (Y_{t-2}, \dots, Y_{t-k-1})' \varepsilon_t$, 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 $T^{-1} \sum_{t=1}^T \varepsilon_t^2 \xrightarrow{P} \sigma_\varepsilon^2$ by an LLN, $T^{-1} \sum_{t=1}^T \nabla \boldsymbol{\mu}_{t-1} \Gamma(L) \Delta \mu_t = (0_{k \times 1}, \mathbf{I}_k) \mathbf{M}(1, -\gamma)' \xrightarrow{w} -\gamma [\mathcal{P}_\lambda]$ by (16) and CMT, and $T^{-1} \sum_{t=1}^T (\Gamma(L) \Delta \mu_t)^2 = \text{tr}((1, -\gamma)'(1, -\gamma) \mathbf{M}) \xrightarrow{w} [\mathcal{P}_\lambda] (1 + \gamma' \gamma)$ by (16) 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 (14) and (15). ■

From Lemma 9 it follows that, with $\Gamma_\infty := \Gamma(1) + \mathbf{1}'_k (\Omega_{\nabla \nabla}^P)^{-1} \gamma [\mathcal{P}_\lambda]$,

$$\begin{aligned} T^{-1} A_T &\xrightarrow{w} \Gamma_\infty \sigma^2 \int_0^1 H_c(s) dH_c(s) + (\Gamma_\infty + \Omega_{1\nabla} (\Omega_{\nabla \nabla}^P)^{-1} \gamma - 1) [\mathcal{P}_\lambda], \\ T^{-2} B_T &\xrightarrow{w} \sigma^2 \int_0^1 H_c(s)^2 ds. \end{aligned}$$

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

$$\hat{\Gamma}(1) = \Gamma(1) - \mathbf{1}'_k \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} \nabla \mathbf{X}'_{t-1} \right)^{-1} \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} \tilde{\varepsilon}_t \right) + o_P(1).$$

From Lemma 9 we obtain $\hat{\Gamma}(1) \xrightarrow{w} \Gamma_\infty$. Taken together with $ADF_{\hat{\alpha}} = T(\hat{\alpha} - 1)/\hat{\Gamma}(1)$, these results yield the limit of $ADF_{\hat{\alpha}}$ in (6) with $\omega_0 := (1 + (\Omega_{1\nabla}(\Omega_{\nabla\nabla}^P)^{-1}\gamma - 1)\Gamma_\infty^{-1})[\mathcal{P}_\lambda]/\sigma^2$.

The proof for the t statistics follows similarly. Specifically, as $A_T^2/T^2 = o_P(1)$, see above, the following equalities hold:

$$\begin{aligned} ADF_t &= A_T/T (B_T C_T/T^3 - A_T^2/T^3)^{-1/2} = \frac{A_T/T}{B_T/T^2} (B_T/T^2)^{1/2} (C_T/T)^{-1/2} + o_P(1) \\ &= T(\hat{\alpha} - 1) (B_T/T)^{1/2} (C_T/T)^{-1/2} + o_P(1) \end{aligned} \quad (17)$$

with (Chang and Park, 2002)

$$C_T := \sum_{t=1}^T \tilde{\varepsilon}_t^2 - \left(\sum_{t=1}^T \tilde{\varepsilon}_t \nabla \mathbf{X}'_{t-1} \right) \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} \nabla \mathbf{X}'_{t-1} \right)^{-1} \left(\sum_{t=1}^T \nabla \mathbf{X}_{t-1} \tilde{\varepsilon}_t \right). \quad (18)$$

According to Lemma 9, $T^{-1}C_T \xrightarrow{w} \sigma_\varepsilon^2 + [\mathcal{P}_\lambda](1 + \gamma'\gamma) - \gamma'(\Omega_{\nabla\nabla}^P)^{-1}\gamma[\mathcal{P}_\lambda]^2 =: C_\infty$. Hence, (17) and (18) ensure that

$$ADF_t \xrightarrow{w} \frac{\Gamma_\infty \sigma^2 \int_0^1 H_c(s) dH_c(s) + (\Gamma_\infty + \Omega_{1\nabla}(\Omega_{\nabla\nabla}^P)^{-1}\gamma - 1)[\mathcal{P}_\lambda]}{\sigma C_\infty^{1/2} (\int_0^1 H_c(s)^2 ds)^{1/2}},$$

which is the limit in (6) with $\omega_1 := C_\infty \Gamma_\infty^{-2} \sigma^{-2}$. In the special case $k = 0$, we have $\Gamma_\infty = 1$, $C_\infty = \sigma_\varepsilon^2 + [\mathcal{P}_\lambda]$ and $\sigma^2 = \sigma_\varepsilon^2$, from where the simpler expressions for ω_0 and ω_1 . ■

A.3 Proof of Theorem 4 and related results

A.3.1 Uniform evaluations related to de-jumping

Consistent shift detection. The next lemma provides conditions for uniformly consistent detection of level shifts. It is formulated for compacts \mathbb{A}_T and random functions $\zeta(z)$ satisfying the hypothesis of Lemma 3, and follows the convention adopted there to denote $\sup_{z \in \mathbb{A}_T} f(\zeta(z))$ by $\sup_{\mathbb{A}_T} f(\zeta)$.

Lemma 10 . *Let $\hat{\varepsilon}_t$ be as in Lemma 8 and $\tilde{\delta}_t(\zeta)$ be defined as in (9) with $\hat{\varepsilon}_t$ in place of ΔX_t . If $\sum_{t=1}^T (1 - \delta_t) |\hat{\varepsilon}_t|^{\nu+1} = O_Q(T)$, $Q \in \{P, P_1\}$, then it holds that $\sup_{\mathbb{A}_T} \sum_{t=1}^T \delta_t (1 - \tilde{\delta}_t(\zeta)) \leq O_Q(T^{-(\nu-2)/2})$ and $\sup_{\mathbb{A}_T} \sum_{t=1}^T |\delta_t - \tilde{\delta}_t(\zeta)| \leq O_Q(T^{\max\{-1/2, -(\nu-2)/2\}})$.*

PROOF. The density $\phi_\nu(e; a^2)$ is given by $ga^{-1} (1 + e^2/(\nu a^2))^{-(\nu+1)/2}$, where g is a normalization constant independent of a . The notation $\zeta = (\lambda, \theta^2, \sigma^2)'$ is used for $\zeta(z) = (\zeta^\lambda(z), \zeta^\theta(z), \zeta^\sigma(z))'$, with the dependence on $z \in \mathbb{A}_T$ subsumed.

Conditionally on $N_T = 0$ it holds that $\sum_{t=1}^T \delta_t (1 - \tilde{\delta}_t(\zeta)) = 0$, so we only need to derive the $O_{P_1}(T^{-(\nu-2)/2})$ -property of this sum; the $O_P(T^{-(\nu-2)/2})$ property follows by combining the cases $N_T \geq 1$ and $N_T = 0$. First,

$$\begin{aligned} 1 - \tilde{\delta}_t(\zeta) &= \left(1 - \frac{\lambda}{T} + \frac{\lambda}{T} \frac{\phi_\nu(\hat{\varepsilon}_t; \sigma^2 + T\theta^2)}{\phi_\nu(\hat{\varepsilon}_t; \sigma^2)}\right)^{-1} < \frac{T}{\lambda} \frac{\phi_\nu(\hat{\varepsilon}_t; \sigma^2)}{\phi_\nu(\hat{\varepsilon}_t; \sigma^2 + T\theta^2)} \\ &= \frac{T}{\lambda} \left(1 + T \frac{\theta^2}{\sigma^2}\right)^{1/2} \left(1 + \frac{\hat{\varepsilon}_t^2}{\nu \sigma^2}\right)^{-(\nu+1)/2} \left(1 + \frac{\hat{\varepsilon}_t^2}{\nu(\sigma^2 + T\theta^2)}\right)^{(\nu+1)/2}. \end{aligned}$$

On \mathbb{A}_T the quantity $\lambda^{-1} (1 + T\theta^2/\sigma^2)^{1/2}$ is bounded by $O_{P_1}(T^{1/2})$, and hence,

$$\delta_t(1 - \tilde{\delta}_t(\zeta)) \leq O_{P_1}(T^{3/2})\delta_t \left(1 + \frac{\min_{t:\delta_t=1} \hat{e}_t^2}{\nu\sigma^2}\right)^{-(\nu+1)/2} \left(1 + \frac{\max_{t:\delta_t=1} \hat{e}_t^2}{\nu(\sigma^2+T\theta^2)}\right)^{(\nu+1)/2}.$$

Summing over t and accounting for the assumption that $N_T = O_{P_1}(1)$ yields the evaluation

$$\sum_{t=1}^T \delta_t(1 - \tilde{\delta}_t(\zeta)) \leq O_{P_1}(T^{3/2}) \left(1 + \frac{\min_{t:\delta_t=1} \hat{e}_t^2}{\nu\sigma^2}\right)^{-(\nu+1)/2} \left(1 + \frac{\max_{t:\delta_t=1} \hat{e}_t^2}{\nu(\sigma^2+T\theta^2)}\right)^{(\nu+1)/2}.$$

Since $\sigma^2 \leq O_{P_1}(1)$ on \mathbb{A}_T and, according to Lemma 8, $T^{-1} \min_{t:\delta_t=1} \hat{e}_t^2$ is bounded away from 0 in P_1 -probability, it follows that $T^{-1}(\nu\sigma^2)^{-1} \min_{t:\delta_t=1} \hat{e}_t^2$ has the same property. Hence, $(1 + (\nu\sigma^2)^{-1} \min_{t:\delta_t=1} \hat{e}_t^2)^{-(\nu+1)/2} = O_{P_1}(T^{-(\nu+1)/2})$. On the other hand, $\max_{t:\delta_t=1} \hat{e}_t^2 = O_{P_1}(T)$ again by Lemma 8, and since θ^2 is bounded away from zero in P_1 -probability on \mathbb{A}_T , it follows that $\nu^{-1}(\sigma^2 + T\theta^2)^{-1} \max_{t:\delta_t=1} \hat{e}_t^2 = O_{P_1}(1)$. Combining the two conclusions yields $\sum_{t=1}^T \delta_t(1 - \tilde{\delta}_t(\zeta)) \leq O_{P_1}(T^{-(\nu-2)/2})$, where the dominating sequence is independent of the point $z \in \mathbb{A}_T$ at which $\zeta(z)$ is evaluated.

The order of magnitude of $\sum_{t=1}^T (1 - \delta_t) \tilde{\delta}_t(\zeta)$ is addressed next. By evaluating from below the denominator of (9), it can be concluded that

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

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

$$\begin{aligned} (1 - \delta_t) \tilde{\delta}_t(\zeta) &\leq T^{-3/2} (1 - \delta_t) (O_Q(1) + O_Q(1)|\hat{e}_t|^{\nu+1}) \\ \sum_{t=1}^T (1 - \delta_t) \tilde{\delta}_t(\zeta) &\leq O_Q(T^{-1/2}) + O_Q(T^{-3/2}) \sum_{t=1}^T (1 - \delta_t) |\hat{e}_t|^{\nu+1} \end{aligned} \quad (19)$$

where in the second line the inequality $\sum_{t=1}^T (1 - \delta_t) \leq T$ has been used. Under the hypothesis that $\sum_{t=1}^T (1 - \delta_t) |\hat{e}_t|^{\nu+1} = O_Q(T)$ we get $\sum_{t=1}^T (1 - \delta_t) \tilde{\delta}_t(\zeta) \leq O_Q(T^{-1/2})$. Combining this result with the first part of the lemma gives $\sum_{t=1}^T |\delta_t - \tilde{\delta}_t(\zeta)| = \sum_{t=1}^T (1 - \delta_t) \tilde{\delta}_t(\zeta) + \sum_{t=1}^T \delta_t(1 - \tilde{\delta}_t(\zeta)) \leq O_Q(T^{-1/2}) + O_Q(T^{-(\nu-2)/2})$. ■

Uniform consistency for the autoregressive coefficients. The next lemma evaluates the uniform distance between Y_t and $X_t^d = X_t - \sum_{s=1}^t d_s \Delta X_s$ depending on how close $d = (d_1, \dots, d_T)'$ is to the true $\delta = (\delta_1, \dots, \delta_T)'$, while Lemma 12 establishes the distance between the OLS estimators in the ADF regression based on X_t^d and that based on Y_t . For the reasons that motivated the format of Lemma 3, see Remark 3.3, d is treated as a random function of a real vector. For example, it could be that $d(z) = \tilde{\delta}(\zeta(z))$, where the components of $\tilde{\delta}(\cdot)$ are the functions (9), and $\zeta(z)$ are as in Lemma 3; other choices will be made too. The domain of d is restricted to a set \mathbb{G}_T such that, for some $\beta \geq 0$, $\sup_{z \in \mathbb{G}_T} \sum_{t=1}^T |d_t(z) - \delta_t|$ is dominated by an $O_Q(T^{-1/2+\beta})$ -sequence, $Q \in \{P, P_1\}$. In the example $d(z) = \tilde{\delta}(\zeta(z))$ we could take $\mathbb{G}_T = \mathbb{A}_T$ and $\beta = 0$, under the assumptions of Lemmas 8 and 10, and for $\nu \geq 3$. Below the dependence of d on z is subsumed, and $\sup_{z \in \mathbb{G}_T} f(d(z))$ is abbreviated to $\sup_{\mathbb{G}_T} f(d)$.

Lemma 11 Assume that the domain of d is a set \mathbb{G}_T such that $\sup_{\mathbb{G}_T} \sum_{t=1}^T |d_t - \delta_t| \leq O_P(T^{-1/2+\beta})$ for some $\beta \geq 0$. If $\max_{t \leq T} |\varepsilon_t| = O_Q(T^\tau)$ with $\beta + \tau < 1/2$ and $\beta < \tau$, then for $Q \in \{P, P_1\}$:

- $\sup_{\mathbb{G}_T; t \leq T} |Y_t - X_t^d| \leq O_Q(T^\beta)$ and $\sup_{\mathbb{G}_T; t \leq T} |X_t^d| \leq O_Q(T^{1/2})$;
- $\sup_{\mathbb{G}_T; t \leq T} \sum_{s=1}^t |\Delta Y_s - \Delta X_s^d| \leq O_Q(T^\beta)$, $\sup_{\mathbb{G}_T} |\Delta Y_t - \Delta X_t^d| \leq a_T \delta_t + b_T$ with $a_T = O_Q(T^\beta)$, $b_T = O_Q(T^{-1/2+\beta+\tau})$, and $\sup_{\mathbb{G}_T; t \leq T} |\Delta X_t^d| \leq O_Q(T^\tau)$;
- $\sup_{\mathbb{G}_T} \|A_T \mathbf{X}_t^d - A_T \mathbf{Y}_t\| \leq \tilde{a}_T \boldsymbol{\delta}_t + b_T$ and $\sup_{\mathbb{G}_T; t \leq T} \|A_T \mathbf{X}_t^d\| \leq O_Q(T^\tau)$, with $\boldsymbol{\delta}_t = \sum_{i=0}^{k-1} \delta_{t-i}$ denoting the number of level shifts between periods $t-k+1$ and t , $\tilde{a}_T = O_Q(T^\beta)$ and $b_T = O_Q(T^{-1/2+\beta+\tau})$.

PROOF. Inserting $X_t = Y_t + T^{1/2} \sum_{s=1}^t \delta_s \eta_s$ and $\Delta X_t = \Delta Y_t + T^{1/2} \delta_t \eta_t$ in the definition of X_t^d gives

$$\begin{aligned} X_t^d - Y_t &= - \sum_{s=1}^t \delta_s \Delta Y_s + \sum_{s=1}^t (\delta_s - d_s) (\Delta Y_s + T^{1/2} \delta_s \eta_s), \\ \Delta X_t^d - \Delta Y_t &= -\delta_t \Delta Y_t + (\delta_t - d_t) (\Delta Y_t + T^{1/2} \delta_t \eta_t). \end{aligned} \quad (20)$$

Thus, from the first equation,

$$\begin{aligned} |Y_t - X_t^d| &\leq \sum_{s=1}^t \delta_s |\Delta Y_s| + [\max_{s \leq T} |\Delta Y_s| + T^{1/2} \sum_{s=1}^t \delta_s |\eta_s|] \sum_{s=1}^t |\delta_s - d_s| \\ &\leq O_Q(1) + \left[O_Q(T^\tau) + O_Q(T^{1/2}) \right] O_Q(T^{-1/2+\beta}) = O_Q(T^\beta), \end{aligned} \quad (21)$$

with orders of magnitude taken respectively from Lemmas 7, 6(a), 7, and from the hypothesis. This is the first relation in (a). Together with $\max_{t \leq T} |Y_t| = O_P(T^{1/2})$, it implies the other relation there.

The right side of the first line in (21) is an upper bound also for $\sum_{s=1}^t |\Delta Y_s - \Delta X_s^d|$, and with $t = T$ the first relation in (b) follows from the above argument. Next, from the second equation in (20),

$$|\Delta Y_t - \Delta X_t^d| \leq \delta_t (\max_{t: \delta_t=1} |\Delta Y_t| + T^{1/2} \max_{t: \delta_t=1} |\eta_t| \sum_{t=1}^T |\delta_t - d_t|) + \max_{t \leq T} |\Delta Y_t| \sum_{t=1}^T |\delta_t - d_t|,$$

from where the expressions for a_T and b_T in (b) can be read. Their orders of magnitude follow as in (a). Further, $\sup_{\mathbb{G}_T; t \leq T} |\Delta X_t^d| \leq \sup_{\mathbb{G}_T; t \leq T} |\Delta Y_t - \Delta X_t^d| + \max_{t \leq T} |\Delta Y_t| = O_Q(T^\tau)$ by the above evaluation, Lemma 6(a), and since $\beta < \tau$.

A vector version $\sup_{\mathbb{G}_T} \|\nabla \mathbf{Y}_t - \nabla \mathbf{X}_t^d\| \leq \boldsymbol{\delta}_t a_T + k b_T$ 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 $\max_{t \leq T} \|A_T \mathbf{Y}_t\| \leq \max_{t \leq T} |T^{-1/2} Y_t| + k \max_{t \leq T} |\Delta Y_t| = O_Q(T^\tau)$, it holds further that

$$\sup_{\mathbb{G}_T; t \leq T} \|A_T \mathbf{X}_t^d\| \leq \max_{t \leq T} \|A_T \mathbf{Y}_t\| + \sup_{\mathbb{G}_T; t \leq T} \|A_T (\mathbf{X}_t^d - \mathbf{Y}_t)\| \leq O_Q(T^\tau) + \tilde{a}_T N_T + \tilde{b}_T,$$

which is $O_Q(T^\tau)$ for $\beta < \tau$. ■

Lemma 12 Let $(\hat{\alpha}_d - 1, \hat{\gamma}'_d)'$ and $(\hat{\alpha}_Y - 1, \hat{\gamma}'_Y)'$ be the OLS estimators for the regressions $\Delta X_t^d = (\alpha - 1, \gamma') \mathbf{X}_{t-1}^d + \text{error}_t$ and $\Delta Y_t = (\alpha - 1, \gamma') \mathbf{Y}_{t-1} + \text{error}_t$ respectively, and $\hat{\sigma}_d^2$ and $\hat{\sigma}_Y^2$ be the residual variances. Then, under the assumptions of Lemma 11 with $\beta = 0$, it holds that $\sup_{\mathbb{G}_T} \|(T(\hat{\alpha}_d - \hat{\alpha}_Y), T^{1/2}(\hat{\gamma}_d - \hat{\gamma}_Y)')\| = O_Q(T^{\tau-1/2})$ and $\sup_{\mathbb{G}_T} |\hat{\sigma}_d^2 - \hat{\sigma}_Y^2| = O_Q(T^{\tau-1/2})$.

PROOF. From (1) it follows that $\Delta Y_t = \gamma'_* \nabla \mathbf{Y}_{t-1} + \tilde{e}_t$, $\tilde{e}_t := \varepsilon_t - (c_*/T) \Gamma_*(L) Y_{t-1}$, which upon insertion of $\Delta Y_t = \Delta X_t^d + (\Delta Y_t - \Delta X_t^d)$ and $\nabla \mathbf{Y}_{t-1} = \nabla \mathbf{X}_{t-1}^d + (\nabla \mathbf{Y}_{t-1} - \nabla \mathbf{X}_{t-1}^d)$ becomes $\Delta X_t^d = \gamma'_* \nabla \mathbf{X}_{t-1}^d + \tilde{e}_t + \Gamma_*(L) (\Delta X_t^d - \Delta Y_t)$. Hence,

$$\begin{aligned} (T(\hat{\alpha}_Y - 1), T^{1/2} \hat{\gamma}'_Y)' &= T^{1/2} (0, \gamma'_*)' + (S_{11}^T)^{-1} S_{1e}^T, \\ (T(\hat{\alpha}_d - 1), T^{1/2} \hat{\gamma}'_d)' &= T^{1/2} (0, \gamma'_*)' + (S_{11}^T + R_{11}^T)^{-1} (S_{1e}^T + R_{1e}^T + R_{1\nabla}^T), \end{aligned} \quad (22)$$

with matrices $S_{11}^T := T^{-1} \sum_{t=1}^T A_T \mathbf{Y}_{t-1} (A_T \mathbf{Y}_{t-1})'$, $S_{1e}^T := T^{-1/2} \sum_{t=1}^T A_T \mathbf{Y}_{t-1} \Delta Y_t$, $R_{11}^T := T^{-1} A_T [\sum_{t=1}^T \mathbf{X}_{t-1}^d (\mathbf{X}_{t-1}^d)' - \sum_{t=1}^T \mathbf{Y}_{t-1} (\mathbf{Y}_{t-1})'] A_T$, $R_{1e}^T := T^{-1/2} A_T \sum_{t=1}^T (\mathbf{X}_{t-1}^d - \mathbf{Y}_{t-1}) \tilde{e}_t$ and $R_{1\nabla}^T := T^{-1/2} A_T \sum_{t=1}^T \mathbf{X}_{t-1}^d \Gamma_*(L) (\Delta X_t^d - \Delta Y_t)$. We argue for the existence of sequences $r_{ij}^T = O_Q(T^{\tau-1/2})$ such that $\sup_{\mathbb{G}_T} \|R_{ij}^T\| \leq r_{ij}^T$, and then use that S_{11}^T and S_{1e}^T are well-known to be normalized properly for non-degenerate convergence.

First, $\|R_{11}^T\| \leq (\max_{t \leq T} \|A_T \mathbf{X}_{t-1}^d\| + \max_{t \leq T} \|A_T \mathbf{Y}_{t-1}\|) \cdot T^{-1} \sum_{t=1}^T \|A_T \mathbf{X}_{t-1}^d - A_T \mathbf{Y}_{t-1}\|$. It is seen from Lemmas 6(a) and 11 (with $\beta = 0$) that the two maxima are $O_Q(T^\tau)$, while the summation is bounded by $T^{-1/2} \sum_{t=1}^T |X_{t-1}^d - Y_{t-1}| + k \sum_{t=1}^T |\Delta Y_t - \Delta X_t^d| \leq O_Q(T^{1/2})$, all uniformly on \mathbb{G}_T . Hence, $\sup_{\mathbb{G}_T} \|R_{11}^T\| \leq O_Q(T^{\tau-1/2})$. Second,

$$\|R_{1e}^T\| \leq T^{-1} \left| \sum_{t=1}^T (X_{t-1}^d - Y_{t-1}) \tilde{e}_t \right| + T^{-1/2} \left\| \sum_{t=1}^T (\nabla \mathbf{X}_{t-1}^d - \nabla \mathbf{Y}_{t-1}) \tilde{e}_t \right\|.$$

The difference under the absolute value can be written as

$$\sum_{t=1}^T (X_{t-1}^d - Y_{t-1}) \tilde{e}_t = (X_{T-1}^d - Y_{T-1}) \tilde{e}_t - \sum_{t=1}^{T-1} (\Delta X_t^d - \Delta Y_t) \sum_{s=1}^t \tilde{e}_t$$

by partial summation. Therefore, the absolute value itself is bounded by

$$\max_{t \leq T} |\tilde{e}_t| \max_{t \leq T} |X_t^d - Y_t| + \max_{t \leq T} \left| \sum_{s=1}^t \tilde{e}_t \right| \sum_{t=1}^T |\Delta X_t^d - \Delta Y_t| \leq O_Q(T^{1/2})$$

since $\max_{t \leq T} |\tilde{e}_t| \leq \max_{t \leq T} |\varepsilon_t| + (\|\gamma_*\|_1 + 1) (c_*/T) \max_{t \leq T} |Y_t| = O_Q(T^\tau)$, $\max_{t \leq T} |X_t^d - Y_t| \leq O_Q(1)$ by Lemma 11(a),

$$\max_{t \leq T} \left| \sum_{s=1}^t \tilde{e}_t \right| \leq \max_{t \leq T} \left| \sum_{s=1}^t \varepsilon_t \right| + (\|\gamma_*\|_1 + 1) (c_*/T) \max_{t \leq T} \left| \sum_{s=1}^t Y_{t-1} \right| = O_Q(T^{1/2}),$$

and, by Lemma 11(b), $\sup_{\mathbb{G}_T} \sum_{t=1}^T |\Delta X_t^d - \Delta Y_t| \leq O_Q(1)$. Further, still by Lemma 11,

$$\left\| \sum_{t=1}^T (\nabla \mathbf{X}_{t-1}^d - \nabla \mathbf{Y}_{t-1}) \tilde{e}_t \right\| \leq \max_{t \leq T} |\tilde{e}_t| \sum_{t=1}^T \|\nabla \mathbf{X}_{t-1}^d - \nabla \mathbf{Y}_{t-1}\| \leq O_Q(T^\tau) O_Q(1)$$

uniformly on \mathbb{G}_T . Combining the above evaluations gives $\sup_{\mathbb{G}_T} \|R_{1e}^T\| \leq O_Q(T^{\tau-1/2})$. The boundedness of $\sup_{\mathbb{G}_T} \|R_{1\nabla}^T\|$ is obtained similarly:

$$\|R_{1\nabla}^T\| \leq T^{-1/2} (\|\gamma_*\|_1 + 1) \max_{t \leq T} \|A_T \mathbf{X}_{t-1}^d\| \sum_{t=1}^T |\Delta X_t^d - \Delta Y_t| = O_Q(T^{\tau-1/2}).$$

Returning to (22) and accounting for 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 obtains that

$$(T(\hat{\alpha}_d - \hat{\alpha}_Y), T^{1/2}(\hat{\gamma}_d - \hat{\gamma}'_Y)) = (S_{11}^T)^{-1} (R_{1e}^T + R_{1\nabla}^T) - (S_{11}^T + R_{11}^T)^{-1} R_{11}^T (S_{11}^T)^{-1} (S_{1e}^T + R_{1e}^T + R_{1\nabla}^T).$$

To evaluate the norm of this expression, note that for outcomes such that $\|R_{11}^T\| \|(S_{11}^T)^{-1}\| \leq 1$, the inequality $\|(S_{11}^T + R_{11}^T)^{-1}\| \leq \|(S_{11}^T)^{-1}\| (1 - \|R_{11}^T\| \|(S_{11}^T)^{-1}\|)^{-1}$ holds.¹³ Define $\mathcal{F}_T := \{\omega : r_{11}^T \leq \|(S_{11}^T)^{-1}\|^{-1}\}$; they satisfy $Q(\mathcal{F}_T) \rightarrow 1$ since $r_{11}^T = O_Q(T^{-1/2+\tau})$ and S_{11} converges to a positive definite limit. Now

$$\|(S_{11}^T)^{-1}\| (r_{1e}^T + r_{1\nabla}^T) - r_{11}^T \|(S_{11}^T)^{-1}\| (\|S_{1e}^T\| + r_{1e}^T + r_{1\nabla}^T) / (1 - r_{11}^T \|(S_{11}^T)^{-1}\|)$$

is an $O_Q(T^{-1/2+\tau})$ -sequence dominating $\sup_{\mathbb{G}_T} \|(T(\hat{\alpha}_d - \hat{\alpha}_Y), T^{1/2}(\hat{\gamma}_d - \hat{\gamma}_Y)')\|$ for outcomes in \mathcal{F}_T .

For the residual variances we have

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

and since $\|\sum_{t=1}^T [(\Delta X_t^d)^2 - \Delta Y_t^2]\| \leq (\max_{t \leq T} |\Delta X_t^d| + \max_{t \leq T} |\Delta Y_t|) \sum_{t=1}^T |\Delta X_t^d - \Delta Y_t| \leq O_Q(T^\tau)$ uniformly on \mathbb{G}_T , see Lemma 11, from the previous argument it follows that, for outcomes in \mathcal{F}_T , $\sup_{\mathbb{G}_T} |\hat{\sigma}_d^2 - \hat{\sigma}_Y^2|$ is dominated by an $O_Q(T^{-1/2+\tau})$ sequence. ■

A.3.2 Proof of Lemma 3

The proof of items (a) to (c) is straightforward from the results in the previous section. We start by verifying that $\hat{e}_t = \Delta X_t$ satisfies the hypotheses of Lemmas 8 and 10, which yields item (a) and the first relation in item (b) of Lemma 3, provided that $3 \leq \nu \leq r-1$. Indeed, in view of assumption $\mathcal{M}(b)$, $\max_{t \leq T} |\varepsilon_t| = O_Q(T^\tau)$ with $\tau = r^{-1} \leq 1/4$, and by Lemma 6(a) also $\max_{t \leq T} |u_t| = O_Q(T^\tau)$. Together with $\max_{t \leq T} |Y_{t-1}| = O_Q(T^{1/2})$, this ensures that $\max_{t \leq T} |\Delta X_t - \varepsilon_t - T^{1/2} \delta_t \eta_t| \leq T^{-1} c \max_{t \leq T} |Y_{t-1}| + \max_{t \leq T} |u_t| = o_Q(T^{1/2})$ as required in Lemma 8. From here it is straightforward that $\max_{t \leq T} |\Delta X_t| = O_P(T^{1/2})$, which will be used below. Further, $(1 - \delta_t) |\Delta X_t|^{\nu+1} = (1 - \delta_t) |\Delta Y_t|^{\nu+1}$ and $\sum_{t=1}^T (1 - \delta_t) |\Delta X_t|^{\nu+1} \leq \sum_{t=1}^T |\Delta Y_t|^{\nu+1} = O_Q(T)$ by the hypothesis, Lemma 6(b) and due to the choice $\nu + 1 \leq r$, implying that Lemma 10 is applicable too.

In view of item (a), we can invoke Lemma 12 with $d(z) = \tilde{\delta}(\zeta(z))$, $\mathbb{G}_T = \mathbb{A}_T$, to get item (c). The two remaining relations in (b) obtain as follows. First,

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

since $\max_{t \leq T} |\Delta X_t| = O_P(T^{1/2})$. Substituting next $(1 - \delta_t) \Delta X_t^2 = (1 - \delta_t) \Delta Y_t^2$ and $\delta_t \Delta X_t = \delta_t (\Delta Y_t + T^{1/2} \eta_t)$ and passing to the limit gives the relations.

Consider now the Jacobian $\Phi_T'(\zeta)$. It is checked directly that, for $\tilde{\delta}_t(\zeta)$ defined in (9), $\partial \tilde{\delta}_t(\zeta) / \partial \zeta' = \tilde{\delta}_t(\zeta) (1 - \tilde{\delta}_t(\zeta)) v_t$, with

$$v_t = \left(\frac{1}{\lambda(1-T^{-1}\lambda)}, \frac{\partial \ln[\phi_\nu(\hat{e}_t; \sigma^2 + T\theta^2) / \phi_\nu(\hat{e}_t; \sigma^2)]}{\partial \sigma^2}, \frac{\partial \ln \phi_\nu(\hat{e}_t; \sigma^2 + T\theta^2)}{\partial \theta^2} \right)'$$

¹³It follows e.g. by taking norms on both sides of $(S_{11}^T + R_{11}^T)^{-1} = (S_{11}^T)^{-1} - (S_{11}^T)^{-1} R_{11}^T (S_{11}^T + R_{11}^T)^{-1}$, evaluating the norm of the right side, and solving the resulting inequality for $\|(S_{11}^T + R_{11}^T)^{-1}\|$.

From here $\Phi'_T(\zeta)$ can be written as $\Phi'_T(\zeta) = T^{-1} \sum_{t=1}^T \tilde{\delta}_t(\zeta) (1 - \tilde{\delta}_t(\zeta)) w_t v'_t$, with $w_t = (T, -\Delta X_t^2, \Delta X_t^2)'$, and can be evaluated as follows:

$$\|\Phi'_T(\zeta)\| \leq 3 \max_{t \leq T} \|v_t\|_\infty \max_{t \leq T} \|T^{-1} w_t\|_\infty \sum_{t=1}^T \tilde{\delta}_t(\zeta) (1 - \tilde{\delta}_t(\zeta)), \quad (23)$$

where $\|\cdot\|_\infty$ is the max-norm in \mathbb{R}^3 . To show item (d), we discuss the factors on the right side separately. Start from the components of v_t . First,

$$\frac{\partial \ln \phi_\nu(\hat{e}_t; \sigma^2 + T\theta^2)}{\partial(\sigma^2, \theta^2)} = -\frac{1}{2} \frac{1}{\sigma^2 + T\theta^2} \left[1 - (\nu + 1) \frac{\hat{E}_t^2}{1 + \hat{E}_t^2} \right] \begin{pmatrix} 1 \\ T \end{pmatrix},$$

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

$$\left| \frac{\partial \ln \phi_\nu(\hat{e}_t; \sigma^2 + T\theta^2)}{\partial \theta^2} \right| < \frac{\nu+2}{2\theta^2}, \quad \left| \frac{\partial \ln \phi_\nu(\hat{e}_t; \sigma^2 + T\theta^2)}{\partial \sigma^2} \right| < \frac{\nu+2}{2\sigma^2} \quad \text{and} \quad \left| \frac{\partial \ln \phi_\nu(\hat{e}_t; \sigma^2)}{\partial \sigma^2} \right| < \frac{\nu+2}{2\sigma^2}, \quad (24)$$

with all right sides $O_Q(1)$ on \mathbb{A}_T . Also $\lambda^{-1} (1 - T^{-1}\lambda)^{-1} \leq (\min_{\mathbb{A}_T} \lambda)^{-1} (1 - T^{-1} \max_{\mathbb{A}_T} \lambda)^{-1} = O_Q(1)$ on \mathbb{A}_T . Hence, $\sup_{\mathbb{A}_T; t \leq T} \|v_t\|_\infty \leq O_Q(1)$. The other terms on the right side of (23) are $\sum_{t=1}^T \tilde{\delta}_t(\zeta) (1 - \tilde{\delta}_t(\zeta)) \leq \sum_{t=1}^T |\delta_t - \tilde{\delta}_t(\zeta)| \leq O_Q(T^{-1/2})$ uniformly on \mathbb{A}_T by Lemma 3(a), and $\max_{t \leq T} \|T^{-1} w_t\|_\infty = \max\{1, T^{-1} \max_{t \leq T} \Delta X_t^2\} = O_Q(1)$, see above. Item (d) obtains by inserting the obtained orders of magnitude into (23). ■

A.3.3 Proof of Theorems 2 and 4

EXISTENCE. To be able to invoke a standard fixed-point theorem, we set up an auxiliary fixed-point problem for a new mapping defined by centering, normalizing and truncating Φ_T in a way such that it maps a compact onto itself. Existence of a random fixed point for that mapping is classical, and it turns out that with P_1 -probability approaching one, the same point is a fixed point of Φ_T .

Let $\mathbb{A}_T = \mathbb{A}_T^\lambda \times \mathbb{A}_T^\theta \times \mathbb{A}_T^\sigma$ with $\mathbb{A}_T^\lambda = [-1/2, 1/2]$, $\mathbb{A}_T^\theta = [1/2, 3/2]$, and $\mathbb{A}_T^\sigma = [\sigma_u^2/2, 2\sigma_u^2]$. Further, let $\zeta : \mathbb{A}_T \rightarrow \mathbb{R}^3$ act on $z = (z^\lambda, z^\theta, z^\sigma)'$ according to $\zeta(z) = (z^\lambda + N_T, H_T z^\theta, z^\sigma)'$, and let $\Theta_T = \zeta^{-1} \circ \Phi_T \circ \zeta$. Under P_1 we have $H_T \neq 0$ a.s., and hence, Θ_T is well-defined a.s. Further, let $\Theta_T^{tr} : \mathbb{A}_T \rightarrow \mathbb{A}_T$ be defined component-wise: for $v \in \{\lambda, \theta, \sigma\}$, $\Theta_T^{tr, v}(z) = \Theta_T^v(z)$ 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 $\tilde{z}_T \in \mathbb{A}_T$. This is guaranteed for example by Theorem 10 in Bharucha-Reid (1976).

Since, in view of the evaluations in Lemma 3(b), $\Theta_T(\tilde{z}_T) = \zeta^{-1}((N_T, H_T, \sigma_u^2)') + o_{P_1}(1) = (0, 1, \sigma_u^2)'$, and since the point $(0, 1, \sigma_u^2)'$ is interior for \mathbb{A}_T , it follows that $P_1(\Theta_T(\tilde{z}_T) \in \mathbb{A}_T) \rightarrow 1$. Note next that for outcomes such that $\Theta_T(\tilde{z}_T) \in \mathbb{A}_T$ we have $\tilde{z}_T = \Theta_T^{tr}(\tilde{z}_T) = \Theta_T(\tilde{z}_T)$, i.e. $\tilde{\zeta}_T = \zeta(\tilde{z}_T)$ is a fixed point of Φ_T . Hence, \tilde{z}_T is a random fixed point of Φ_T on \mathbb{A}_T with P_1 -probability approaching one. ■

UNIQUENESS. This part of the proof uses a contraction argument. Let $\tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$ have the properties of $\tilde{\zeta}_T$ from the existence part of the theorem. Fix $\epsilon > 0$. Since H_T , $\tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$ are bounded away from 0 in P_1 -probability (by Lemma 8(a) applied to η_t^{-1} for the first of them),

there exist a constant $\underline{l} > 0$ and a set \mathcal{A}_1 with $P_1(\mathcal{A}_1) > 1 - \epsilon/10$ such that for outcomes in \mathcal{A}_1 it holds that $H_T \geq 2\underline{l}$, $\tilde{\zeta}_T^1 \geq \underline{l}$ and $\tilde{\zeta}_T^2 \geq \underline{l}$, where for $\tilde{\zeta}_T^i$ the inequalities are satisfied by each component. Similarly, since $H_T, N_T, \tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$ are $O_{P_1}(1)$ (again by Lemma 8(a) for the first of them), there exist a constant $\bar{l} > 2\underline{l}$ and a set \mathcal{A}_2 with $P_1(\mathcal{A}_2) > 1 - \epsilon/10$ such that for outcomes in \mathcal{A}_2 , $H_T \leq \bar{l}$, $N_T \leq \bar{l}$, $\tilde{\zeta}_T^1 \leq \bar{l}$ and $\tilde{\zeta}_T^2 \leq \bar{l}$.

Define the compact \mathbb{K} as $[\min(\underline{l}, 1/2), 2\bar{l}] \times [\underline{l}, 2\bar{l}] \times [\min(\sigma_u^2/2, \underline{l}), \max(2\sigma_U^2, \bar{l})]$. This choice ensures two things for outcomes in $\mathcal{A}_1 \cap \mathcal{A}_2$. First, that $\tilde{\zeta}_T^1 \in \mathbb{K}$ and $\tilde{\zeta}_T^2 \in \mathbb{K}$, and second, that the point $(N_T, H_T, \sigma_u^2)'$ is interior for \mathbb{K} and at distance from $\partial\mathbb{K}$ bounded from below by a positive number. In view of the latter, it follows from Lemma 3(b) applied to $\mathbb{A}_T = \mathbb{K}$ and $\zeta(z)$ equal to the identity function that there exist an $\mathcal{A}_3 \subset \mathcal{A}_1 \cap \mathcal{A}_2$ with $P_1(\mathcal{A}_3) > 1 - \epsilon/4$ and an integer T_3 such that for outcomes in \mathcal{A}_3 and $T > T_3$, $\Phi_T(\mathbb{K}) \subset \mathbb{K}$. By item (d) of the same lemma, there exist $T_4 > T_3$ and a set $\mathcal{A}_4 \subset \mathcal{A}_3$, with $P_1(\mathcal{A}_4) > 1 - \epsilon/2$, such that for $T > T_4$ and on \mathcal{A}_4 , $\sup_{\zeta \in \mathbb{K}} \|(\Phi_T)'_{\zeta}\| < 1/2$. Thus, under the same conditions Φ_T is a contraction, and for each outcome in \mathcal{A}_4 it has a unique fixed point by Banach's fixed point theorem.

Consider finally T_5 and \mathcal{A}_5 with $P_1(\mathcal{A}_5) > 1 - \epsilon/2$ such that for $T > T_5$ and outcomes in \mathcal{A}_5 , $\Phi_T(\tilde{\zeta}_T^1) = \tilde{\zeta}_T^1$ and $\Phi_T(\tilde{\zeta}_T^2) = \tilde{\zeta}_T^2$. This is possible by the choice of $\tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$. Then for outcomes in $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_5$ and $T > T_5$ both $\tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$ are fixed points of Φ_T on \mathbb{K} , while for outcomes in $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_4 \cap \mathcal{A}_5 = \mathcal{A}_4 \cap \mathcal{A}_5$ and $T > \max(T_3, T_4, T_5)$, $\tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$ must be equal to the unique fixed point of Φ_T on \mathbb{K} . Since $P_1(\mathcal{A}_4 \cap \mathcal{A}_5) > 1 - \epsilon$ and ϵ is arbitrary, this means that $P_1(\tilde{\zeta}_T^1 = \tilde{\zeta}_T^2) \rightarrow 1$. ■

COMPUTABILITY. This proof uses a contraction argument too. Let $\tilde{\mathbb{K}} := \mathbb{K} \cup \{\zeta_{T_0}\}$, where \mathbb{K} was defined in the previous proof. As in the previous proof, there exist a set $\mathcal{B}_1 \subset \mathcal{A}_1 \cap \mathcal{A}_2$ with $P_1(\mathcal{B}_1) > 1 - \epsilon/2$ and an integer T_1 such that for outcomes in \mathcal{B}_1 and $T > T_1$, Φ_T is a contraction of $\tilde{\mathbb{K}}$ onto $\tilde{\mathbb{K}}$ and $\tilde{\zeta}_T \in \tilde{\mathbb{K}}$, with $\tilde{\zeta}_T$ standing for any of $\tilde{\zeta}_T^1$ and $\tilde{\zeta}_T^2$. Further, there exist a set \mathcal{B}_2 with $P_1(\mathcal{B}_2) > 1 - \epsilon/2$ and an integer T_2 such that for outcomes in \mathcal{B}_2 and $T > T_2$ it holds that $\Phi_T(\tilde{\zeta}_T) = \tilde{\zeta}_T$. Hence, for outcomes $\omega \in \mathcal{B}_1 \cap \mathcal{B}_2$, with $P_1(\mathcal{B}_1 \cap \mathcal{B}_2) > 1 - \epsilon$, and for $T > \max(T_1, T_2)$, the sequence of iterates of Φ_T converges to its unique fixed point $\tilde{\zeta}_T(\omega)$ on $\tilde{\mathbb{K}}$. Convergence is again implied by Banach's fixed point theorem. ■

The remaining statements in Theorem 2 and Theorem 4 follow in a straightforward way from Lemmas 3 (with \mathbb{A}_T as in the existence proof) and 12 (with $d(z) = \tilde{\delta}(\zeta(z))$, $\mathbb{G}_T = \mathbb{A}_T$), and from (2).

A.4 Proof of Theorem 5 and related results

The role of the compact \mathbb{A}_T from the proof of Theorem 2 is played here by $\mathbb{D}_T \subset \mathbb{R}^{T+k+4}$ defined as $\mathbb{D}_T = \prod_{v \in V} \mathbb{D}_T^v$, $V = \{\delta, \lambda, a, \gamma, \theta, \sigma\}$, with factors $\mathbb{D}_T^\delta = \{x^\delta \in \mathbb{R}^T : \sum_{t=1}^T |x_t^\delta| \leq T^{-1/2+\rho}\}$, $\mathbb{D}_T^\lambda = [-1/2, 1/2]$, $\mathbb{D}_T^a = [-T^{-1/2+\rho}, T^{-1/2+\rho}]$, $\mathbb{D}_T^\gamma = \{x^\gamma \in \mathbb{R}^k : \|x^\gamma - \gamma_*\| \leq 1\}$, $\mathbb{D}_T^\theta = [1/2, 3/2]$ and $\mathbb{D}_T^\sigma = [\sigma_\varepsilon^2/2, 2\sigma_\varepsilon^2]$, where $\rho > 0$ will be chosen as close to 0 as convenient. The point at which Ψ_T is evaluated is specified as a transformation $\xi(x)$ with components $d(x) = x^\delta + \delta$ with $\delta = (\delta_1, \dots, \delta_T)'$, $\lambda(x) = x^\lambda + N_T$, $(a, \gamma)'(x) = (x^a, (x^\gamma)')$, $\theta^2(x) = H_T x^\theta$ and $\sigma^2(x) = x^\sigma$. It will be shown that there exists a random sequence $x_T \in \mathbb{D}_T$ such that $P_1(\Psi_T(\xi(x_T)) = \xi(x_T)) \rightarrow 1$. The argument uses properties in the spirit of Lemma 3 that are worked out next. The notation $\sup_{x \in \mathbb{D}_T} f(\xi(x)) = \sup_{\mathbb{D}_T} f(\xi)$ is employed, and the

argument x of ξ and its components is subsumed.

A.4.1 More lemmas

Recall the notation $X_t^\Psi := X_t - \sum_{s=1}^t \Psi_{T_s}^\delta \Delta X_s$ ($X_t^d := X_t - \sum_{s=1}^t d_s \Delta X_s$) and $e_t^\Psi := \Delta X_t - \Psi_T^{\alpha, \gamma} A_T \mathbf{X}_{t-1}^\Psi$ ($e_t^d := \Delta X_t - (a, \gamma') A_T \mathbf{X}_{t-1}^d$).

Lemma 13 *Let $3 \leq \nu \leq r - 1$, $\rho < \tau$, $\rho + \tau < 1/2$ and $2\rho(\nu + 1) \leq 1$, where $\max_{t \leq T} |\varepsilon_t| = O_{P_1}(T^\tau)$. Then:*

- $\sup_{\mathbb{D}_T} \sum_{t=1}^T |\delta_t - \Psi_{Tt}^\delta| \leq O_{P_1}(T^{-1/2})$, $\sup_{\mathbb{D}_T} \sum_{t=1}^T \delta_t (1 - \Psi_{Tt}^\delta) \leq O_{P_1}(T^{-(\nu-2)/2})$;
- $\sup_{\mathbb{D}_T} \sum_{t=1}^T \boldsymbol{\delta}_{t-1} \Psi_{Tt}^\delta (1 - \Psi_{Tt}^\delta) \leq O_{P_1}(T^{-1} + T^{-(\nu-2)/2})$, where $\boldsymbol{\delta}_{t-1} = \sum_{i=1}^k \delta_{t-i}$;
- $\sup_{\mathbb{D}_T} \|\Psi_T^{\alpha, \gamma} - (T^{1/2}(\hat{\alpha}_Y - 1), \hat{\gamma}'_Y)'\| = O_{P_1}(T^{\tau-1})$;
- $\sup_{\mathbb{D}_T} \sum_{t=1}^T \Psi_{Tt}^\delta |e_t^\Psi| = O_{P_1}(T^{1/2})$;
- $\sup_{\mathbb{D}_T} \|(\Psi_T^\theta - H_T, \Psi_T^\sigma - \sigma_\varepsilon^2)\| = O_{P_1}(T^{-1/2})$.

PROOF. Note that $\max_{t \leq T} |\varepsilon_t| = O_{P_1}(T^{1/4})$ since $E|\varepsilon_1|^4 < \infty$ under P_1 . Thus, the assumptions on ρ are not restrictions on the studied processes, since in the definition of \mathbb{D}_T arbitrarily small positive ρ can be chosen.

To get item (a), we apply Lemma 10 to $\hat{e}_t = e_t^d$, $\mathbb{A}_T = \mathbb{D}_T^\lambda \times \mathbb{D}_T^\sigma \times \mathbb{D}_T^\theta$ and $\zeta(x) = (x^\lambda + N_T, H_T x^\theta, x^\sigma)'$. To claim that the resulting evaluation of $\sum_{t=1}^T |\delta_t - \Psi_{Tt}^\delta|$ is uniform over the whole \mathbb{D}_T , we use that e_t^d do not depend on $(\lambda, \sigma^2, \theta^2)$ and have the properties required in Lemmas 8 and 10 uniformly on $\mathbb{D}_T^{\delta a \gamma} := \mathbb{D}_T^\delta \times \mathbb{D}_T^a \times \mathbb{D}_T^\gamma$.

First, $e_t^d - \varepsilon_t - T^{1/2} \delta_t \eta_t - (\gamma_* - \gamma)' \nabla \mathbf{Y}_{t-1}$ are evaluated. They equal $-T^{-1/2} a Y_{t-1} - T^{-1} c_* \Gamma_*(L) Y_{t-1} - (a, \gamma') A_T (\mathbf{X}_{t-1}^d - \mathbf{Y}_{t-1})$, where $|T^{-1/2} a| \leq O(T^{-1+\rho})$ and $\|(a, \gamma')\| \leq O(1)$ on $\mathbb{D}_T^a \times \mathbb{D}_T^\gamma$, while $\max_{t \leq T} |Y_{t-1}| = O_{P_1}(T^{1/2})$, and to $\|A_T (\mathbf{X}_{t-1}^d - \mathbf{Y}_{t-1})\|$ Lemma 11(c) with $d(x) = x^\delta + \delta$, $x \in \mathbb{D}_T$ and $\beta = \rho$ can be applied. Combining the orders of magnitude gives

$$\sup_{\mathbb{D}_T^{\delta a \gamma}} |e_t^d - \varepsilon_t - T^{1/2} \delta_t \eta_t - (\gamma_* - \gamma)' \nabla \mathbf{Y}_{t-1}| \leq O_{P_1}(T^\rho) \boldsymbol{\delta}_{t-1} + O_{P_1}(T^{-1/2+\rho+\tau}). \quad (25)$$

Since $\max_{t \leq T} \|\nabla \mathbf{Y}_{t-1}\| = O_{P_1}(T^\tau)$ by Lemma 6(b), it follows that $\sup_{\mathbb{D}_T^{\delta a \gamma}} |e_t^d - \varepsilon_t - T^{1/2} \delta_t \eta_t| = O_{P_1}(T^{1/2})$ given the assumptions on ρ and τ . This is a uniform version of the hypothesis of Lemma 8.

Second, we verify that $\sup_{\mathbb{D}_T^{\delta a \gamma}} \sum_{t=1}^T (1 - \delta_t) |e_t^d|^{\nu+1} = O_{P_1}(T)$ as required by Lemma 10. Indeed, according to (25) it holds uniformly on $\mathbb{D}_T^{\delta a \gamma}$ that $(1 - \delta_t) |e_t^d|^{\nu+1} \leq (1 - \delta_t) 4^\nu [|\varepsilon_t|^{\nu+1} + \|\gamma_* - \gamma\|^{\nu+1} \|\nabla \mathbf{Y}_{t-1}\|^{\nu+1} + (\boldsymbol{\delta}_{t-1})^{\nu+1} O_{P_1}(T^{1/2}) + o_{P_1}(1)]$ due to the choices $\rho(\nu + 1) \leq 1/2$ and $\rho + \tau < 1/2$. Further, with $\nu + 1 \leq r$ we have $E|\varepsilon_t|^{\nu+1} < \infty$, so that $\sum_{t=1}^T |\varepsilon_t|^{\nu+1} = O_{P_1}(T)$ and $\sum_{t=1}^T \|\nabla \mathbf{Y}_{t-1}\|^{\nu+1} = O_{P_1}(T)$ by Lemma 6(b), while $\sum_{t=1}^T (\boldsymbol{\delta}_{t-1})^{\nu+1} \leq k^{\nu+1} N_T = O_{P_1}(1)$. These give the required uniform order of $\sum_{t=1}^T (1 - \delta_t) |e_t^d|^{\nu+1}$. Having verified the conditions of Lemma 10, we obtain the orders of magnitude of the suprema in (a).

In item (b), $0 \leq \sum_{t=1}^T \delta_t \boldsymbol{\delta}_{t-1} \Psi_{Tt}^\delta (1 - \Psi_{Tt}^\delta) \leq N_T \sum_{t=1}^T \delta_t (1 - \Psi_{Tt}^\delta) \leq O_{P_1}(T^{-(\nu-2)/2})$ by (a). Thus, it remains to evaluate $\sum_{t=1}^T (1 - \delta_t) \boldsymbol{\delta}_{t-1} \Psi_{Tt}^\delta (1 - \Psi_{Tt}^\delta) \leq \sum_{t=1}^T (1 - \delta_t) \boldsymbol{\delta}_{t-1} \Psi_{Tt}^\delta$. From the first equation in (19) written for Ψ_{Tt}^δ and e_t^d , multiplied by $\boldsymbol{\delta}_{t-1}$, and summed over

t , the last summation is seen not to exceed

$$O_{P_1}(T^{-3/2}) \sum_{t=1}^T (1 - \delta_t) \boldsymbol{\delta}_{t-1} + O_{P_1}(T^{-3/2}) \sum_{t=1}^T (1 - \delta_t) \boldsymbol{\delta}_{t-1} |e_t^d|^{\nu+1},$$

where $\sum_{t=1}^T (1 - \delta_t) \boldsymbol{\delta}_{t-1} \leq kN_T = O_{P_1}(1)$ and $\sum_{t=1}^T (1 - \delta_t) \boldsymbol{\delta}_{t-1} |e_t^d|^{\nu+1}$ is dominated by

$$4^\nu \left[\max_{t: \boldsymbol{\delta}_{t-1} > 0} |\varepsilon_t|^{\nu+1} + \|\gamma_* - \gamma\|^{\nu+1} \max_{t: \boldsymbol{\delta}_{t-1} > 0} \|\nabla \mathbf{Y}_{t-1}\|^{\nu+1} \right] N_T + O_{P_1}(T^{\rho(\nu+1)}) (N_T)^{\nu+2} + o_{P_1}(1),$$

which is $O_{P_1}(T^{\rho(\nu+1)})$, $\rho(\nu+1) \leq 1/2$, according to Lemma 7. Combining the above orders of magnitude delivers item (b).

Item (c) with $o_{P_1}(1) = O_{P_1}(T^{\tau-1/2})$ follows from item (a) and Lemma 12 with $d(x) = \Psi_T^\delta(\xi(x))$, $\mathbb{G}_T = \mathbb{D}_T$. Further, item (c) and the standard properties of $(\hat{\alpha}_Y, \hat{\gamma}'_Y)'$ imply that $\sup_{\mathbb{D}_T} \|\Psi_T^{\alpha, \gamma}\| = O_{P_1}(1)$, $\sup_{\mathbb{D}_T} |\Psi_T^\alpha| = O_{P_1}(T^{\tau-1/2})$ and $\sup_{\mathbb{D}_T} \|\gamma_* - \Psi_T^\gamma\| = O_{P_1}(T^{-1/2})$. These are useful in a derivation similar to that of (25), but this time invoking for $\sup_{\mathbb{D}_T} \|A_T(\mathbf{X}_{t-1}^\Psi - \mathbf{Y}_{t-1})\|$ Lemma 11 with $d(x) = \Psi_T^\delta(\xi(x))$, $\mathbb{G}_T = \mathbb{D}_T$ and $\beta = 0$:

$$\begin{aligned} \sup_{\mathbb{D}_T} |e_t^\Psi - \varepsilon_t - T^{1/2} \delta_t \eta_t| &\leq O_{P_1}(1) \boldsymbol{\delta}_{t-1} + O_{P_1}(T^{\tau-1/2}) + \sup_{\mathbb{D}_T} \|\gamma_* - \Psi_T^\gamma\| \max_{t \leq T} \|\nabla \mathbf{Y}_{t-1}\| \\ &= O_{P_1}(1) \boldsymbol{\delta}_{t-1} + O_{P_1}(T^{\tau-1/2}), \end{aligned} \quad (26)$$

since $\max_{t \leq T} \|\nabla \mathbf{Y}_{t-1}\| = O_{P_1}(T^\tau)$ by Lemma 6(a). From here $\sup_{\mathbb{D}_T, t \leq T} |e_t^\Psi| = O_{P_1}(T^{1/2})$, and

$$\begin{aligned} \sum_{t=1}^T \Psi_{Tt}^\delta |e_t^\Psi| &\leq \sum_{t=1}^T |\Psi_{Tt}^\delta - \delta_t| |e_t^\Psi| + \sum_{t=1}^T \delta_t |e_t^\Psi| \\ &\leq \max_{t \leq T} |e_t^\Psi| (\sum_{t=1}^T |\Psi_{Tt}^\delta - \delta_t| + N_T) = O_{P_1}(T^{1/2}) (O_{P_1}(T^{-1/2}) + O_{P_1}(1)) \end{aligned}$$

uniformly on \mathbb{D}_T , from where item (d).

Finally, item (e) can be derived starting from

$$\begin{aligned} \sup_{\mathbb{D}_T} \left\| \left(\Psi_T^\theta, \Psi_T^\sigma \right) - T^{-1} \sum_{t=1}^T (\delta_t, 1 - \delta_t) (e_t^\Psi)^2 \right\| &= \sqrt{2} T^{-1} \sup_{\mathbb{D}_T} \left| \sum_{t=1}^T (\Psi_{Tt}^\delta - \delta_t) (e_t^\Psi)^2 \right| \\ &\leq \sqrt{2} T^{-1} \sup_{\mathbb{D}_T, t \leq T} (e_t^\Psi)^2 \sup_{\mathbb{D}_T} \sum_{t=1}^T |\Psi_{Tt}^\delta - \delta_t| = O_{P_1}(T^{-1/2}). \end{aligned}$$

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 (26) to evaluate $|u - v|$, it is seen that

$$\sup_{\mathbb{D}_T} \left\| T^{-1} \sum_{t=1}^T (1 - \delta_t, \delta_t) \left[(e_t^\Psi)^2 - (\varepsilon_t + T^{1/2} \delta_t \eta_t)^2 \right] \right\| = O_{P_1}(T^{2\tau-1}).$$

Inserting $\tau = r^{-1} \leq 1/4$ and combining with the previous display concludes the proof of (e). \blacksquare

Lemma 14 *Let Assumption $\mathcal{M}(b)$ be satisfied with $r \geq 5$, and let the arguments of Ψ_T be collected in $W = \{d_t, a, \gamma_i, \lambda, \theta^2, \sigma^2 : t = 1, \dots, T; i = 1, \dots, k\}$. Then, for $4 \leq \nu \leq r - 1$ and $\rho < 3/5$, $\|(\Psi_T)'_\xi\|_1 = \sum_{w \in W} \sum_{v \in W} |(\Psi_T^v)'_w|$ satisfies $\sup_{x \in \mathbb{D}_T} \|(\Psi_T)'_\xi|_{\xi=\xi(x)}\|_1 \leq o_{P_1}(1)$, where $\xi(x)$ is the function defined in the introduction to section A.4.*

In the proof the symbols $(\cdot)'_w$ and $\partial(\cdot)/\partial w$ are used interchangeably for Jacobian matrices, and preference is given to the latter symbol when there is a possibility of confusion with matrix transposition. All evaluations are uniform on \mathbb{D}_T , but for brevity sup-signs are omitted. The hypothesis with $r \geq 5$ implies that $\max_{t \leq T} |\varepsilon_t| = O_{P_1}(T^\tau)$ with $\tau = 1/5$.

PROOF. 1. *Derivatives of Ψ_{Tt}^δ* : $\|(\Psi_{Tt}^\delta)'_\xi\|_1 = \sum_{w \in W} \sum_{t=1}^T |(\Psi_{Tt}^\delta)'_w| \leq O_{P_1}(T^{\rho+\tau-1/2})$. In the expressions for the derivatives, $\Psi_{Tt}^\delta(1 - \Psi_{Tt}^\delta)$ appear frequently and are abbreviated to ψ_{Tt}^δ . As a corollary of Lemma 13(a), $\sum_{t=1}^T \psi_{Tt}^\delta \leq O_{P_1}(T^{-1/2})$, which is essential for obtaining convergence of the derivatives to 0.

It is checked directly that $(\Psi_{Tt}^\delta)'_\lambda = \psi_{Tt}^\delta T \lambda^{-1} (T - \lambda)^{-1}$, so that $\|(\Psi_{Tt}^\delta)'_\lambda\|_1 = \sum_{t=1}^T |(\Psi_{Tt}^\delta)'_\lambda| \leq O_{P_1}(T^{-1/2})$ on \mathbb{D}_T . Further, for an argument v different from λ ,

$$(\Psi_{Tt}^\delta)'_v = \psi_{Tt}^\delta (\ln f_\nu)'_v, \quad (27)$$

where $f_\nu = \phi_\nu(e_t^d; \sigma^2 + T\theta^2) / \phi_\nu(e_t^d; \sigma^2)$. The cases $v = (\sigma^2, \theta^2)$ and $v = (d', a, \gamma')$ are considered separately. First, for $v = (\sigma^2, \theta^2)$ the derivatives in $(\ln f_\nu)'_v$ are $O_{P_1}(1)$ on \mathbb{D}_T as in (24), and thus, $\sum_{t=1}^T \|(\Psi_{Tt}^\delta)'_v\| \leq O_{P_1}(1) \sum_{t=1}^T \psi_{Tt}^\delta \leq O_{P_1}(T^{-1/2})$. Second,

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

where, with $h(x, y) = x/y[1 + x^2/y]^{-1}$, the factor in front of the derivative is bounded in absolute value by $(\nu + 1) \{(\nu \sigma^2)^{-1/2} + [\nu(\sigma^2 + T\theta^2)]^{-1/2}\}/2$ due to the inequality $|h(x, y)| \leq (4y)^{-1/2}$, $y > 0$, and on \mathbb{D}_T does not exceed the constant $K = 2(\nu + 1)/\sqrt{\nu}/\sigma_\varepsilon^2$, while the derivative has components

$$(e_t^d)'_{(a, \gamma')} = A_T \mathbf{X}_{t-1}^d \text{ and } (e_t^d)'_{d_s} = (T^{-1/2} a, \gamma') \mathbb{I}_{ts} \Delta X_s, \quad s = 1, \dots, T, \quad (28)$$

with \mathbb{I}_{ts} the vector of indicators $(\mathbb{I}_{\{s \leq t-1\}}, \mathbb{I}_{\{s=t-1\}}, \dots, \mathbb{I}_{\{s=t-k\}})'$. Hence, recalling (27),

$$\|(\Psi_{Tt}^\delta)'_{(a, \gamma')}\|_1 = \sum_{t=1}^T \|(\Psi_{Tt}^\delta)'_{(a, \gamma')}\|_1 \leq K \max_{t \leq T} \|A_T \mathbf{X}_{t-1}^d\|_1 \sum_{t=1}^T \psi_{Tt}^\delta \leq O_{P_1}(T^{\tau-1/2})$$

since the maximum is uniformly bounded by $O_{P_1}(T^\tau)$, Lemma 11(c, $\beta = \rho$). Similarly, by retaining only terms corresponding to unit entries of \mathbb{I}_{ts} ,

$$\|(\Psi_{Tt}^\delta)'_d\|_1 = \sum_{t=1}^T \sum_{s=1}^T |(\Psi_{Tt}^\delta)'_{d_s}| \leq K \sum_{t=1}^T \psi_{Tt}^\delta \left[|a| T^{-1/2} \sum_{s \leq t-1} |\Delta X_s| + \|\gamma\| \sum_{s=t-k}^{t-1} |\Delta X_s| \right]. \quad (29)$$

Further details are worked out using $\Delta X_s = \Delta Y_s + T^{1/2} \delta_s \eta_s$ and $|\Delta X_s| \leq \max_{t \leq T} |\Delta Y_t| + T^{1/2} \delta_s |\eta_s|$. On the one hand, $\sum_{s \leq T-1} |\Delta X_s| \leq O_{P_1}(T^{\tau+1}) + T^{1/2} \sum_{t=1}^T \delta_s |\eta_s| = O_{P_1}(T^{\tau+1})$, so that $\sum_{t=1}^T \psi_{Tt}^\delta \sum_{s \leq t-1} |\Delta X_s| \leq O_{P_1}(T^{\tau+1}) \sum_{t=1}^T \psi_{Tt}^\delta \leq O_{P_1}(T^{\tau+1/2})$, and since $|a| \leq T^{-1/2+\rho}$ on \mathbb{D}_T , the contribution to (29) of terms involving this factor is $O_{P_1}(T^{\rho+\tau-1/2})$.

On the other hand, $\sum_{s=t-k}^{t-1} |\Delta X_s| \leq k \max_{t \leq T} |\Delta Y_t| + T^{1/2} \delta_{s-1} \max_{t: \delta_t=1} |\eta_t| = O_{P_1}(T^\tau) + \delta_{s-1} O_{P_1}(T^{1/2})$, so that $\sum_{t=1}^T \psi_{Tt}^\delta \sum_{s=t-k}^{t-1} |\Delta X_s| \leq O_{P_1}(T^\tau) \sum_{t=1}^T \psi_{Tt}^\delta + O_{P_1}(T^{1/2}) \sum_{t=1}^T \delta_{t-1} \psi_{Tt}^\delta \leq O_{P_1}(T^{\tau-1/2})$ by Lemma 13(b) and the choice $\nu \geq 4$. Combining the obtained orders of magnitude gives $\|(\Psi_{Tt}^\delta)'_\xi\|_1 \leq O_{P_1}(T^{\rho+\tau-1/2})$ uniformly on \mathbb{D}_T .

At a later step it is used that $\sum_{w \in W} \sum_{t=1}^T \delta_t |(\Psi_{Tt}^\delta)'_w| \leq O_{P_1}(T^{\rho-1/2})$: everywhere above $\sum_{t=1}^T \psi_{Tt}^\delta \leq O_{P_1}(T^{-1/2})$ becomes $\sum_{t=1}^T \delta_t \psi_{Tt}^\delta \leq \sum_{t=1}^T \delta_t |1 - \Psi_{Tt}^\delta| \leq O_{P_1}(T^{-(\nu-2)/2})$, $\nu \geq 4$.

2. Since $\Psi_T^\lambda = \sum_{t=1}^T \Psi_{Tt}^\delta$, it holds that $\sum_{w \in W} |(\Psi_T^\lambda)'_w| \leq \sum_{w \in W} \sum_{t=1}^T |(\Psi_{Tt}^\delta)'_w| \leq O_{P_1}(T^{\rho+\tau-1/2})$.

3. *Derivatives of $\Psi_T^{a,\gamma}$.* Recall that $\Psi_T^{a,\gamma} = [\sum_{t=1}^T A_T \mathbf{X}_{t-1}^\Psi (A_T \mathbf{X}_{t-1}^\Psi)']^{-1} \sum_{t=1}^T A_T \mathbf{X}_{t-1}^\Psi \Delta X_t^\Psi$. The components of $\Psi_T^{a,\gamma}$ depend on ξ only through X_t^Ψ . With $S_{XX}^\Psi = \sum_{t=1}^T A_T \mathbf{X}_{t-1}^\Psi (A_T \mathbf{X}_{t-1}^\Psi)'$ and with derivatives $\partial(\cdot)/\partial d_s$ that here happen not to depend on the point d they are evaluated at, it is checked that $(\Psi_T^{a,\gamma})'_w$ equals $[S_{XX}^\Psi]^{-1}$ postmultiplied by

$$\begin{aligned} & A_T \sum_{t=1}^T \sum_{s=1}^T \left\{ \frac{\partial \mathbf{X}_{t-1}^d}{\partial d_s} [\Delta X_t^\Psi - (A_T \mathbf{X}_{t-1}^\Psi)' \Psi_T^{a,\gamma}] + \mathbf{X}_{t-1}^\Psi \left[\frac{\partial \Delta X_t^d}{\partial d_s} - \frac{\partial (\mathbf{X}_{t-1}^d A_T)'}{\partial d_s} \Psi_T^{a,\gamma} \right] \right\} \frac{\partial \Psi_{Ts}^\delta}{\partial w} \\ &= \sum_{s=1}^T \left\{ A_T \sum_{t=s+1}^T [\mathbb{I}_{ts} [(A_T \mathbf{X}_{t-1}^\Psi)' \Psi_T^{a,\gamma} - \Delta X_t^\Psi] + \mathbf{X}_{t-1}^\Psi \mathbb{I}'_{ts} A_T \Psi_T^{a,\gamma}] - \mathbf{X}_{s-1}^\Psi \right\} \Delta X_s \frac{\partial \Psi_{Ts}^\delta}{\partial w} \end{aligned}$$

(a prime denotes transposition). Since $T^{-1} S_{XX}^\Psi$ converges to a non-singular limit (see the proof of Lemma 12), the order of magnitude of $(\Psi_T^{a,\gamma})'_w$ is determined by T^{-1} times the expression above. Thus, if Z_s denotes the term in braces in this expression, it follows that

$$\|(\Psi_T^{a,\gamma})'_w\| \leq O_{P_1}(T^{-1}) \sum_{s=1}^T \|Z_s\| |\Delta X_s| |(\Psi_{Ts}^\delta)'_w|, \quad (30)$$

where $\|Z_s\| \leq 2 \|\sum_{t=s+1}^T A_T \mathbb{I}_{ts} (A_T \mathbf{X}_{t-1}^\Psi)'\| \|\Psi_T^{a,\gamma}\| + \|\sum_{t=s+1}^T A_T \mathbb{I}_{ts} \Delta X_t^\Psi\| + \|A_T \mathbf{X}_{s-1}^\Psi\|$, and $\sum_{t=s+1}^T A_T \mathbb{I}_{ts} (A_T \mathbf{X}_{t-1}^\Psi)' = (T^{-1/2} \sum_{t=s+1}^T \mathbf{X}_{t-1}^\Psi, \mathbf{X}_s^\Psi, \mathbf{X}_{s+1}^\Psi \mathbb{I}_{\{s+1 \leq T\}}, \dots, \mathbf{X}_{s+k-1}^\Psi \mathbb{I}_{\{s+k-1 \leq T\}})' A_T$ has norm bounded by $(T^{1/2} + k) \max_{s \leq T} \|A_T \mathbf{X}_s^\Psi\|$, while the norm of $\sum_{t=s+1}^T A_T \mathbb{I}_{ts} \Delta X_t^\Psi = A_T (\sum_{t=s+1}^T \Delta X_t^\Psi, \Delta X_{s+1}^\Psi \mathbb{I}_{\{s+1 \leq T\}}, \dots, \Delta X_{s+k}^\Psi \mathbb{I}_{\{s+k \leq T\}})'$ does not exceed $(k+2) \max_{s \leq T} \|A_T \mathbf{X}_s^\Psi\|$. Hence, accounting also for the relation $\|\Psi_T^{a,\gamma}\| \leq O_{P_1}(1)$ implied by Lemma 13(c), it obtains that $\max_{s \leq T} \|Z_s\| \leq O_{P_1}(T^{1/2}) \max_{s \leq T} \|A_T \mathbf{X}_s^\Psi\|$. Inserting this into (30) together with $|\Delta X_t| \leq |\Delta Y_t| + T^{1/2} \delta_t |\eta_t|$ gives

$$\|(\Psi_T^{a,\gamma})'_w\| \leq O_{P_1}(T^{-1/2}) \max_{s \leq T} \|A_T \mathbf{X}_s^\Psi\| \left\{ \max_{s \leq T} |\Delta Y_s| \sum_{s=1}^T |(\Psi_{Ts}^\delta)'_w| + T^{1/2} \max_{t: \delta_s=1} |\eta_s| \sum_{s=1}^T \delta_s |(\Psi_{Ts}^\delta)'_w| \right\}.$$

Since $\max_{s \leq T} \|A_T \mathbf{X}_s^\Psi\| \leq O_{P_1}(T^\tau)$ (Lemma 11(c), $\beta = 0$), $\max_{s \leq T} |\Delta Y_s| = O_{P_1}(T^\tau)$ (Lemma 6) and $\max_{t: \delta_s=1} |\eta_s| = O_{P_1}(1)$ (Lemma 7), while $\sum_{w \in W} \|(\Psi_T^\delta)'_w\|_1 \leq O_{P_1}(T^{\rho+\tau-1/2})$ and $\sum_{w \in W} \sum_{s=1}^T \delta_s |(\Psi_{Ts}^\delta)'_w| \leq O_{P_1}(T^{\rho-1/2})$, it obtains that $\sum_{w \in W} \|(\Psi_T^{a,\gamma})'_w\| \leq O_{P_1}(T^{\rho+\tau-1/2})$. This property is shared by $\|(\Psi_T^{a,\gamma})'_\xi\|_1$.

4. *Derivatives of Ψ_T^θ and Ψ_T^σ .* First, $(\Psi_T^\theta)'_w = T^{-1} \sum_{t=1}^T (\Psi_{Tt}^\delta)'_w (e_t^\Psi)^2 + 2T^{-1} \sum_{t=1}^T \Psi_{Tt}^\delta e_t^\Psi (e_t^\Psi)'_w$, where

$$\begin{aligned} & |T^{-1} \sum_{t=1}^T (\Psi_{Tt}^\delta)'_w (e_t^\Psi)^2| \leq T^{-1} \max_{t \leq T} (e_t^\Psi)^2 \sum_{t=1}^T |(\Psi_{Tt}^\delta)'_w| = O_{P_1}(1) \|(\Psi_T^\delta)'_w\|_1, \text{ and} \\ & |T^{-1} \sum_{t=1}^T \Psi_{Tt}^\delta e_t^\Psi (e_t^\Psi)'_w| \leq T^{-1/2} \sum_{t=1}^T \Psi_{Tt}^\delta |e_t^\Psi| T^{-1/2} \max_{t \leq T} |(e_t^\Psi)'_w| \leq O_{P_1}(T^{-1/2}) \max_{t \leq T} |(e_t^\Psi)'_w|, \end{aligned}$$

the latter inequality by Lemma 13(d). From these evaluations and $\|(\Psi_T^\delta)'_\xi\|_1 \leq O_{P_1}(T^{\rho+\tau-1/2})$ it follows by summation over $w \in W$ that

$$\|(\Psi_T^\theta)'_\xi\|_1 \leq O_{P_1}(T^{-1/2}) \sum_{w \in W} \max_{t \leq T} |(e_t^\Psi)'_w| + O_{P_1}(T^{\rho+\tau-1/2}). \quad (31)$$

Thus, it remains to study $(e_t^\Psi)'_w$. With $W' = W \setminus \{\lambda, \sigma^2, \theta^2\}$, these are $(e_t^\Psi)'_w = \sum_{v \in W'} (e_t^d)'_v (\Psi_T^v)'_w$ and $(e_t^d)'_v$ are components of the derivatives (28) evaluated at $d = (\Psi_{T1}^\delta, \dots, \Psi_{TT}^\delta)'$ and $(a, \gamma)' = \Psi_T^{\alpha, \gamma}$. For any $w \in W$,

$$\begin{aligned} \left| \sum_{h=1}^T (e_t^d)'_{dh} (\Psi_{Th}^\delta)'_w \right| &= |T^{-1/2} \Psi_T^a \sum_{h \leq t-1} \Delta X_h (\Psi_{Th}^\delta)'_w + \sum_{h=t-k}^{t-1} (\Psi_T^\gamma)_{t-h} \Delta X_h (\Psi_{Th}^\delta)'_w| \\ &\leq \max_{h \leq T} |\Delta X_h| \left[T^{-1/2} |\Psi_T^a| \sum_{h=1}^T |(\Psi_{Th}^\delta)'_w| + \|\Psi_T^\gamma\| \sum_{h=t-k}^{t-1} |(\Psi_{Th}^\delta)'_w| \right], \end{aligned}$$

where $\max_{h \leq T} |\Delta X_h| \leq \max_{s \leq T} |\Delta Y_s| + T^{1/2} \max_{s, \delta_s=1} |\eta_s| = O_{P_1}(T^{1/2})$, $|\Psi_T^a| \leq O_{P_1}(T^{-1/2})$ and $\|\Psi_T^\gamma\| \leq O_{P_1}(1)$ on \mathbb{D}_T , so that

$$\sum_{w \in W} \max_{t \leq T} \left| \sum_{h=1}^T (e_t^d)'_{dh} (\Psi_{Th}^\delta)'_w \right| \leq O_{P_1}(T^{1/2}) [O_{P_1}(T^{-1}) + k \|\Psi_T^\gamma\|] \sum_{w \in W} \sum_{h=1}^T |(\Psi_{Th}^\delta)'_w| \leq O_{P_1}(T^{\rho+\tau}). \quad (32)$$

In view of the expression for $(e_t^\Psi)'_w$ and of (31),

$$\|(\Psi_T^\theta)'_\xi\|_1 \leq O_{P_1}(T^{-1/2}) \max_{t \leq T} \|(e_t^d)'_{(a, \gamma)} (\Psi_T^{\alpha, \gamma})'_\xi\|_1 + O_{P_1}(T^{\rho+\tau-1/2}), \quad (33)$$

where $(e_t^d)'_{(a, \gamma)}$ is read as a row vector. The maximum is bounded by $O_{P_1}(T^{\rho+2\tau-1/2})$ since it does not exceed $\|(\Psi_T^{\alpha, \gamma})'_\xi\|_1 \max_{t \leq T} \|A_T \mathbf{X}_{t-1}^\Psi\|_1$ and the two factors are bounded respectively by $O_{P_1}(T^{\rho+\tau-1/2})$ and $O_{P_1}(T^\tau)$. Hence, $\|(\Psi_T^\theta)'_\xi\|_1 \leq O_{P_1}(T^{\rho+2\tau-1} + T^{\rho+\tau-1/2}) = O_{P_1}(T^{\rho+\tau-1/2})$.

The order of magnitude of $\|(\Psi_T^\sigma)'_\xi\|$ follows from a modification of the argument for Ψ_T^θ . 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 summation has already been shown to be $O_{P_1}(T^{\rho+\tau-1/2})$. For the second one, the following evaluation can be used:

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

The maximum does not exceed $\max_{t: \delta_t=0} |e_t^\Psi| + \max_{t \leq T} |e_t^\Psi| \sum_{t=1}^T |\delta_t - \Psi_{Tt}^\delta| \leq O_{P_1}(T^\tau)$, see (26). Regarding the summation, instead of (32) and (33), now

$$\begin{aligned} \sum_{w \in W} \sum_{t=1}^T \left| \sum_{h=1}^T (e_t^d)'_{dh} (\Psi_{Th}^\delta)'_w \right| &\leq O_{P_1}(T^{1/2}) [O_{P_1}(1) + k \|\Psi_T^\gamma\|] \|(\Psi_T^\delta)'_\xi\|_1 = O_{P_1}(T^{\rho+\tau}), \\ \|(\Psi_T^\sigma)'_\xi\|_1 &\leq O_{P_1}(T^{\tau-1}) \sum_{t=1}^T \|(e_t^d)'_{(a, \gamma)} (\Psi_T^{\alpha, \gamma})'_\xi\|_1 + O_{P_1}(T^{\rho+\tau-1/2}), \end{aligned}$$

so that $\|(\Psi_T^\sigma)'_\xi\|_1 \leq T^\tau \|(\Psi_T^{a,\gamma})'_\xi\|_1 T^{-1} \sum_{t=1}^T \|A_T \mathbf{X}_{t-1}^\Psi\|_1 + O_{P_1}(T^{\rho+\tau-1/2}) = O_{P_1}(T^{\rho+2\tau-1/2})$ since $\sum_{t=1}^T \|A_T \mathbf{X}_{t-1}^\Psi\|_1 \leq \sum_{t=1}^T \|A_T \mathbf{Y}_{t-1}\|_1 + \sum_{t=1}^T |X_{t-1}^\Psi - Y_{t-1}| + k \sum_{t=1}^T |\Delta Y_t - \Delta X_t^\Psi| \leq O_{P_1}(T)$ by Lemma 11(a),(b) with $\beta = 0$.

To summarize, among the considered finitely many blocks of the Jacobian matrix the highest magnitude order is $O_{P_1}(T^{\rho+2\tau-1/2})$, which for $\rho < 3/5$ and $\tau = 1/5$ is $o_{P_1}(1)$. ■

A.4.2 Proof of Theorem 5

The argument is similar to that for Theorem 2. Let $\rho > 0$ be chosen sufficiently small for the hypotheses of Lemmas 13 and 14 to be satisfied. For $x \in \mathbb{D}_T$ and $\xi(x)$ defined in the introduction to section A.4, let $\Xi_T = \xi^{-1} \circ \Psi_T \circ \xi$; it is well-defined P_1 -a.s. Further, let $\Xi_T^{tr} : \mathbb{D}_T \rightarrow \mathbb{D}_T$ 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 $(\Xi_T^{tr})^\delta(x) = T^{\rho-1/2} \Xi_T^\delta(x) / \sum_{t=1}^T |\Xi_T^\delta(x)|$ 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 $\Xi_T^\lambda, \Xi_T^a, \Xi_T^\theta, \Xi_T^\sigma$ defined similarly to the components of Θ_T^{tr} in the proof of Theorems 2 and 4, see A.3. Then Ξ_T^{tr} has a random fixed point \varkappa_T on \mathbb{D}_T by Th. 10 in Bharucha-Reid (1976). For outcomes such that $\Xi_T(\varkappa_T)$ is interior for \mathbb{D}_T no truncation takes place in the calculation of Ξ_T^{tr} , and hence, $\Psi_T(\xi(\varkappa_T)) = \xi(\varkappa_T)$, i.e. $\xi_T(\varkappa_T)$ is a fixed point of Ψ_T . From Lemma 13(a),(c),(e) we have that $\Psi_T(\xi(\varkappa_T)) = \xi(\varkappa_\infty) + o_{P_1}(1)$ and $\Xi_T(\varkappa_T) = \varkappa_\infty + o_{P_1}(1)$, with $\varkappa_\infty := (0_{T+2}, \gamma'_*, 1, \sigma_\varepsilon^2)'$ belonging to the interior of \mathbb{D}_T . Moreover, according to the same lemma, $\Xi_T^\delta(\varkappa_T)$ and $\Xi_T^a(\varkappa_T)$ tend to zero at rates faster than the shrinkage rates of the components \mathbb{D}_T^δ and \mathbb{D}_T^a . Therefore, $\Xi_T(\varkappa_T)$ is interior for \mathbb{D}_T with P_1 -probability approaching 1, and hence, $P_1(\Psi_T(\xi(\varkappa_T)) = \xi(\varkappa_T)) \rightarrow 1$. The sequence ξ_T whose existence is asserted in (i) can now be defined as $\xi_T := \xi(\varkappa_T)$.

Items (ii) and (iii) follow from $\varkappa_T \in \mathbb{D}_T$ and Lemma 13. Item (d) follows once noticed that the ADF statistics based on the de-jumped data are $ADF_\alpha^\Psi = T^{1/2} \Psi_T^a(\xi_T) (1 - \sum_{i=1}^k \Psi_T^{\gamma_i}(\xi_T))^{-1}$ and $ADF_t^\Psi = \Psi_T^a(\xi_T) (\hat{\sigma}_\Psi^2(\xi_T) v_{11})^{-1/2}$, where v_{11} is the first entry of the matrix $[\sum_{t=1}^T \mathbf{X}_{t-1}^\Psi (\mathbf{X}_{t-1}^\Psi)']^{-1}$ evaluated at ξ_T . According to Lemma 12, $\varkappa_T \in \mathbb{D}_T$ implies that these statistics have the same asymptotic distribution as those based on ΔY_t .

Uniqueness is addressed next. Fix an arbitrary $\epsilon \in (0, 1)$. In view of Lemmas 14 and 7, $\|(\Xi_T)'_x\|_1 \leq \|(\Psi_T)'_\xi\|_1 \max(H_T, H_T^{-1}) \leq o_{P_1}(1)$ uniformly on \mathbb{D}_T . Therefore, we can find a set \mathcal{B}_1 with $P_1(\mathcal{B}_1) > 1 - \epsilon/4$ and an integer T_1 such that, for $T > T_1$ and outcomes in \mathcal{B}_1 , $\|(\Xi_T)'_x\|_1 \leq 1/2$ for all $x \in \mathbb{D}_T$. Say that $x_1, x_2 \in \mathbb{D}_T$ are fixed points of Ξ_T for some outcome ω . Since \mathbb{D}_T is convex, 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 $(\Xi_T)'_x$ each row of which is evaluated at some point in \mathbb{D}_T , in general varying across rows. From here and the fixed-point property, $\|x_1 - x_2\|_\infty = \|\Xi_T(x_1) - \Xi_T(x_2)\|_\infty \leq \|x_1 - x_2\|_\infty \sup_{x \in \mathbb{D}_T} \|(\Xi_T)'_x\|_1$, and for $\omega \in \mathcal{B}_1$ and $T > T_1$ it follows that $x_1 = x_2$, i.e. Ξ_T has at most one fixed point on \mathbb{D}_T . According the first part of the proof we can find a set $\mathcal{B}_2 \subset \mathcal{B}_1$ with $P_1(\mathcal{B}_2) > 1 - \epsilon/3$ and an integer $T_2 \geq T_1$ such that, on \mathcal{B}_2 and for $T > T_2$, \varkappa_T is the unique fixed point of Ξ_T on \mathbb{D}_T .

Further, if ξ_T^1 and ξ_T^2 satisfy item (i), there exist a set \mathcal{B}_3 with $P_1(\mathcal{B}_3) > 1 - \epsilon/3$ and an integer T_3 such that, on \mathcal{B}_3 and for $T > T_3$, it holds that $\Psi_T(\xi_T^i) = \xi_T^i$, $i = 1, 2$. If they also satisfy items (ii) to (iv), so that $\Psi_T(\xi_T^i) = \xi(\varkappa_\infty) + o_{P_1}(1)$ again at rate faster than the shrinkage rate for \mathbb{D}_T^δ and \mathbb{D}_T^a , then we can find a set \mathcal{B}_4 with $P_1(\mathcal{B}_4) > 1 - \epsilon/3$ and an integer T_4 such that, for $T > T_4$ and outcomes in \mathcal{B}_4 , $\xi^{-1}(\xi_T^i) \in \mathbb{D}_T$. On $\mathcal{B}_3 \cap \mathcal{B}_4$ and for $T > \max(T_3, T_4)$, $\xi^{-1}(\xi_T^i)$ are fixed points of Ξ_T on \mathbb{D}_T , while on $\cap_{i=2}^4 \mathcal{B}_i$ with $P_1(\cap_{i=2}^4 \mathcal{B}_i) > 1 - \epsilon$ and for

$T > \max(T_2, T_3, T_4)$, they must be equal to \varkappa_T and to each other. Since ϵ is arbitrary, uniqueness with P_1 -probability approaching one is proved.

Finally, consider computation. In view of Lemma 13, there exist a set \mathcal{B}_5 with $P_1(\mathcal{B}_5) > 1 - \epsilon/3$ and an integer T_5 such that, on \mathcal{B}_5 and for $T > T_5$, $\Xi_T(\mathbb{D}_T) \subset \mathbb{D}_T$. On the other hand, from Theorems 2 and 4, $x_{T0} := \xi^{-1}(\xi_{T0}) \xrightarrow{P_1} \varkappa_\infty$ at faster rate than the shrinkage rate for \mathbb{D}_T^δ and \mathbb{D}_T^a . Thus, on some $\mathcal{B}_6 \subset \mathcal{B}_5$ with $P_1(\mathcal{B}_6) > 1 - \epsilon/2$ and for T greater than some T_6 , we have that $x_{T0} \in \mathbb{D}_T$. From the first part of the proof we know that we can find $T_7 \geq T_2$ and $\mathcal{B}_7 \subset \mathcal{B}_2$ with $P_1(\mathcal{B}_7) > 1 - \epsilon/2$ such that, on \mathcal{B}_7 and for $T > T_7$, Ξ_T is Lipschitz with modulus $1/2$. Then, on $\mathcal{B}_6 \cap \mathcal{B}_7$ and for $T > \max(T_6, T_7)$, Ξ_T is a contraction of \mathbb{D}_T with unique fixed point \varkappa_T , and by Banach's fixed point theorem $x_{Ti} \rightarrow \varkappa_T$ as $i \rightarrow \infty$. Hence, also $\xi_{Ti} = \xi(x_{Ti}) \rightarrow \xi(\varkappa_T) = \xi_T$ as $i \rightarrow \infty$. ■

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
- Chang Y. and J.Y. Park (2002), On the asymptotics of ADF tests for unit roots, *Econometric Reviews* 21, 431-447.
- Dempster A., Laird N. and D. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* 39, 1-38
- Doornik J., D. Hendry and B. Nielsen (1998), Inference in Cointegrated Models: UK M1 Revisited, *Journal of Economic Surveys* 12, 533-572.
- 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.
- Hansen B.E. (1992), Convergence to stochastic integrals for dependent heterogeneous processes, *Econometric Theory* 8, 489–500.
- He S.W., J.G. Wang (1995), The convergence of partial sum processes of Markov chains, in Ma, Rockner, Yan e de Gruyter J.A (eds.), *Dirichlet Forms and Stochastic Processes*, 175–184.
- Jacod J. and A. Shiryaev (1987), *Limit Theorems for Stochastic Processes*, Springer.
- Kapetanios G. (2005), Unit-root testing against the alternative hypothesis of up to m structural breaks, *Journal of Time Series Analysis* 26, 123–133.
- 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 (2002new), Comparison of unit root tests for time series with level shifts, *Journal of Time Series Analysis* 23, 667–685.
- 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 (2001) Lag length selection and the construction of unit root tests with good size and power, *Econometrica* 69, 1519–1554.
- Nielsen, H. B. (2004), Cointegration analysis in the presence of outliers, *Econometrics Journal* 7, 249–271.

- 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.
- Phillips P.C.B. (1987), Toward a unified asymptotic theory for autoregression, *Biometrika* 74, 535–547.
- Saikkonen P. (2001), Consistent Estimation in Cointegrated Vector Autoregressive Models with Nonlinear Time Trends in Cointegrating Relations, *Econometric Theory* 17, 296-326
- 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.
- White H. (2001), *Asymptotic theory for econometricians*, Academic press
- 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.

T	γ	Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
		$ADF_{\hat{\alpha}}$	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^δ and ADF^Ψ denote the ADF tests under rough de-jumping and finer de-jumping. In both cases for estimation 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 power of standard ADF tests (ADF) and of the modified ADF tests (ADF^δ and ADF^Ψ). Raw data.

T	γ	Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
		$ADF_{\hat{\alpha}}$	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	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: Power is evaluated at $\alpha=1-c/T$ with $c=7$. See also Table 1.

Table 3. Empirical size of standard ADF tests (ADF) and of the modified ADF tests (ADF^δ and ADF^Ψ). GLS de-trended data

T	γ	Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
		$ADF_{\hat{\alpha}}$	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 4. Size-adjusted power of standard ADF tests (ADF) and of the modified ADF tests (ADF^δ and ADF^Ψ). GLS de-trended data.

T	γ	Model S_0 (no level shifts)						Model S_4 (four level shifts at fixed fractions)						Model S_r (random level shifts)					
		$ADF_{\hat{\alpha}}$	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	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 $\alpha=1-c/T$ with $c=-13.5$. See also Table 1.

Table 5. 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 $t(10)$ innovations

T	γ	$ADF_{\hat{\alpha}}$	ADF_t	$\nu = 8$				$\nu = \infty$			
				$ADF_{\hat{\alpha}}^\delta$	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	0.0	0.0	0.2	0.2	0.3	0.3	8.7	8.8	11.0	11.5
	0	0.5	0.5	1.3	1.3	1.5	1.5	17.0	16.9	17.4	17.6
	0.5	13.7	13.4	14.8	14.6	15.2	15.2	24.6	25.3	27.1	27.9
200	-0.5	0.0	0.0	0.8	0.7	1.6	1.6	17.1	16.9	20.4	20.7
	0	0.6	0.5	5.7	5.7	5.8	5.8	29.1	30.0	29.8	29.8
	0.5	13.7	13.1	18.6	18.3	19.8	20.2	35.2	35.1	36.7	37.7
400	-0.5	0.0	0.0	8.4	8.4	16.6	16.4	26.9	26.7	28.1	28.0
	0	0.5	0.4	26.9	27.4	27.1	27.7	36.4	36.5	36.5	36.7
	0.5	15.0	15.0	30.2	30.8	35.9	35.9	43.1	44.3	42.9	44.3

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