Robust Inference in Autoregressions with Multiple Outliers

Giuseppe Cavaliere
Università di Bologna

Iliyan Georgiev
Universidade Nova de Lisboa

August 2007

Abstract

We consider robust methods for estimation and unit root [UR] testing in autoregressions with innovational outliers whose number, size and location can be random and unknown. We show that in this setting standard inference based on OLS estimation of an augumented Dickey-Fuller [ADF] regression may not be reliable, since (i) clusters of outliers may lead to inconsistent estimation of the autoregressive parameters, and (ii) large outliers induce a jump component in the asymptotic null distribution of UR test statistics. In the benchmark case of known outlier location, we discuss why the augmentation of the ADF regression with appropriate dummy variables not only ensures consistent parameter estimation, but also gives rise to UR tests with significant power gains, growing with the number and the size of the outliers. In the case of unknown outlier location, the dummy based approach is compared with a robust, mixed Gaussian, Quasi Maximum Likelihood [QML] inference approach, novel in this context. It is proved that, when the ordinary innovations are Gaussian, the QML and the dummy based approach are asymptotically equivalent, yielding UR tests with the same asymptotic size and power. Moreover, the outlier dates can be consistently estimated as a by-product of QML. When the innovations display tails fatter than Gaussian, the QML approach seems to ensure further power gains over the dummy based method. A number of Monte Carlo simulations show that the QML ADF-type t-test, in conjunction with standard Dickey-Fuller critical values, yields the best combination of finite sample size and power.

1 Introduction

Over the past decade econometricians have seriously entertained the question of how to improve the power of autoregressive [AR] unit root [UR] and cointegration tests. A first major strand of this literature draws on the seminal paper by Elliott et al. (1996), who show that massive power improvement can be obtained by considering point-optimal tests against a fixed alternative. A second, important strand of this literature focuses on the distributional properties of the data, in two respects. First, in the presence of non-Gaussian data – mainly, excess kurtosis – the asymptotic power envelope generally differs from the Gaussian envelope. Important papers in this area are Lucas (1995a,b), Rothenberg and Stock (1997), Hodgson (1998a,b), Abadir and Lucas (2004), Boswijk (2005) and Jansson (2006). Second, econometric techniques based on M estimation, including non-Gaussian quasi maximum likelihood [QML], may benefit from substantial power gains over Gaussian QML inference methods; see Lucas (1995a,b), Franses and Lucas (1998), Lucas (1997, 1998), Franses et al. (1999) and Boswijk and Lucas (2002). An attempt to compare the two strands of the literature is made by Thompson (2004).

A prominent case of departure from the Gaussian framework arises when data are characterized by innovational outliers [IO].[1] The effects of outlying events on UR and cointegration testing have been extensively studied in the literature; see, inter alia, Perron (1989, 1990), Perron and Vogelsang (1992), Fransen and Haldrup (1994), Lucas (1997), Lanne et al. (2002), Bohn Nielsen (2004) and Xiao and Lima (2004); see also Burridge and Taylor (2006) for a recent reference. Rothenberg and Stock (1997, p.282) implicitly consider an innovational outlier model and show that Gaussian QML inference leads to UR tests with power far below the power envelope. Also Lucas (1995b) clearly shows that in the IO case there is room for power gains when UR tests are based on the optimization of non-Gaussian criterion functions. In particular, the robust QML methods proposed in Lucas (1997,1998), Franses and Lucas (1998) and Franses et al. (1999) allow one to obtain important power gains in the presence of innovational outliers.

The good efficiency and power properties of robust QML techniques somewhat contrast with the 'common practice' of accounting for IOs through the inclusion of impulse dummies in the model; see, among many others, Box and Tiao (1975), Hendry and Juselius (2001) and Bohn-Nielsen (2004). The dummy-variable approach can be viewed as an extreme case of robust inference methods, where outlying observations – given that the outlier dates are known to the econometrician – are implicitly eliminated by the inclusion of the dummies. Nevertheless, as far as we are aware, no study has been undertaken in order to assess whether a dummy-based approach to estimation and UR inference in the presence of outliers allows to obtain power gains comparable with those of the robust procedures proposed in the literature.

A first aim of this paper is to answer the previous question. In particular, by using both asymptotic arguments and Monte Carlo simulations, we aim at showing that, when the ordinary shocks are Gaussian, the dummy-based approach is comparable to robust inference methods, both in terms of size and power. This result suggests that the use of appropriate dummy variables may represent a compelling way to increase the power of UR tests, in view of the further advantages that (i) no new critical values are needed, and (ii) it allows the practitioner to address the economic interpretation of the outlying events.

Given that the inclusion of impulse dummies is in general unfeasible in practice (unless the dates of the outlying events are known to the econometrician), we discuss a robust QML estimator that allows one to construct UR tests with the same asymptotic size and power properties as the UR tests obtained using the dummy variables approach. Hence, the new robust QML tests benefit from the power gains associated to the latter (unfeasible) approach. Moreover, the robust QML method delivers estimators of the model parameters which are asymptotically unaffected by outliers of relevant size. The QML estimator weights each observation according to how likely it is an outlier to have occurred at the corresponding date. In contrast with the dummy variables approach, no *a priori* information on either the location or the number of outliers is required, as QML implicitly performs consistent estimation of the dates where outliers occur. In this respect, a further contribution of this approach is that it bridges the gap between the robust statistics approach, which, similarly to ours, requires no identification of the outlier dates and applies continuous weights to the observations, and the (unfeasible) dummy variable approach.

[1] Following, inter alia, Franses and Lucas (1998) and Lanne et al. (2002), we focus on IOs, as, with respect to other types of outliers (e.g., additive), they are more likely to affect economic and financial time series, see e.g. Lucas (1995b, p.169).

A second aim of the paper is to shed some light on the mechanics behind the power gains under local alternatives. To accomplish it, we compare the large sample representations of standard UR test statistics and of statistics constructed using dummy variables. We argue that power gains are due to the intuitive fact that impulse dummy variables account for the effect of outliers on the first differences of the data, but not for the long run effect on the levels. By asymptotic equivalence, the same conclusion applies to the QML estimator. Furthermore, this result possibly applies to robust estimators in general, as they tend to downweight the observations corresponding to periods with large innovations, while remaining sensitive to the long-run effect of such innovations. Notice that, consistently with this conclusion, in the case of additive outliers, where the long run effect on the levels is zero, the use of dummy variables and QML leads to no power gains, similarly to what Lucas (1995b) found about other robust approaches.

Finally, we show that, in the (empirically relevant[2]) case where outliers cluster together, the coefficients of the stable regressors of the reference AR model may not be estimated consistently by OLS, with the unfortunate consequence that the usually employed AR estimators of the long run variance are not necessarily consistent. The proposed robust QML approach is also able to fix this problem, as it restores Gaussian asymptotic inference on the short-run coefficients.

The outlier model we consider is quite different from those considered in the earlier literature, in several respects. Specifically, under this model, (i) outliers occur randomly over time; (ii) the number of outliers is unknown, and only needs to be bounded in probability; (iii) outliers need not occur independently over time and, in particular, may cluster together; (iv) the sizes of the outliers are random and of larger magnitude order than the ordinary shocks driving the AR dynamics; (v) outliers do not need to be independent of the ordinary shocks.

Notice that (i)-(v) above are rather general. No restrictions or *a priori* knowledge of the number or the location of the outliers is assumed. Differently from a strand of the literature where the number of outliers diverges with the sample size (cf. Balke and Fomby, 1991; Franses and Haldrup, 1994), here this number is kept bounded, hence allowing us to distinguish between frequent, ordinary shocks and rare, outlying events. A further important feature of our model is that outliers are large in size, when compared to the ordinary shocks. This allows us to develop an asymptotic framework that renders the outliers asymptotically influential, both under the UR null hypothesis and under the alternative, cf. Leybourne and Newbold (2000a,b) and Müller and Elliott (2003).

The structure of the paper is as follows. In section 2 we present the reference model and its assumptions. In section 3 we discuss how outliers affect the asymptotic distributions of the standard OLS estimator of the model parameters and of the associated standard UR tests. In section 4 we turn to the analysis of the dummy-based approach under the assumption that the outlier dates are known. Finite sample comparisons are reported in section 5. The robust QML approach and the resulting UR tests are proposed and analyzed in sections 6 (asymptotic properties) and 7 (finite sample simulation). Section 8 extends the QML approach to general deterministic time trends. Some concluding comments are collected in section 9. All proofs are placed in the Appendix. The following notation is used: '$\xrightarrow{w}$' denotes weak convergence and '$\xrightarrow{P}$' convergence in P-probability, with $O_P(1)$ denoting boundedness

[2]Cf. Balke and Fomby (1994, section 4.2).

in P-probability; $\mathbb{I}(\cdot)$ is the indicator function; $\mathbf{I}_k$ and $\mathbf{1}_k$ are the $k \times k$ identity matrix and the $k \times 1$ vector of ones. With '$x := y$' ('$x =: y$') we indicate that x is defined by y (y is defined by x), and $\lfloor \cdot \rfloor$ signifies the largest integer not greater than its argument. With $\mathcal{D}$ we denote the space of càdlàg functions on $[0,1]$, endowed with the Skorohod topology. For a vector $x \in \mathbb{R}^n$, $\|x\| := (x'x)^{1/2}$ stands for its Euclidean norm, whereas for a matrix A, $\|A\| := [tr(A'A)]^{1/2}$, where $tr(\cdot)$ is the trace operator. For brevity, integrals such as $\int_0^1 X(s\text{-})\, dY(s)$ and $\int_0^1 X(s)\, Y(s)\, ds$ are written as $\int XdY$ and $\int XY$, respectively.

2 The model

We consider parameter estimation and tests of the UR null hypothesis $\mathsf{H}_0 : \alpha = 1$ against local alternatives $\mathsf{H}_c : \alpha = 1 - c/T$ ($c > 0$) and fixed stable alternatives $\mathsf{H}_s : \alpha = \alpha^*$ ($|\alpha^*| < 1$), in the model

$$\begin{aligned} y_t &= \alpha y_{t-1} + u_t, \qquad & t = 1-k, ..., T, \\ u_t &= \textstyle\sum_{i=1}^k \bar{\gamma}_i u_{t-i} + \varepsilon_t + \delta_t \theta_t, \qquad & t = 1, ..., T, \end{aligned} \tag{1}$$

where, for $k \geq 1$, $(u_0, ..., u_{1-k}, y_{-k})'$ may be any random vector (for $k = 0$, y_0 may be any random scalar) whose distribution is fixed and independent of T. The model is completed with Assumptions $\mathcal{M}$ and $\mathcal{S}$ below.

Assumption $\mathcal{M}$. (a) The roots of $\bar{\Gamma}(z) := 1 - \sum_{i=1}^k \bar{\gamma}_i z^i$ have modulus greater than 1; (b) $\{\varepsilon_t\}_{t=1}^\infty$ is IID$(0, \sigma_\varepsilon^2)$, with $\sigma_\varepsilon^2 > 0$.

Assumption $\mathcal{M}$ prevents y_t from being I(2) or seasonally integrated, and ensures that the so-called long-run variance of u_t, hereafter $\sigma^2 := \sigma_\varepsilon^2 \bar{\Gamma}(1)^{-2}$, is well-defined.

The term $\delta_t \theta_t$ in (1) is the outlier component of the model. Specifically, δ_t is an unobservable binary random variable indicating the occurrence of an outlier at time t, with θ_t being the associated (random) outlier size. The (random) number of outliers is given by $N_T := \sum_{t=1}^T \delta_t$. The following condition is imposed $\{\delta_t, \theta_t\}$.

Assumption $\mathcal{S}$. (a) N_T is bounded in probability conditionally on $N_T \geq 1$; (b) $\theta_t = T^{1/2}\eta_t$, where $\{\eta_t\}_{t=1}^T$ and $\{\eta_t^{-1}\}_{t=1}^T$ are $O_P(1)$ sequences as $T \to \infty$; (c) for all T, $\{\delta_t\}_{t=1}^T$ is independent of $\{\varepsilon_t\}_{t=1}^T$,$\{\eta_t\}_{t=1}^T$, y_{-k} and, if $k \geq 1$, of $(u_0, ..., u_{1-k})'$.

For illustrative purposes, we will sometimes strengthen Assumption $\mathcal{S}$ by requiring that the following condition holds.

Assumption $\mathcal{S}'$. Assumption $\mathcal{S}$ holds and, as $T \to \infty$, $C_T(\cdot) := T^{-1/2} \sum_{t=1}^{\lfloor T \cdot \rfloor} \theta_t \delta_t \xrightarrow{w} C(\cdot)$, where C is a piecewise constant process in $\mathcal{D}$.

Remark 2.1. Assumption $\mathcal{S}$ allows us to generalize the single outlier model in several directions. For instance, the number of outliers N_T, instead of being fixed, is only assumed to be bounded in probability. Furthermore, we do not restrict the dependence structure of $\{\delta_t\}$, allowing e.g. for outliers at consecutive dates.

Remark 2.2. By Assumption $\mathcal{S}$(b) the outliers have the same stochastic magnitude order as the levels of y_t under H_0 or H_c. In particular, the effect of outliers does not become negligible in large samples. A similar assumption has been advocated by Perron (1989, p.1372) and employed by Leybourne and Newbold (2000a,b). The magnitude order $T^{1/2}$ has also been used by Müller and Elliott (2003) to model the size of the initial observation of an AR process with a root near to unity (notice that the initial observation can be thought of as a large outlier occurring at the beginning of the sample).

Remark 2.3. Assumption $\mathcal{S}$(c) rules out dependence between the outlier indicators $\{\delta_t\}$ and $\{\varepsilon_t, \eta_t\}$. However, it should be stressed that this is not a strictly necessary assumption for the results of the paper, and is made mainly for technical convenience. For instance, $\mathcal{S}$(c) could be replaced by the assumption that, conditionally on the occurrence of at least one outlier, the quantities $\max_{t:\delta_t=1} |\varepsilon_t| := \max_{t\leq T} |\delta_t\varepsilon_t|$, $\max_{t:\delta_t=1} |\eta_t|$ and $\max_{t:\delta_t=1} |\eta_t^{-1}|$ are bounded in probability.

Remark 2.4. Conditionally on the occurrence of at least one outlier, the smallest jump of the outlier partial-sum process C_T is bounded away from zero in probability; see Assumption $\mathcal{S}$(b). Thus, if the occurrence of at least one outlier has non-vanishing probability (the case where our asymptotic analysis is non-trivial), the tightness condition in Billingsley (1968, Theorem 15.2) implies that C_T has a limit in $\mathcal{D}$ only if the time distance between outliers diverges at the rate of T. Therefore, Assumption $\mathcal{S}'$ rules out, e.g., outliers occurring in adjacent periods, at least in large samples. A simple setup where Assumption $\mathcal{S}'$ is satisfied obtains when $\{\delta_t\}$ is an IID sequence of Bernoulli random variables with $p_T := P(\delta_t = 1) = \lambda/T$, $T > \lambda > 0$, and $\{\eta_t\}$ is an IID sequence as well. In this case the limiting process C is a compound Poisson process with jump intensity λ; see Georgiev (2006).

Remark 2.5. Since $\{\delta_t\}$, $\{\theta_t\}$ and, under H_c, also α of (1) depend on T, we are formally considering a triangular array format for $Y_{T,t}, \delta_{T,t}, \theta_{T,t}$. Unless differently specified, to keep notation simple we drop the 'T' subscript. □

In the analysis of model (1), the following alternative parameterization will be used. Let $\gamma := (\gamma_1, ..., \gamma_k)'$ and $\Gamma = (\pi, \gamma')'$, where, under H_0 and H_c, $\pi := 0$ and $\gamma_i := \bar{\gamma}_i$ $(i = 1, ..., k)$ whereas under H_s the new parameters are defined through the identity $(1 - \alpha z)\bar{\Gamma}(z) = 1 - (\pi + 1)z - \sum_{i=1}^k \gamma_i z^i(1 - z)$. Then Δy_t has the representation

$$\Delta y_t = \pi y_{t-1} + \gamma'\nabla\mathbf{Y}_{t-1} + e_t = \Gamma'\mathbf{Y}_{t-1} + e_t, \qquad t = 1, ..., T, \tag{2}$$

where $\nabla\mathbf{Y}_{t-1} := (\Delta y_{t-1}, ..., \Delta y_{t-k})'$ and $\mathbf{Y}_{t-1} := (y_{t-1}, \nabla\mathbf{Y}_{t-1}')'$. Under H_0 and H_s this is a regression with error term $e_t = \varepsilon_t + \delta_t\theta_t$, whereas under H_c it is an approximate regression whose error term differs from $\varepsilon_t + \delta_t\theta_t$ infinitesimally (see section A.1 of the Appendix). In view of Assumption $\mathcal{M}$, under H_0 or H_c the components of $\nabla\mathbf{Y}_{t-1}$ will be referred to as stable regressors, whereas under H_s the components of $\mathbf{Y}_{t-1}$ will be referred to as such.

3 ADF estimation and testing in the presence of outliers

In this section we discuss the effects of outlying events on the OLS estimator and on the related UR tests in the AR model (1) under the assumptions introduced in the previous section. Recall that ADF tests are based on OLS estimation of the regression equation,

$$\Delta y_t = \pi y_{t-1} + \gamma'\nabla\mathbf{Y}_{t-1} + \mathrm{error}_t, \tag{3}$$

and build on the statistics $ADF_\alpha := T\hat{\pi}/|\hat{\Gamma}(1)| = T(\hat{\alpha} - 1)/|\hat{\Gamma}(1)|$ and $ADF_t := \hat{\pi}/s(\hat{\pi})$, where $\hat{\Gamma}(1) := 1 - \sum_{i=1}^k \hat{\gamma}_i$ (with $\hat{\gamma} := (\hat{\gamma}_1, ..., \hat{\gamma}_k)'$ denoting the OLS estimator of γ), and

$s(\hat{\pi})$ is the (OLS) standard error of $\hat{\pi}$. Under Assumption $\mathcal{M}$ and for $\alpha = 1 - c/T$ ($c \geq 0$), it is well known (see e.g. Chang and Park, 2002, section 3) that in the standard case of no outliers, $\hat{\pi} \xrightarrow{P} 0$ and $\hat{\gamma} \xrightarrow{P} \gamma$. Moreover, the ADF statistics admit the representation

$$ADF_\alpha = -c + \frac{\int B_{c,T} dB_T}{\int B_{c,T}^2} + o_P(1), \qquad ADF_t = -c(\textstyle\int B_{c,T}^2)^{1/2} + \frac{\int B_{c,T} dB_T}{(\int B_{c,T}^2)^{1/2}} + o_P(1), \tag{4}$$

where $B_{c,T}$ of (4) lies in $\mathcal{D}$ and is defined as

$$B_{c,T}(s) := T^{-1/2}\sigma_\varepsilon^{-1} \sum_{i=0}^{\lfloor Ts \rfloor - 1} (1 - c/T)^i \, \varepsilon_{\lfloor Ts \rfloor - i}, \tag{5}$$

and $B_T := B_{0,T}$. Using $B_c(s) := \int_0^s e^{-c(s-z)} dB(z)$ to denote an Ornstein-Uhlenbeck process, B being a standard Brownian motion, when $T \to \infty$ we have that (Phillips, 1987) $B_{c,T} \xrightarrow{w} B_c$, and that

$$ADF_\alpha \xrightarrow{w} -c + \frac{\int B_c dB}{\int B_c^2}, \qquad ADF_t \xrightarrow{w} -c(\textstyle\int B_c^2)^{1/2} + \frac{\int B_c dB}{(\int B_c^2)^{1/2}}. \tag{6}$$

Under the null hypothesis that $c = 0$, $B_c = B$ and the distributions in (6) are the so-called univariate Dickey-Fuller distributions.

We now turn to the analysis of the OLS approach in the presence of multiple outliers, starting from the coefficients of the stable regressors in (3). Specifically, in the following proposition we present some sufficient and necessary conditions for consistent estimation of these coefficients.

Proposition 1 *Let $\tau_T := \min_{1 \leq i < j \leq T} \{j - i : \delta_i \delta_j = 1\}$ denote the smallest time distance between two consecutive outliers, and ∞, if at most one outlier occurs. Then, under Assumptions $\mathcal{M}$ and $\mathcal{S}$, the following results hold as $T \to \infty$.*

a. A sufficient condition for $\hat{\gamma} \xrightarrow{P} \gamma$ (and under H_s, for $\hat{\pi} \xrightarrow{P} \pi$) is that either $\gamma = 0$ (and under H_s, also $\pi = 0$), or $\tau_T \xrightarrow{P} \infty$.

b. If $\gamma \neq 0$ (or under H_s, $\pi \neq 0$), then for $\hat{\gamma} \xrightarrow{P} \gamma$ (and under H_s, for $\hat{\pi} \xrightarrow{P} \pi$) it is necessary that $\tau_T \xrightarrow{P} \infty$ conditionally on:

- the occurrence of exactly *two outliers, if the probability of this event is bounded away from zero;*

- the occurrence of at least *two outliers, if the probability of this event is bounded away from zero, and the variables $\{\eta_t\}$ are jointly independent and non-degenerately distributed.*

Remark 3.1. In the presence of short-run dynamics (i.e., $\gamma \neq 0$) and outliers of non-negligible size, the coefficients $\gamma_1, ..., \gamma_k$ (and π under H_s) associated to the stable regressors $\Delta y_{t-1}, ..., \Delta y_{t-k}$ (and y_{t-1} under H_s) may not be estimated consistently. This result has serious implications on the usual UR testing practice, as it implies that spectral AR estimators of the long run variance such as those suggested in, *inter alia*, Berk (1974), Stock (1994), Chang and Park (2002) and Ng and Perron (2001) may be inconsistent.

Remark 3.2. A condition that ensures consistent estimation of the short run coefficients $\gamma_1, ..., \gamma_k$ (and π under H_s), whatever the number and the size of the outliers are, is that the distance between consecutive outliers diverges with the sample size; see part (a). The condition is obviously satisfied in the case of a single outlier and, according to Remark 2.4, also under Assumption $\mathcal{S}'$. Notice that many econometric techniques for dealing with multiple structural breaks (see Bai and Perron, 1998; Perron, 2005) require the distance between consecutive break dates to diverge with the sample size (that is, $\tau_T \to \infty$ in the notation of Proposition 1).

REMARK 3.3. In the presence of short-run dynamics, the condition $\tau_T \xrightarrow{P} \infty$ becomes necessary for the consistency of $\hat{\gamma}$ (and $\hat{\pi}$ under H_s) under quite general circumstances, involving the occurrence of multiple outliers. The two parts of point (b) are intended to illustrate this claim. For instance the first part of (b) shows that in cases where two outliers occur, consistent estimation of γ through a simple ADF regression is not possible if the distance between the two outliers does not diverge with T.[3] □

For the discussion of the asymptotic properties of the UR tests, it is useful to define the following process in $\mathcal{D}$:

$$C_{c,T}(s) := T^{-1/2} \sum_{i=0}^{\lfloor Ts \rfloor -1} (1-c/T)^i \delta_{\lfloor Ts \rfloor - i} \theta_{\lfloor Ts \rfloor - i},$$

and let $H_{T,c} := B_{c,T} + C_{c,T}/\sigma_\varepsilon$, with $B_{c,T}$ as defined in (5) ($C_{0,T}$ and $H_{0,T}$ will be abbreviated as C_T and H_T, respectively). Should no outliers occur, $H_{c,T} = B_{c,T}$. Notice that if Assumption $\mathcal{S}'$ holds, then $C_{c,T}$ has a weak limit in $\mathcal{D}$; specifically, $C_{c,T} \xrightarrow{w} C_c$, with $C_c(s) := \int_0^s e^{-c(s-z)} dC(z)$ (cf. Kurtz and Protter, 1991, Theorem 2.7). In the latter case, $H_{c,T} \xrightarrow{w} H_c$, where H_c is the jump diffusion $H_c := B_c + C_c/\sigma_\varepsilon$.

We may now obtain large-sample representations of the ADF statistics in the presence of outliers, both under the null hypothesis and under local alternatives. The representations are formulated in terms of the finite-sample process $H_{c,T}$, similarly to (4), because in general the ADF statistics need not have weak limits under Assumption $\mathcal{S}$.

Proposition 2 *Let Assumptions $\mathcal{M}$ and $\mathcal{S}$ be satisfied. Then under H_0 or H_c, $c > 0$, the following results hold as $T \to \infty$.*

a. The ADF statistics have the representation

$$\begin{aligned} ADF_\alpha &= \frac{\Gamma(1)}{|\hat{\Gamma}(1)|}\Big(-c + \frac{\int H_{c,T} dH_T + \varkappa_{0,T}}{\int H_{c,T}^2} \Big) + o_P(1), \\ ADF_t &= \frac{1}{\varkappa_{1,T}^{1/2}}\Big(-c(\textstyle\int H_{c,T}^2)^{1/2} + \frac{\int H_{c,T} dH_T + \varkappa_{0,T}}{(\int H_{c,T}^2)^{1/2}} \Big) + o_P(1), \end{aligned}$$

where the expressions for $\varkappa_{0,T}$ and $\varkappa_{1,T}$ are given in the Appendix, eqs. (A.9) and (A.12).

[3] Still, it is possible to find particular configurations of multiple outliers where consistency obtains although $\tau_T = 1$ for all T. For example, if (i) $k = 0$, (ii) the autoregression is stable, (iii) $\delta_{\lfloor T/3 \rfloor} = \delta_{\lfloor T/3 \rfloor + 1} = \delta_{\lfloor T/2 \rfloor} = \delta_{\lfloor T/3 \rfloor + 1} = 1$ (all other being equal to zero), and (iv) $\eta_{\lfloor T/3 \rfloor} = -\eta_{\lfloor T/3 \rfloor + 1} = \eta_{\lfloor T/2 \rfloor} = \eta_{\lfloor T/3 \rfloor + 1} = 1$ (all other being irrelevant), then a necessary and sufficient condition for consistency (see eq. (A.11) in the Appendix) is satisfied, due to the particular degenerate distribution of η_t.

b. A necessary and sufficient condition for $\varkappa_{0,T} = o_P(1)$ is that $\hat{\gamma} \xrightarrow{P} \gamma$; in this case $\varkappa_{1,T} = 1 + \sigma_\varepsilon^{-2} \sum_{t=1}^{T} \delta_t \eta_t^2$, and

$$ADF_\alpha = -c + \frac{\int H_{c,T} dH_T}{\int H_{c,T}^2} + o_P(1) \ , \qquad ADF_t = \frac{1}{\varkappa_{1,T}^{1/2}} \Big(-c(\textstyle\int H_{c,T}^2)^{1/2} + \frac{\int H_{c,T} dH_T}{(\int H_{c,T}^2)^{1/2}} \Big) + o_P(1).$$

Several remarks are due.

REMARK 3.4. Differently from the standard case, see eq. (4), in the presence of outliers the null and local-to-null representations of the ADF statistics involve the process $H_{c,T}$ (i.e., both the errors ε_t and the outliers θ_t) instead of $B_{c,T}$ alone. Moreover, the contribution of θ_t is asymptotically non-negligible, see also Remark 3.6 below. Unless γ is consistently estimated, also the short-run dynamics has an asymptotically non-negligible effect on the ADF statistics.

REMARK 3.5. In representations (a) and (b), the process $H_{c,T}$ appears both as integrand and as integrator in the term $\int H_{c,T} dH_{c,T}$. An intuitive explanation is that when the standard ADF regression is employed to construct UR tests, then (i) outliers have a 'long run' effect, as they affect (through cumulation) the levels of y_t, hence implying that $H_{c,T}$ appears as integrand; (ii) outliers have a 'short run' effect, as they affect the errors of the ADF regression, hence implying that $H_{c,T}$ appears as integrator.

REMARK 3.6. Under Assumption $\mathcal{S}'$ it holds that $\hat{\gamma} \xrightarrow{P} \gamma$, see Remark 3.2. In this case, a corollary of Proposition 2 is that

$$ADF_\alpha \xrightarrow{w} -c + \frac{\int H_c dH}{\int H_c^2} \ , \qquad ADF_t \xrightarrow{w} \frac{1}{(1 + \sigma_\varepsilon^{-2}[C_c])^{1/2}} \Big(-c(\textstyle\int H_c^2)^{1/2} + \frac{\int H_c dH}{(\int H_c^2)^{1/2}} \Big), \quad (7)$$

where $[\cdot]$ denotes quadratic variation at unity.[4] These asymptotics generalize those obtained in the standard case of no outliers, cf. Stock (1994) *inter alia*. Specifically, the distributions in (7) have the same structure as the univariate Dickey-Fuller distributions, see (6), but with B_c replaced by the jump-diffusion H_c. The asymptotic distribution of the t statistic also depends on $\sigma_\varepsilon^{-2}[C_c]$, which measures the relative importance of the outliers with respect to the innovation variance. Notice also that the result (7) generalizes in several direction Theorem 1 in Leybourne and Newbold (2000a), where the case of a single fixed outlier occurring at a fixed (relative) date is considered under H_0 and in the absence of short run dynamics ($k = 0$ in eq. (1)).

REMARK 3.7. It is not hard to see that, under fixed stable alternatives, a sufficient condition for $ADF_\alpha \xrightarrow{P} -\infty$ and $ADF_t \xrightarrow{P} -\infty$, is that $\hat{\pi}$ is negative with probability approaching one and, in particular, that π is estimated consistently. It is, however, possible to construct examples where clusters of outliers, especially if close to the end of the sample, can create spurious explosiveness. The estimation methods discussed in the sections below are immune to this problem. □

In contrast to the common belief that innovational outliers do not affect inference in autoregressions with a possible unit root (see e.g. Shin et al., 1996, and Bohn-Nielsen, 2004),

[4] Convergence follows from the continuous mapping theorem, from Theorem 2.7 of Kurtz and Protter (1991) and from the well-known result that $\int B_{c,T} dB_{c,T} \xrightarrow{w} \int B_c dB_c$.

the results of this section suggest that innovational outliers of large size actually *do* affect the asymptotic properties of autoregression estimation and UR testing. Notice that this result is in line with previous findings for stationary time series: for instance, Tsay (1988) clearly recognizes that 'The effect of multiple IOs, (...), could be serious'.

A further, more important result, is that the presence of outliers, when properly accounted for, may be exploited in order to boost the power of UR tests. This crucial issue is investigated in the next section.

4 Dummy variables accounting for outliers

In this section we examine estimation and UR testing based on an ADF regression augmented with the inclusion of one impulse dummy variable for each outlier. Unless in cases where the outlier indicators δ_t are observable, see Lütkepohl et al. (2001) and Lanne et al. (2002) for a discussion, the results of the section are mostly of theoretical interest, and serve as a benchmark for the estimator we introduce in section 6. The key result we provide is that, by properly accounting for the outliers, not only is it possible to ensure consistent parameter estimation, but also to boost the power of UR tests beyond that attainable under standard conditions.

The 'dummy augmented' ADF regression has the form

$$\Delta y_t = \pi y_{t-1} + \gamma' \nabla \mathbf{Y}_{t-1} + \varphi' \mathbf{D}_t + \text{error}_t, \tag{8}$$

where $\mathbf{D}_t := (D_{1,t}, ..., D_{N_T,t})'$ is the vector of impulse dummies, one for each outlier. The ADF tests are based on the statistics $ADF_\alpha^D := T\tilde{\pi}/|\tilde{\Gamma}(1)|$ and $ADF_t^D := \tilde{\pi}/s(\tilde{\pi})$, where the superscript '$\sim$' now indicates that estimates are computed upon the inclusion of the vector of dummy indicators in the ADF regression.

As in (2), let $\Gamma := (\pi, \gamma')'$ and $\mathbf{Y}_{t-1} := (y_{t-1}, \nabla \mathbf{Y}'_{t-1})'$. The dummy variable estimators of Γ and σ_ε^2 are given by

$$\begin{aligned} \tilde{\Gamma} &:= \Big(\sum_{t=1}^{T} (1-\delta_t)(\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}) \Big)^{-1} \sum_{t=1}^{T} (1-\delta_t)(\mathbf{Y}_{t-1}\Delta y_t) \\ \tilde{\sigma}_\varepsilon^2 &:= \Big(\sum_{t=1}^{T} (1-\delta_t) \Big)^{-1} \sum_{t=1}^{T} (1-\delta_t)(\Delta y_t - \tilde{\Gamma}'\mathbf{Y}_{t-1})^2 \end{aligned} \tag{9}$$

As $\sum_{t=1}^{T} \delta_t = O_P(1)$, the inverses in both lines are well-defined with probability approaching one. The counterpart of Propositions 1 and 2 for the dummy ADF approach is given next.

Proposition 3 *Let Assumptions $\mathcal{M}$ and $\mathcal{S}$ be satisfied. Then the following results hold as $T \to \infty$.*

a. $\tilde{\gamma} \xrightarrow{P} \gamma$*, and under* H_s, $\tilde{\pi} \xrightarrow{P} \pi$.

b. Under H_0 *or* H_c, $c > 0$*, the ADF statistics have the following representation:*

$$ADF_\alpha^D = -c + \frac{\int H_{c,T} dB_T}{\int H_{c,T}^2} + o_P(1) \ , \quad ADF_t^D = -c(\textstyle\int H_{c,T}^2)^{1/2} + \frac{\int H_{c,T} dB_T}{(\int H_{c,T}^2)^{1/2}} + o_P(1) \ . \tag{10}$$

c. Under H_s, $ADF_\alpha^D \xrightarrow{P} -\infty$ *and* $ADF_t^D \xrightarrow{P} -\infty$.

Remark 4.1. In contrast with Proposition 1, upon the inclusion of a set of impulse dummy variables (one for each outlier) the estimator of the short-run parameters is consistent, even in the case of clustering outliers. As a consequence, under H_0 or H_c the ADF^D statistics are asymptotically independent of the short-run dynamics (i.e., of $\gamma_1, ..., \gamma_k$), while under H_s UR tests based on these statistics are consistent.

Remark 4.2. Similarly to standard ADF tests, see Remark 3.5, also when impulse dummies are included in the estimated regression, the large-sample representations of the ADF statistics involve the process $H_{c,T}$ instead of $B_{c,T}$ alone. However, now $H_{c,T}$ appears as an integrand only, and not as an integrator. The reason is that the inclusion of the dummy variables cancels the short run effect of the outliers, but not their long run effect on the levels of y_t.

Remark 4.3. Under assumption $\mathcal{S}'$, from Proposition 3 it follows that the dummy-based ADF^D statistics have asymptotic distributions

$$ADF_\alpha^D \xrightarrow{w} -c + \frac{\int H_c dB}{\int H_c^2} \quad \text{and} \quad ADF_t^D \xrightarrow{w} -c(\textstyle\int H_c^2)^{1/2} + \frac{\int H_c dB}{(\int H_c^2)^{1/2}}.$$

under the null and local alternatives. □

A further important issue about UR testing in ADF regressions which incorporate impulse dummies is related to the power of UR tests. Specifically, since the dummy approach is a special case of the robust approach (where the effect of outlying observations is trimmed down by adding impulse dummies to the estimated model), we expect it to benefit from the power gains featured by the robust approaches to UR testing in the presence of non-Gaussian data (Lucas, 1995, 1997). To shed some more light on this intuition, we now carry out an analytical experiment where the influence of outliers is taken to the extreme. A related exercise is made by Lucas (1995b, p.156-7) for the case of a single outlier with fixed location.

Let Assumption $\mathcal{S}'$ hold, implying that the ADF statistics have limiting distributions. These distributions were given in Remarks 3.5 and 4.3, and are now collected in the second column of Table 1, the first column reporting the standard case where no outliers occur. In the limiting distributions we replace the process C by hC, and let $h \to \infty$, conditioning on the occurrence of at least one outlier. This is a simple way to make the process C dominant in the limit. The obtained h-limits are collected in the third column of Table 1; details on their derivation are provided in the Appendix.

[Table 1 about here]

The following points can be made about this analysis.

Remark 4.4. The most striking qualitative difference in the h-limits occurs under local alternatives. Whatever the critical value is, under local alternatives the probability of rejecting the UR null hypothesis converges to 1 as $h \to \infty$ if the dummy-based ADF_t^D statistic is used, and the same holds for the coefficient statistic ADF_α^D if $-c$ is smaller than the critical value. This is in contrast with the standard OLS-based statistics ADF_α, ADF_t, whose corresponding rejection probabilities are bounded away from 1. It suggests that, in the presence of outliers, the dummy-based tests can have an advantage in terms of power over the standard tests, with power gains increasing with the size (and possibly with the number) of outliers.

REMARK 4.5. The power gains of the ADF_t^D test are formally due to the fact that outliers, through the long-run effect process C_c, make $\int H_c^2$ large, which upon the inclusion of dummy variables is not offset by an analogous effect on the estimator of the residual variance. A similar phenomenon occurs with the ADF_α^D test. This means that, in terms of power, we have no interest in eliminating the long-run effect of outliers from the asymptotic distributions. For this reason we do not discuss estimation with step dummy variables, which do cancel the long-run effect of outliers and (as is well known from the UR literature under trend breaks, cf. Perron, 2005) may cause a power loss.

REMARK 4.6. In terms of size, if standard Dickey-Fuller asymptotic critical values (see Fuller, 1976) are used, the ADF_t^D test can be expected to behave better than the ADF_α^D test, which may be undersized. This is because in the h-limit ADF_t^D approaches a $N(0,1)$ distribution (assuming independence of $B(\cdot)$ and $C(\cdot)$), whereas the coefficient statistic ADF_α^D tends to 0. Regarding the size of standard ADF tests, their size distortions are expected to decrease as the number of outliers increases, since the terms $(\int C^2)^{-1/2}\int CdC$ and $(\int C^2)^{-1}\int CdC$ equal 0 for a single outlier (implying 0 size as $h \to \infty$), and have distribution approaching the Dickey-Fuller counterparts $(\int B^2)^{-1/2}\int BdB$ and $(\int B^2)^{-1}\int BdB$ when the number of outliers grows. □

5 Finite sample comparisons

In this section we present a Monte Carlo study of standard and dummy-based ADF tests under a variety of innovation outlier models. Specifically, we want to assess whether (i) the power gains predicted in the previous section for the dummy-based tests are of relevant magnitude in finite samples, and (ii) size distortions for inference based on DF asymptotic critical values are substantial.

The employed DGPs are as follows. Data are generated for sample sizes of $T = 100, 200, 400$ observations according to model (1) with $k = 1$, $\bar{\gamma} := \bar{\gamma}_1 \in \{-0.5, 0, 0.5\}$, $y_0 = 0$ and u_0 drawn from the stationary distribution induced by the equation $v_t = \bar{\gamma} v_{t-1} + \varepsilon_t$. We consider the UR case, which obtains by setting $\alpha = 1$ in (1), and the sequence of local alternatives $\alpha = 1 - c/T$ with $c := 7$.

In addition to the case of no outliers ($\delta_t = 0$ for all t) – denoted with S_0 in the following – we consider four models for the outlier component:

- S_2 (two fixed outliers): two outliers occurring at fixed sample fractions t_i, $i = 1, 2$, with $t_1 := \lfloor 0.2T \rfloor$ and $t_2 := \lfloor 0.6T \rfloor$, and with size magnitudes $\theta_{t_1} := -0.4T^{1/2}$ and $\theta_{t_2} := 0.35T^{1/2}$;
- S_4 (four fixed outliers): four outliers occurring at fixed sample fractions t_i, $i = 1, ..., 4$, with t_1,t_2 as in S_2 above, $t_3 := \lfloor 0.4T \rfloor$ and $t_4 := \lfloor 0.8T \rfloor$; the corresponding size magnitudes are $\theta_{t_1}, \theta_{t_2}$ as in S_2 above, $\theta_{t_3} := -0.35T^{1/2}$ and $\theta_{t_4} := -0.4T^{1/2}$.
- S_r (random outliers): the number of outliers is $N_T \sim 3 + B(7/T, T)$, ($B(\cdot,\cdot)$ denoting a Binomial distribution), i.e. at least 3 and 10 on average, and their positions $\delta_1, ..., \delta_{N_T}$ are independent uniformly distributed on $\{1, ..., T\}$; the outlier magnitudes sizes θ_t are independent and distributed as a Gaussian r.v. with mean 0 and variance $0.09T$.

- S_c (cluster of three outliers): three consecutive outliers at positions $t_1 := \lfloor T/2 \rfloor$, $t_2 = t_1 + 1$, $t_3 = t_1 + 2$, all of magnitude $-0.35T^{1/2}$.

For our selection of T, models S_2, S_4 and S_c generate outliers of size between 4 and 8 standard deviations of the ordinary shocks. For model S_r, the random size of the outliers has standard deviation between 3 and 6 times the standard deviation of the ordinary shocks. These outlier magnitudes, although large, are not unrealistic; see the discussion in Vogelsang and Perron (1998, p.1090).

The innovations are zero-mean, unit-variance IID r.v. following either a $N(0,1)$ distribution or a standardized $t(5)$ distribution.

We consider both standard ADF tests (ADF_α,ADF_t) and the dummy-augmented tests (ADF_α^D,ADF_t^D), the latter being based on the assumption that the outlier locations are known. All tests are performed at the 5% (asymptotic) nominal level, with critical values taken from Fuller (1976, Tables 10.A.1 and 10.A.2). Computations are based on 10,000 Monte Carlo replications and are carried out in Ox v. 3.40, Doornik (2001). Results are reported in Table 2 (Gaussian errors) and in Table 3 (Student t errors).

[Tables 2–3 about here]

The following facts are worth noting.

(i) For outlier models S_2, S_4 and S_r, under which the representations in Proposition 2(b) hold, the presence of outliers does not seem to affect the *size* of standard ADF tests. This is in line with, e.g., the findings of Lucas (1995, Table 1). On the other hand, for model S_c, under which outliers cluster together, the size of ADF tests appears to be bounded away from the nominal level. The tests tend to be undersized (resp. oversized) for negative (resp. positive) values of γ. This dependence on the short run dynamics agrees with the representations in Proposition 2(a).

(ii) The presence of outliers does not substantially affect the size of the dummy-based ADF_t^D test, even when outliers cluster together. In all the cases considered, size is about 5%. In contrast, outliers do affect the size of the ADF_α^D test, which appears to be undersized. This finding is in line with the predictions based on the h-limits of the previous section.

(iii) The local (size-adjusted) power of ADF tests is slightly affected by the outliers, especially in small samples. For models S_2, S_4 and S_r power is generally below the approximate 50% power characterizing the tests in the absence of outliers. Interestingly, when outliers cluster together (model S_c), ADF tests display power slightly above 50%. This is of little practical importance, however (given the size distortions of ADF tests, the empirical rejection frequencies drop to as low as 25% for $\gamma = -0.5$). In general, there are no significant differences between the ADF_α and the ADF_t tests.

(iv) The use of impulse dummies *substantially* increases the local power, again as predicted previously. The power gains increase with the number of outliers. For instance, under model S_2 the addition of the dummy variables increases the local power of ADF tests from about 50% (no outliers) to above 60%. Under S_4, power increases to above 75%. In general, the ADF_t^D test performs slightly better than the ADF_α^D test in terms of local power. Differences between ADF_t^D and ADF_α^D tests, however, becomes substantial when the empirical rejection frequencies are considered, mainly because the ADF_α^D test is undersized. These results show that the ADF_t^D test is largely preferable over the ADF_α^D test.

(v) The dummy-based tests perform very well under model S_c (a cluster of outliers), again as predicted by the theoretical analysis of section 4. Although the ADF_α^D test has slightly higher power than the ADF_t^D test, in terms of the empirical rejection frequencies the latter test is clearly more appealing.

(vi) Results for the case of t innovations do not substantially differ from those obtained in the Gaussian case.

In summary, our Monte Carlo experiment shows that the inclusion of dummy variables which account for the short run effects of outliers is an important device for boosting the power of unit root tests. The ADF_t^D statistic used in conjunction with standard critical values, gives rise to a test with good size properties and with considerably higher power than the standard ADF tests which neglect the presence of outliers. As far as we are aware, these power gains have not been discussed extensively in the literature.

With respect to robust inference methods, an obvious drawback of the dummy-based approach is that it is unfeasible in practice, except in cases where the outlier dates are known. In the next section we will obtain a feasible t test based on a robust QML procedure, and will discuss an important connection between this robust method and the dummy based approach.

6 Robust QML estimation and UR testing

In this section we discuss a robust inference technique, based on Quasi Maximum Likelihood [QML], for autoregression estimation and UR testing. In contrast with the dummy-based approach, QML can be used when there is no *a priori* information on either the location or the number of outliers, mainly because QML implicitly involves consistent estimation of the outlier dates. In addition, our robust method attains the same asymptotic power gains as the (unfeasible) dummy-based estimators discussed earlier.

The proposed robust inference method is based on a quasi likelihood which places more probability mass in the tails of the error distribution. As is standard in outlier robust statistics, each observation is implicitly 'reweighted' on the basis of how likely it fits the postulated model (cf. Lucas, 1996, Ch.1): the less an observation fits the model, the less weight is assigned to that particular observation. In this respect, our QML is close to the robust techniques advocated in Lucas (1997), Franses and Lucas (1998), Lucas (1998) and Franses et al. (1999). On the other hand, our approach differs in several directions. First, the quasi distribution of the innovations is a mixture distribution, where the two mixing components have different orders of magnitude. This allows us to study robustification with respect to outliers of relevant size. Second, we provide a full asymptotic analysis of both parameter estimators and the corresponding UR test statistics. Finally, we are able to establish the relation between our robust inference method and the unfeasible dummy variable approach.

In the next subsection the estimator is defined; its asymptotic analysis is reported in subsection 6.2. The finite sample properties are analyzed in section 7.

6.1 Definition

Our robust QML method builds on the observation that the innovation term of the reference model, see eq. (1), has a mixture distribution, with mixing variable δ_t and mixture components ε_t (when $\delta_t = 0$) and $\varepsilon_t + \theta_t$ (when $\delta_t = 1$). Notice that in Assumption $\mathcal{S}$ no parametric hypothesis on the joint process $\{\varepsilon_t, \theta_t\}$ is made. Nevertheless, it is still possible to jointly estimate the outlier indicators and the parameters of interest in a QML framework.

Specifically, consider a QML estimator based on the following 'quasi distributions': (i) the innovations ε_t are normally distributed; (ii) the outlier indicators δ_t are Bernoulli random variables with $P(\delta_t = 1) = \lambda/T$, $T > \lambda > 0$; (iii) the outlier magnitudes η_t are Gaussian with mean 0 and finite variance σ_η^2; (iv) $\{\varepsilon_t\}$, $\{\delta_t\}$ and $\{\eta_t\}$ are IID and mutually independent. Notice that (i)-(iv) do not necessarily hold in general under Assumption $\mathcal{S}$.

Let $\theta := (\Gamma', \sigma_\varepsilon^2, \sigma_\eta^2, \lambda)'$. Under (i)–(iv) and conditional on the initial values, the quasi likelihood function is, up to an additive constant, given by

$$\Lambda(\theta) := \sum_{t=1}^{T} \ln\Big(\frac{\lambda}{T} l_t(\theta, 1) + \big(1 - \frac{\lambda}{T}\big) l_t(\theta, 0)\Big), \tag{11}$$

where

$$l_t(\theta, i) := \frac{1}{(\sigma_\varepsilon^2 + Ti\sigma_\eta^2)^{1/2}} \exp\Big(-\frac{(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{2(\sigma_\varepsilon^2 + Ti\sigma_\eta^2)}\Big), \qquad i = 0, 1.$$

In the following we will make use of the weights

$$d_t(\theta) := \frac{\lambda l_t(\theta, 1)}{\lambda l_t(\theta, 1) + (T - \lambda) l_t(\theta, 0)}, \tag{12}$$

which under (i)–(iv) correspond to the expectation of δ_t (i.e., to the probability of occurrence of an outlier at time t) conditional on the data.

By equating to zero the derivatives of $\Lambda(\theta)$ and rearranging terms we find the normal equations

$$\theta = \Phi(\theta) \tag{13}$$

where $\Phi := (\Phi^\Gamma, \Phi^\varepsilon, \Phi^\eta, \Phi^\lambda)' : \mathbb{R}^{k+4} \to \mathbb{R}^{k+4}$ is the random map with components

$$\Phi^\Gamma(\theta) := \sum_{t=1}^{T} w_t(\theta)(\Delta y_t \mathbf{Y}_{t-1}') \Big[\sum_{t=1}^{T} w_t(\theta)(\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}')\Big]^{-1}, \qquad \Phi^\lambda(\theta) := \sum_{t=1}^{T} d_t(\theta)$$

$$\Phi^\varepsilon(\theta) := \frac{\sum_{t=1}^{T}(1 - d_t(\theta))(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{\sum_{t=1}^{T}(1 - d_t(\theta))}, \qquad \Phi^\eta := \frac{\sum_{t=1}^{T} d_t(\theta)(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{T\sum_{t=1}^{T} d_t(\theta)} - \frac{\sigma_\varepsilon^2}{T}$$

and $w_t(\theta) := d_t(\theta)/(\sigma_\varepsilon^2 + T\sigma_\eta^2) + (1 - d_t(\theta))/\sigma_\varepsilon^2$.

A QML estimator could be computed, e.g., by iterating the map Φ in (13). After the QML estimates are computed, the ADF statistics obtain as $ADF_\alpha^Q := T\check{\pi}/|\check{\Gamma}(1)|$ and $ADF_t^Q := \check{\pi}/s(\check{\pi})$, where $s(\check{\pi}) := \{[\sum_{t=1}^{T} w_t(\check{\theta})(\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}')]^{-1}\}_{11}^{1/2}$.

Remark 6.1. If $\check{\theta}$ is a stationary point of Λ such that $\{d_t(\check{\theta})\}$ are sufficiently close to $\{\delta_t\}$, then $\check{\theta}$ could be expected to be close to the dummy-variables estimator $\tilde{\theta} := (\tilde{\Gamma}', \tilde{\sigma}_\varepsilon^2, \tilde{\sigma}_\eta^2, \tilde{\lambda})'$, with $\tilde{\Gamma}$ and $\tilde{\sigma}^2_\varepsilon$ defined in (9), and $\tilde{\sigma}^2_\eta := \sum_{t=1}^T \delta_t(\Delta y_t - \tilde{\Gamma}'\mathbf{Y}_{t-1})^2/(TN_T)^{-1}$ (conditionally on $N_T \geq 1$), $\tilde{\lambda} := N_T$. Since $\{\delta_t\}$ is unobservable, $\tilde{\theta}$ is empirically unfeasible; however, its relationship with $\check{\theta}$ is useful in the asymptotic analysis of $\check{\theta}$, see the next section.

Remark 6.2. The quasi likelihood function could be based on a mixture of non-Gaussian distribution for ε_t and η_t; e.g., a mixture of Student t distributions. This would allow the asymptotic analysis in the next section to be carried out without assuming normality of ε_t. Extensive Monte Carlo simulations have shown that in practice the normality assumption allows to obtain good results under a various range of distributions for the errors. Thus, for ease of exposition, we stick to the Gaussian distribution in what follows. □

6.2 Asymptotic analysis

In this section we discuss various asymptotic results for the QML approach. Asymptotics are derived under the assumption that the errors ε_t are normally distributed; deviations from normality are investigated by Monte Carlo simulation in the next section.

First, we discuss the properties of the QML estimator $\check{\theta}$ of the parameter θ, and its relation to the dummy-based OLS estimator discussed in section 4. In addition, we discuss an important by-product of the QML approach; that is, an associated estimator of the outlier indicators based on the weights $d_t(\check{\theta})$.

The main results are presented in the following theorem, where with a subscript '0' we denote the true parameter values.

Theorem 1 *Let Assumptions $\mathcal{M}$ and $\mathcal{S}$ be satisfied, with $\{\varepsilon_t\}$ being normally distributed. Let P denote the induced probability measure conditional on the occurrence of at least one outlier. Introduce also $D_T := \mathrm{diag}(T^{-1/2}, 1, ..., 1)$ under H_0 or H_c, and $D_T = \mathbf{I}_{k+1}$ under stable alternatives, H_s. Then there exists a random $(k+4)\times 1$-vector sequence $\check{\theta}_T$ (abbreviated to $\check{\theta}$) with the following properties as $T \to \infty$.*

a. $\check{\theta}$ is a local maximizer of $\Lambda(\theta)$ with P-probability approaching one.

b. $\sum_{t=1}^T |d_t(\check{\theta}) - \delta_t| = O_P(T^{\rho-1/2})$ for all $\rho > 0$.

c. $T^{1/2}D_T^{-1}(\check{\Gamma} - \Gamma_0) = T^{1/2}D_T^{-1}(\tilde{\Gamma} - \Gamma_0) + o_P(1)$;

d. $(\check{\lambda}, \check{\sigma}^2_\varepsilon, \check{\sigma}^2_\eta) = (N_T, \sigma^2_{\varepsilon 0}, Q_T) + o_P(1)$, where $Q_T := N_T^{-1}\sum_{t=1}^T \delta_t \eta_t^2$.

Some remarks are in order.

Remark 6.3. By part (a), we refer to $\check{\theta}$ as a QML estimator. As is also the case with other robust approaches, see e.g. Lucas (1995b), the quasi likelihood function may have multiple local maximizers, and parts (b) to (d) refer to one which is sufficiently close to the true value. Differently from Lucas (1995b), we prove the existence of such a maximizer instead of assuming it; notice, however, that we use more specific assumptions than Lucas (1995b). Possible multiplicity of maximizers created no difficulties in the simulations of section 7.

Remark 6.4. According to part (b) of Theorem 1, the sequence $\{d_t(\check{\theta})\}$ is a consistent estimator of the outlier indicators $\{\delta_t\}$. This estimator is not binary, rather, $d_t(\check{\theta})$ can be interpreted as measuring how likely it is an outlier to have occurred in period t, given the data. A binary estimator can be constructed by setting $\check{d}_t := \mathbb{I}(d_t(\check{\theta}) > \kappa)$ for some $\kappa \in (0,1)$, or for a sequence κ_T such that $1 - \kappa_T = O_P(T^{\rho-1/2})$ for some $\rho > 0$. By inverting $d_t(\cdot)$, this estimator can be written in the form $\check{d}_t := \mathbb{I}(|\Delta y_t - \check{\Gamma}'\mathbf{Y}_{t-1}| > \phi(\check{\theta}))$ for some threshold $\phi(\check{\theta})$, which is the traditional form of a residual-based outlier detection rule (see e.g. Tsay, 1988, and Chang et al., 1998). Theorem 1(b) implies that $\hat{d}_t$ are consistent for δ_t in the sense that, with probability approaching one, $\hat{d}_t = \delta_t$ for all $t = 1, ..., T$. Although this is an important by-product of our QML approach, it is worth stressing that the QML approach itself does not require the choice of any threshold for its implementation.

Remark 6.5. The main result is given in part (c) of the Theorem, where it is asserted that $D_T^{-1}(\check{\Gamma} - \tilde{\Gamma}) = o_P(T^{-1/2})$, $\tilde{\Gamma}$ being the (consistent) dummy-based estimator of the autoregressive parameter Γ, see eq. (2). This means that the QML estimator $\check{\Gamma}$ is asymptotically equivalent to $\tilde{\Gamma}$. In particular, $\check{\Gamma}$ is also consistent for Γ, and asymptotic inference on Γ is the same in the QML and the dummy-based approach. This statement is made more precise in Corollary 1 below.

Remark 6.6. Part (d) of the theorem states that the estimators $\check{\lambda}$, $\check{\sigma}_\varepsilon^2$ and $\check{\sigma}_\eta^2$ are consistent respectively for the number of outliers N_T, for the variance of the ordinary shocks $\sigma_{\varepsilon 0}^2$, and for the sample second moment of the outlier sequence, $N_T^{-1} \sum_{t=1}^T \delta_t \eta_t^2$. □

We are now ready to formulate the inferential implications of Theorem 1.

Corollary 1 *Under the conditions of Theorem 1 and under the measure P introduced there:*
a. $ADF_\alpha^Q = ADF_\alpha^D + o_P(1)$ *and* $ADF_t^Q = ADF_t^D + o_P(1)$*;*
b. $\check{\gamma} \xrightarrow{P} \gamma$ *and, if* $\{\varepsilon_t\}_{t=1}^T$ *is independent of* $\{\delta_t \eta_t\}_{t=1}^T$*, then* $\check{\gamma}$ *is asymptotically Gaussian. Under* H_s*, the same result holds for* $\check{\pi} \xrightarrow{P} \pi$*.*

Remark 6.7. According to Corollary 1, in the presence of outliers of the very general form defined through Assumption $\mathcal{S}$, the QML approach delivers ADF UR tests with the same asymptotic properties as obtained by using the unfeasible dummy-augmented ADF regression. In particular, ADF UR tests based on the QML estimates enjoy the same asymptotic power gains as the corresponding dummy-based tests. A further advantage of the QML approach is that asymptotic normality of the estimators of the 'short term' parameters γ allows one to use standard econometric techniques for lag order determination.

Remark 6.8. It is important to keep in mind that the asymptotic equivalence between QML UR test and the dummy-based UR is proved in Theorem 1 and Corollary 1 under the assumption of Gaussian innovations. This result may not hold in general: for instance, if the innovations are not normally distributed, the two approached may not deliver the same asymptotic power function. The Monte Carlo simulations reported in the next section provide some support to this statement.

Remark 6.9. Given the asymptotic equivalence of the QML-based and the unfeasible dummy-based UR statistics, under the null hypothesis the ADF^Q statistics do not have Dickey-Fuller asymptotic distributions. However, since the size distortions experienced by the dummy-based tests are in general negligible (see section 4), we advise – in line with what suggested by Lucas (1995b) – to use the QML approach in conjunction with standard Dickey-Fuller critical values.[5] This choice is supported by the finite sample results that will be presented in the next section. □

[5] An alternative approach for asymptotic critical value determination is to use Monte Carlo methods based on the QML residuals and on the estimated quasi expectations, $d_t(\check{\theta})$, of the outlier indicators δ_t. However, for a wide range of economically plausible models, we have found no significant size improvement over standard asymptotic critical values when Monte Carlo methods are implemented.

7 Finite sample properties of QML

In this section we analyze the finite sample properties of the robust QML UR tests of the previous section. In addition, the QML tests are compared with the robust 'M' t test proposed by Lucas (1995b), ADF_t^L hereafter.[6] Although Lucas (1995b) does not discuss a coefficient version of this test, we introduce it for comparison with the ADF_α^Q test, and denote it by ADF_α^L. The same Monte Carlo design as in section 5 is used. QML estimates are computed by iterating the map Φ, see eq. (13), until convergence, starting from OLS initial values. For the ADF^Q tests, we use the standard Dickey-Fuller critical values as reported in Fuller (1976, Tables 10.A.1 and 10.A.2), for the ADF^L tests the asymptotic critical values are simulated by the authors along the lines suggested in Lucas (1995b). The nominal level is 5%. Results are reported in Table 4 (Gaussian innovations) and in Table 5 (Student t innovations).

[Tables 4–5 about here]

The following points are worth noting; points (i)–(v) compare the size and power properties of the robust ADF^Q tests with those obtained for the dummy-based ADF^D tests (as well as for the standard ADF tests), while point (vi) discusses the differences between the ADF^Q and the ADF^L tests.

(i) Under the null hypothesis, for samples of $T = 100$ observations the QML-based tests are only marginally more liberal than the dummy ADF tests. In the case of the coefficient test ADF_α^Q, this partially offsets the size distortion of the dummy-based ADF_α^D test. For samples of $T = 200, 400$ observations, the size of the ADF^Q tests gets close to that of the corresponding ADF^D tests, and in particular, the ADF_t^Q test has very good size properties.

(ii) As noticed for the ADF^D tests in sections 4 and 5, in the presence of outliers the ADF^Q tests exhibit (size-adjusted) power gains over standard ADF tests. Under Gaussian errors, in terms of empirical rejection frequencies there is essentially no difference between the ADF^D tests and the ADF^Q tests.

(iii) There are no substantial differences in terms of (size-adjusted) power between the ADF_α^Q and the ADF_t^Q tests. However, since the former test tends to be undersized, the latter one is largely preferable, see the empirical rejection frequencies.

(iv) Some interesting properties can be noticed in the case of no outliers. Under Gaussian errors, the size and power of QML tests are roughly the same as those of standard ADF tests. That is, the use of robust QML tests instead of standard ADF tests does not imply deteriorated finite sample properties. Under t errors, the size of ADF^Q tests is quite close to the nominal level, with the ADF_α^Q test slightly undersized. However, under t errors the ADF^Q tests (in particular, the ADF_t^Q test) dominate the standard ADF tests in terms of power. This evidence suggests that the proposed QML approach can exhibit power gains when the innovations are not normally distributed, even if there are no outliers in the sense of Assumption $\mathcal{S}$.

(v) An important finding, related to what was noticed in point (iv) above, concerns the relation between the power of dummy-based and robust QML tests. In the Gaussian case,

[6] The choice of Lucas' test as a benchmark follows from the results in Thompson (2004).

it was proved in section 6 (and confirmed by the finite sample results in tables 2 and 4) that the dummy-based ADF^D tests attain the same asymptotic power as the ADF^Q tests. That is, the use of dummy variables, given that the econometrician is able to identify the outlier dates correctly, allows to obtain the same power as if the robust inference method was employed. This result – which obviously favors the 'common practice' of using dummy variables to account for outlying observations – seems not to hold when the errors are not Gaussian. Specifically, by comparing the results in tables 3 and 5, it can be seen that QML tests (in particular, the ADF_t^Q test) are *more* powerful than their ADF^D counterparts when the innovations are t distributed. This evidence holds for all the model considered in our Monte Carlo exercise.

(vi) In terms of (size-adjusted) power, the behavior of the robust M tests of Lucas (1995b) – ADF^L in Tables 4 and 5 – is quite close to that of the ADF^Q tests. However, for the models considered here both the coefficient version and the t version of the M tests tend to be undersized, in particular as the number of outliers grows. As a consequence, under local alternatives the empirical rejection frequencies of the ADF^L tests are much lower than those obtained using the ADF_t^Q test. Once again, a UR test based on the ADF_t^Q statistic seems to constitute the best compromise in terms of size and (size adjusted and raw) power.

To sum up, under Gaussian innovations the robust QML tests have size and power properties similar to those of the unfeasible dummy-based tests, in agreement with the theoretical discussion of the previous section. Under Student t innovations, however, the dummy-variable approach tends to be inferior to the robust QML tests in terms of power, although QML exploits no preliminary information on the outlier dates. For a variety of models, the 't' version of the robust test, ADF_t^Q, has very good size properties when used in conjunction with standard Dickey-Fuller critical values; hence, no new tables of critical values are needed in practice. The use of a robust inference method such as the ADF_t^Q test seems to constitute a better practice than the use of dummy variables, unless innovations, once having been cleaned from the outlying events, are approximately Gaussian.

8 Robust QML under deterministic time trends

Thus far we have assumed that the process of interest has no deterministic components. However, it is not difficult to generalize the robust QML approach to the case where the data are generated according to $y_t^* := d_t + y_t$, where y_t is as previously defined in (1), and $d_t := \psi' z_t$, z_t being a vector of deterministic components. As in Ng and Perron (2001), we now consider the pth order trend function, $z_t = (1, t, ..., t^p)'$, with special focus on the leading case of a linear trend ($p = 1$), although the analysis remains valid for more general cases, including, for example, the broken intercept and trend models discussed in Perron (1989, 1990) and Perron and Vogelsang (1992); cf. Phillips and Xiao (1998).

In order to improve power against local alternatives, instead of augmenting the ADF regression with the deterministic terms, we suggest a sequential procedure where initially, along the lines suggested, e.g., in Ng and Perron (2001), y_t^* is replaced by its detrended counterpart, say $\hat{y}_t$, and subsequently the robust QML approach is applied to the detrended series $\hat{y}_t$. This approach allows us to use detrending methods different from the OLS detrending method, which implicitly obtains when the ADF regression is augmented by the inclusion of z_t among the regressors.

In details, and restricting our attention to GLS detrending (Elliott et al., 1996), our suggested procedure is as follows:

1. a new series $\hat{y}_t$ is constructed by GLS-detrending y_t^* using standard methods[7] (i.e., ignoring the presence of the outliers);

2. robust QML estimation is carried out using $\hat{y}_t$ instead of y_t^*;

3. the robust ADF statistics, ADF_α^Q and ADF_t^Q are computed accordingly to the estimates of step 2.

We do not report a formal asymptotic analysis of the model. However, in finite samples results do not substantially differ from those reported in section 7 for the case of no deterministics, as it can be noticed from Table 6. In the table we evaluate the properties of the tests using pseudo-GLS detrending at $\bar{\alpha} := 1 - \bar{c}/T$, with $\bar{c} = 13.5$; size adjusted power and raw power are computed under $c = 13.5$. For samples of size $T = 100$ and $T = 200$ critical values are taken respectively from tables 3 ($T = 100$) and 7 ($T = 200$) in Xiao and Phillips (1998); for samples of size $T = 400$, asymptotic critical values as reported in Ng and Perron (2001), Table I, are used. For space constraints, results are reported for models S_0 (no outliers) and S_4 (four outliers, see section 5) only; the full set of results is available from the authors upon request.

For $T = 200$ and $T = 400$, the behavior of the ADF^Q tests is quite close to the behavior of its unfeasible dummy-based counterpart, ADF^D. The size of the test is largely acceptable and the tests allow to obtain sensible power gains with respect to standard ADF UR tests. Again, the ADF_t^Q test is preferable over the ADF_α^Q test in terms of size and empirical rejection frequencies. For $T = 100$ the ADF_t^Q test is slightly oversized, while under local alternatives its size-adjusted power is slightly inferior to the power of the unfeasible dummy tests, ADF^D. The coefficient test, ADF_α^Q, has good size for $T = 100$, but lower power with respect to its t-based counterpart, ADF_t^Q.

[Table 6 about here]

Finally, it is worth noting that the size distortions of the standard ADF tests decrease as the number of outliers increases, as predicted in the theoretical discussion based on the assumption of no deterministic components.

9 Concluding remarks

In this paper we have analyzed the effect of (random) outliers on (i) inference on the presence of a single UR, and (b) inference on the coefficients of the stable regressors in a finite-order autoregression, with or without a UR. With respect to the existing literature, our assumptions on the outlier process is rather general, allowing for multiple random outliers occurring at

[7]Given a time series x_t, $t = 0, 1, ..., T$, the pseudo-GLS detrended series at $\overline{\alpha} := 1 - \overline{c}/T$ ($\overline{c} \geq 0$) is defined as $\tilde{x}_t^{\overline{\alpha}} := x_t^{\overline{\alpha}} - \hat{\varphi}^{\overline{\alpha}\prime} z_t^{\overline{\alpha}}$, where $(x_0^{\overline{\alpha}}, x_t^{\overline{\alpha}}) := (x_0, (1 - \overline{\alpha}L)\, x_t)$, $(z_0^{\overline{\alpha}}, z_t^{\overline{\alpha}}) := (z_0, (1 - \overline{\alpha}L)\, z_t)$ and $\hat{\varphi}^{\overline{\alpha}}$ minimizes $S(\hat{\varphi}^{\overline{\alpha}}) := \sum_t (x_t^{\overline{\alpha}} - \hat{\varphi}^{\overline{\alpha}\prime} z_t^{\overline{\alpha}})^2$.

unknown dates and possibly clustering together. Despite the generality of our model, we have been able to show three general results. First, that in the presence of outliers the null and local-to-null asymptotic distributions (when they exist) of ADF-type statistics are expressed as functionals of a Wiener process and a jump process. Second, that clusters of outliers (e.g., outliers at consecutive dates) in general lead to inconsistent OLS estimation of the coefficients of stable regressors. Third, the addition of impulse dummies to the ADF regression allows one not only to estimate consistently the coefficients of the stable regressors, but also to obtain UR tests with high power. Notice that the dummy-based approach is unfeasible in practice, unless the outliers dates are known to the econometrician or, at least, detected correctly.

In the light of these results, we have proposed a feasible, robust QML approach to autoregression estimation and UR testing which permits to obtain (asymptotically) the same consistency and power gains as with the dummy approach but without requiring the knowledge of either the number of outliers or the outlier dates. Two further advantages of the QML approach is that it can be used in conjunction with standard Dickey-Fuller critical values, and it allows the practitioner to focus on the economic interpretation of the outlying events, since a by-product of QML is the consistent estimation of the outlier dates. The QML approach seems to work quite well in finite samples as well.

Throughout the paper, we have assumed that the lag order of the reference autoregressive process is known. This assumption should not be viewed as too restrictive. Specifically, since the autoregressive parameters are estimated consistently by QML when the employed lag order is not lower than the actual order, standard general-to-specific modeling strategies such as the sequential Wald test discussed in Ng and Perron (1995) may be used. Simulation results[8] (not reported) confirm this claim.

Finally, we believe that the interest of the results obtained for the robust QML approach goes beyond its ability of delivering UR tests with good size and improved power. Specifically, in the econometric and statistical literature on modeling outlying events there is often an opposition between dummy methods and robust methods. The results obtained here show that, in some circumstances, this opposition is actually inexistent, as the robust QML approach and the dummy-based approach are asymptotically equivalent. Similarly, while dummy methods are often considered handy and ad hoc methods without deep roots in statistical theory, here we show that a dummy-based approach has solid foundations as it arises naturally as the limit of a (Q)ML approach.

A Appendix

A.1 Preliminaries

First, we note the following direct consequence of the assumption that $N_T = O_P(1)$.

Lemma A.1 *For any sequence of random variables* $\{z_t\}$ *which is bounded in* P*-probability, also* $\max_{t\leq T}\{\delta_t|z_t|\} = O_P(1)$ *as* $T\to\infty$.

Next, we introduce the companion form version of representation (2). Denote by $\mathbf{Z}_{t-1}$ the stable regressors in (2), i.e., $\mathbf{Z}_{t-1} := (\Delta y_{t-1}, ..., \Delta y_{t-k})'$ under H_c $(c \geq 0)$, and $\mathbf{Z}_{t-1} := \mathbf{Y}_{t-1}$

[8]These are available from the authors upon request.

under H_s. Then

$$\mathbf{Z}_t = \Pi \mathbf{Z}_{t-1} + \mathbf{i} e_t, \qquad t = 1, ..., T, \tag{A.1}$$

where, with $\mathbf{0} := 0_{(k-1)\times 1}$, we have defined $\Pi := (\gamma, (\mathbf{I}_{k-1}\colon \mathbf{0})')'$, $\mathbf{i} := (1\colon \mathbf{0})'$ and $e_t := \varepsilon_t + \delta_t \theta_t - (c/T)\bar{\Gamma}(L) y_{t-1}$ under H_c $(c \geq 0)$, and $\Pi := ((\alpha, \gamma')', \Gamma, (\mathbf{0} : \mathbf{I}_{k-1}\colon \mathbf{0})')'$, $\mathbf{i} := (1, 1, \mathbf{0}')'$ and $e_t := \varepsilon_t + \delta_t \theta_t$ under H_s. The different meaning of some symbols under H_c $(c \geq 0)$ and H_s should cause no confusion in what follows.

A.2 Standard OLS approach

Lemma A.2 *Let Assumptions $\mathcal{M}$ and $\mathcal{S}$ be satisfied. Then, as $T \to \infty$, the following representations hold under H_c ($c \geq 0$) and H_s, unconditionally and conditionally on the occurrence of at least one outlier:*

a. $S_{zz} := T^{-1} \sum_{t=0}^{T-1} \mathbf{Z}_t \mathbf{Z}_t' = F_T + o_P(1)$, where $\lambda_{\min}(F_T)$ is bounded away from 0 in probability and $F_T := \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \Pi^i \mathbf{i} (\Pi^i \mathbf{i})' + \sum_{t=1}^{T-1} (\sum_{i=0}^{t-1} \Pi^i \mathbf{i} \delta_{t-i} \eta_{t-i})(\sum_{i=0}^{t-1} \Pi^i \mathbf{i} \delta_{t-i} \eta_{t-i})'$.

b. $S_{ze} := T^{-1} \sum_{t=1}^{T} \mathbf{Z}_{t-1} e_t = G_T + o_P(1)$, where $G_T := \sum_{t=1}^{T} (\sum_{i=1}^{t-1} \Pi^{i-1} \mathbf{i} \delta_{t-i} \eta_{t-i})(\delta_t \eta_t)$.

Further, the following representations hold under H_c ($c \geq 0$):

c. $S_{yy} := T^{-2} \sum_{t=1}^{T} y_{t-1}^2 = \sigma^2 \int H_{c,T}^2 + o_P(1)$.

d. $S_{zy} := T^{-1} \sum_{t=0}^{T-1} \mathbf{Z}_t y_t = \mathbf{1}_k \sigma^2 \int H_{c,T} dH_{c,T} - F_T (\mathbf{I} - \Pi')^{-1} \gamma + J_T + o_P(1)$, where $J_T = O_P(1)$ is defined before eq. (A.8).

e. $S_{ye} := T^{-1} \sum_{t=1}^{T} y_{t-1} e_t = \sigma^2 \bar{\Gamma}(1) \int H_{c,T} dH_{c,T} - \gamma' (\mathbf{I} - \Pi)^{-1} G_T + o_P(1)$.

f. $S_{ee} := T^{-1} \sum_{t=1}^{T} e_t^2 = \sigma_\varepsilon^2 + Q_T' + o_P(1)$, where $Q_T' := \sum_{t=1}^{T} \delta_t \eta_t^2$.

Proof. We present the derivations under H_c $(c \geq 0)$; those under H_s are analogous. For convenience initial values are set to zero in this proof.

Let $\mathbf{U}_t := (u_t, ..., u_{t-k+1})'$ and $\iota(L) := (L, ..., L^k)'$. With $g' := \bar{\gamma}' (\mathbf{I} - \Pi)^{-1}$ under H_c, the following representations are implied by the model equations (1)-(2): $\mathbf{Z}_t = \mathbf{U}_t - (c/T)\iota(L) y_t$,

$$\begin{aligned} \mathbf{U}_t &= \sum_{i=0}^{t-1} \Pi^i \mathbf{i} (\varepsilon_{t-i} + \delta_{t-i} \theta_{t-i}) = (\bar{\Gamma}(1))^{-1} \mathbf{1}_k (\varepsilon_t + \delta_t \theta_t) - \Pi (\mathbf{I} - \Pi)^{-1} \Delta \mathbf{U}_t, & \text{(A.2)} \\ y_t &= \sigma T^{1/2} H_{c,T}(t/T) - g' \mathbf{U}_t + (c/T) \upsilon_t, & \text{(A.3)} \end{aligned}$$

where $\upsilon_t := \sum_{i=0}^{t-1} (1 - c/T)^i g' \mathbf{U}_{t-1-i}$. Introduce also $\mathbf{U}_t^\varepsilon := \sum_{i=0}^{t-1} \Pi^i \mathbf{i} \varepsilon_{t-i}$ and $\mathbf{U}_t^\theta := \sum_{i=0}^{t-1} \Pi^i \mathbf{i} \delta_{t-i} \theta_{t-i}$, so that $\mathbf{U}_t = \mathbf{U}_t^\varepsilon + \mathbf{U}_t^\theta$, and observe that for a scalar sequence a_t,

$$\begin{aligned} \| T^{-1/2} \sum_{t=1}^{T} \mathbf{U}_t^\theta a_t \| &\leq (\max_{t \leq T} |a_t|) \| \sum_{t=0}^{T} \sum_{i=0}^{t-1} \| \Pi^i \| \delta_{t-i} |\eta_{t-i}| \leq (\max_{t \leq T} |a_t|)(\max_{t: \delta_t = 1} |\eta_t|) N_T (\sum_{i=0}^{\infty} \| \Pi^i \|) \\ &= O_P(\max_{t \leq T} |a_t|). & \text{(A.4)} \end{aligned}$$

The following magnitude orders hold too: $\max_{t \leq T} \| \mathbf{U}_t^\varepsilon \| = o_P(T^{1/2})$ by, e.g., (B.17) of Johansen (1996), $\max_{t \leq T} \| \mathbf{U}_t^\theta \| = O_P(T^{1/2})$ by (A.4), $\max_{t \leq T} |\upsilon_t| = O_P(T^{1/2})$ by the weak convergence of $\max_{t \leq T} \| T^{-1/2} \sum_{i=0}^{t-1} (1 - c/T)^i \mathbf{U}_{t-1-i}^\varepsilon \|$ and by (A.4) for $\max_{t \leq T} \| \sum_{i=0}^{t-1} (1 - c/T)^i \mathbf{U}_{t-1-i}^\theta \|$. Similarly, $\max_{s \in [0,1]} |H_{c,T}(s)| = O_P(1)$, and by combining the previous conclusions with (A.3), $\max_{t \leq T} |y_t| = O_P(T^{1/2})$.

Item (a) follows from the relations $T^{-1}\sum_{t=1}^{T}\mathbf{U}_t^{\varepsilon}(\mathbf{U}_t^{\varepsilon})' \xrightarrow{P} \sigma_\varepsilon^2\sum_{i=0}^{\infty}\Pi^i\mathbf{i}(\Pi^i\mathbf{i})' = Var(\mathbf{U}_t^{\varepsilon})$ (with the latter matrix strictly positive definite), $T^{-1}\sum_{t=1}^{T}\mathbf{U}_t^{\theta}(\mathbf{U}_t^{\theta})' = F_T - Var(\mathbf{U}_t^{\varepsilon})$, $T^{-1}\sum_{t=1}^{T}\mathbf{U}_t^{\theta}(\mathbf{U}_t^{\varepsilon})' = o_P(1)$ and $T^{-2}\sum\mathbf{U}_t^{\theta}[\iota(L)y_t]' = o_P(1)$ (both by (A.4), since $\max_{t\le T}|y_t| = O_P(T^{1/2})$ and $\max_{t\le T}\|\mathbf{U}_t^{\varepsilon}\| = o_P(T^{1/2})$), $T^{-3}\sum[\iota(L)y_t][\iota(L)y_t]' = o_P(1)$ and $T^{-2}\sum\mathbf{U}_t^{\varepsilon}[\iota(L)y_t]' = o_P(1)$ (by the same uniform evaluations of y_t and $\mathbf{U}_t^{\varepsilon}$).

We write S_{ze} as $G_T + \kappa_\varepsilon + \kappa_\theta - \kappa_y$, where (i) $G_T = T^{-1}\sum_{t=1}^{T}\mathbf{U}_{t-1}^{\theta}\delta_t\theta_t$; (ii) $\kappa_\varepsilon := T^{-1}(\sum_{t=1}^{T}\mathbf{U}_{t-1}^{\varepsilon}\varepsilon_t + \sum_{t=1}^{T}\mathbf{U}_{t-1}^{\varepsilon}\delta_t\theta_t - (c/T)\sum_{t=0}^{T-1}\mathbf{U}_t^{\varepsilon}\bar{\Gamma}(L)y_t) = o_P(1)$ respectively by an LLN, by Lemma A.1, and since $\max_{t\le T}\|\mathbf{U}_t^{\varepsilon}\| = o_P(T^{1/2})$ and $\max_{t\le T}|y_t| = O_P(T^{1/2})$; (iii) $\kappa_\theta := T^{-1}(\sum_{t=1}^{T}\mathbf{U}_{t-1}^{\theta}\varepsilon_t - (c/T)\sum_{t=0}^{T-1}\mathbf{U}_t^{\theta}\bar{\Gamma}(L)y_t) = o_P(1)$ by (A.4); (iv) $\kappa_y := \kappa_{y\varepsilon} + \kappa_{y\theta}$, $\kappa_{y\varepsilon} := (c/T^2)\sum_{t=1}^{T}\iota(L)y_t(\varepsilon_t - (c/T)\bar{\Gamma}(L)y_{t-1}) = o_P(1)$ since $\max_{t\le T}|\varepsilon_t| = o_P(T^{1/2})$ and $\max_{t\le T}|y_t| = O_P(T^{1/2})$, whereas $\kappa_{y\theta} := (c/T^2)\sum_{t=1}^{T}\iota(L)y_t\delta_t\theta_t = o_P(1)$ by Lemma A.1. Thus, $S_{ze} = G_T + o_P(1)$ as asserted in (b).

Further, from (A.3) it follows that

$$\begin{aligned} S_{yy} - \sigma^2\textstyle\int H_{c,T}^2 &= T^{-2}|\sum_{t=0}^{T-1} g'\mathbf{U}_t(g'\mathbf{U}_t - 2\sigma T^{1/2}H_{c,T}(t/T))| + o_P(1) \\ &= T^{-3/2}(2\sigma)|g'\sum_{t=0}^{T-1}\mathbf{U}_t^{\theta}H_{c,T}(t/T))| + o_P(1) = o_P(1), \end{aligned}$$

the first equality since $\max_{t\le T}\|\mathbf{U}_t\|$, $\max_{t\le T}|\upsilon_t|$ and $\max_{s\in[0,1]}|T^{1/2}H_{c,T}(s)|$ are $O_P(T^{1/2})$, the second one since $\max_{t\le T}\|\mathbf{U}_t^{\varepsilon}\| = o_P(T^{1/2})$, and the last one from (A.4). This proves (c).

Next, as $\sum_{t=1}^{T}\upsilon_{t-1}\varepsilon_t = o_P(T^2)$, $\sum_{t=0}^{T-1}\upsilon_t\bar{\Gamma}(L)y_t = O_P(T^2)$ and $\sum_{t=1}^{T}\upsilon_{t-1}\delta_t\theta_t = O_P(T)$, the former two since $\max_{t\le T}|\upsilon_t| = O_P(T^{1/2})$, $\max_{t\le T}|y_t| = O_P(T^{1/2})$ and $\max_{t\le T}|\varepsilon_t| = o_P(T^{1/2})$, and the latter one by Lemma A.1, it holds that, up to an $o_P(1)$ term,

$$\begin{aligned} S_{ye} &= T^{-1}\sum_{t=1}^{T}\big(T^{1/2}\sigma H_{c,T}((t-1)/T) - g'\mathbf{U}_{t-1}\big)(\varepsilon_t + \delta_t\theta_t) - (c/T^2)\sum_{t=0}^{T-1}y_t\bar{\Gamma}(L)y_t \\ &= \bar{\Gamma}(1)[\sigma^2\textstyle\int H_{c,T}dH_T - cS_{yy}] - g'G_T + O_P(T^{-1}S_{zy}) + o_P(1) \end{aligned} \tag{A.5}$$

by an LLN for $T^{-1}\sum_{t=1}^{T}\mathbf{U}_{t-1}^{\varepsilon}\varepsilon_t$, by Lemma A.1 for $T^{-1}\sum_{t=1}^{T}\mathbf{U}_{t-1}^{\varepsilon}\delta_t\theta_t$, by evaluation (A.4) for $T^{-1}\sum_{t=1}^{T}\mathbf{U}_{t-1}^{\theta}\varepsilon_t$, and since $T^{-1}\sum_{t=0}^{T-1}y_t(\bar{\Gamma}(L)-\bar{\Gamma}(1))y_t$ is a linear transformation of S_{zy}.

Still further, we find using (A.3) that

$$\begin{aligned} S_{zy} &= T^{-1}\sum_{t=0}^{T-1}\mathbf{Z}_t(T^{1/2}\sigma H_{c,T}(t/T) - g'\mathbf{Z}_t - (c/T)g'\iota(L)y_t + (c/T)\upsilon_t) \\ &= T^{-1/2}\sigma\sum_{t=0}^{T-1}\mathbf{U}_tH_{c,T}(t/T) - T^{-3/2}\sigma c\sum_{t=0}^{T-1}\iota(L)y_tH_{c,T}(t/T) - S_{zz}g + o_P(1) \end{aligned} \tag{A.6}$$

since $T^{-1}\sum_{t=0}^{T-1}\mathbf{Z}_tg'\iota(L)y_t$ is a linear transformation of S_{zz} and S_{zy}. Here

$$T^{-3/2}\sum_{t=0}^{T-1}\iota(L)y_tH_{c,T}(t/T) = \mathbf{1}_k\sigma\textstyle\int H_{c,T}^2 + o_P(1) \tag{A.7}$$

similarly to S_{yy}, and

$$T^{-1/2}\sum_{t=0}^{T-1}\mathbf{U}_tH_{c,T}(t/T)=T^{-1/2}\big(\sum_{t=0}^{T-1}\mathbf{U}_{t+1}H_{c,T}(t/T)-\sum_{t=0}^{T-1}\Delta\mathbf{U}_{t+1}H_{c,T}(t/T)\big)$$
$$=T^{-1/2}\big(\sum_{t=0}^{T-1}((\bar\Gamma(1))^{-1}\mathbf{1}_k(\varepsilon_{t+1}+\delta_{t+1}\theta_{t+1})-(\Pi(\mathbf{I}-\Pi)^{-1}+\mathbf{I})\Delta\mathbf{U}_{t+1})H_{c,T}(t/T)\big)$$
$$=\mathbf{1}_k\sigma\int H_{c,T}dH_T-T^{-1/2}(\mathbf{I}-\Pi)^{-1}\sum_{t=0}^{T-1}\Delta\mathbf{U}_{t+1}H_{c,T}(t/T),$$

the first equality by (A.2). The term $T^{-1/2}\sum_{t=0}^{T-1}\Delta\mathbf{U}_{t+1}H_{c,T}(t/T)$ equals

$$T^{-1/2}\mathbf{U}_TH_{c,T}(1)-T^{-1}\sigma_\varepsilon^{-1}\sum_{t=1}^{T}\mathbf{U}_t(\varepsilon_t+\delta_t\theta_t)+cT^{-3/2}\sum_{t=1}^{T}\mathbf{U}_tH_{c,T}((t-1)/T)$$
$$=\quad T^{-1/2}\mathbf{U}_T^\theta H_{c,T}(1)-\sigma_\varepsilon\mathbf{i}-T^{-1/2}\sigma_\varepsilon^{-1}\sum_{t=1}^{T}\mathbf{U}_t^\theta\delta_t\eta_t+o_P(1)$$

since $\mathbf{U}_T^\varepsilon=o_P(T^{1/2})$, $T^{-1}\sum_{t=1}^T\mathbf{U}_t^\varepsilon\varepsilon_t\xrightarrow{P}\sigma_\varepsilon^2\mathbf{i}$ by an LLN, $T^{-1}\sum_{t=1}^T\mathbf{U}_t^\varepsilon\delta_t\theta_t=o_P(1)$ by Lemma A.1, $T^{-1}\sum_{t=1}^T\mathbf{U}_t^\theta\varepsilon_t=o_P(1)$ by (A.4), $T^{-3/2}\sum_{t=1}^T\mathbf{U}_t^\varepsilon H_{c,T}((t-1)/T)=o_P(1)$ by evaluating the summands uniformly, $T^{-3/2}\sum_{t=1}^T\mathbf{U}_t^\theta H_{c,T}((t-1)/T)=o_P(1)$ by (A.4). Introducing $J_T:=\sigma^2\mathbf{1}_k-(\mathbf{I}-\Pi)^{-1}\sigma T^{-1/2}[\mathbf{U}_T^\theta H_{c,T}(1)-\sigma_\varepsilon^{-1}\sum_{t=1}^T\mathbf{U}_t^\theta\delta_t\eta_t]$, we find that

$$T^{-1/2}\sum_{t=0}^{T-1}\mathbf{U}_tH_{c,T}(t/T)=\mathbf{1}_k\sigma\int H_{c,T}dH_T+\sigma^{-1}J_T+o_P(1),\tag{A.8}$$

which in conjunction with (A.6) and (A.7) gives representation (d). In particular, $S_{zy}=O_P(1)$, and returning to (A.5) we obtain also (e).

Finally, $S_{ee}-T^{-1}\sum_{t=1}^T\varepsilon_t^2-Q_T'$ equals

$$T^{-1/2}\sum_{t=1}^{T}(\varepsilon_t-(c/T)\bar\Gamma(L)y_{t-1})\delta_t\theta_t-(c/T^2)\sum_{t=1}^{T}(e_t-\delta_t\theta_t)\bar\Gamma(L)y_{t-1}=o_P(1),$$

as can be seen from Lemma A.1 and the relations $\max_{t\le T}|\varepsilon_t|=o_P(T^{1/2})$ and $\max_{t\le T}|y_t|=O_P(T^{1/2})$. ■

Proof of Proposition 1. We start by deriving a large sample representation of $T\hat\pi$ under the hypothesis H_c ($c\ge 0$); it will be useful also in the proof of Proposition 2. Then we discuss, simultaneously under H_c and H_s, how the coefficients to the stable regressors are estimated.

Under H_c ($c\ge 0$), we defined $\mathbf{Z}_t=(\Delta y_t,...,\Delta y_{t-k+1})'$, so that $\hat\pi=M_{1,T}/M_{2,T}$, with

$$T^{-1}M_{1,T}:=S_{ye}-S_{zy}'S_{zz}^{-1}S_{ze}\quad\text{and}\quad T^{-2}M_{2,T}:=S_{yy}-T^{-1}S_{zy}'S_{zz}^{-1}S_{zy}=S_{yy}+o_P(1),$$

the magnitude order by Lemma A.2(a,d). Introduce

$$\varkappa_{0,T}:=-\bar\Gamma(1)^{-1}[\mathbf{1}_k\int H_{c,T}dH_{c,T}+\sigma^{-2}J_T]'F_T^{-1}G_T.\tag{A.9}$$

Inserting the expressions for $M_{1,T}$ and $M_{2,T}$ into $\hat\pi=M_{1,T}/M_{2,T}$, and applying Lemma A.2 to the terms of these expressions, we get

$$T\hat\pi=\bar\Gamma(1)[(\int H_{c,T}dH_T+\varkappa_{0,T})(\int H_{c,T}^2)^{-1}-c]+o_P(1),\tag{A.10}$$

since S_{yy} is bounded away from zero in probability. The last expression is $O_P(1)$, and hence, $\hat{\pi} = O_P(T^{-1})$.

Let Ξ collect the coefficients to the stable regressor $\mathbf{Z}_{t-1}$ in (3) under both H_c ($c \geq 0$) and H_s. We have under these hypotheses that

$$(\hat{\Xi} - \Xi)' = S_{zz}^{-1}(S_{ze} - T^{-1} \textstyle\sum_{t=1}^{T} \mathbf{Z}_{t-1} r_t),$$

where $\hat{\Xi}$ is the OLS estimator of Ξ from the regression of Δy_t on $\mathbf{Y}_{t-1}$ ($\mathbf{Y}_t = \mathbf{Z}_t$ under H_s and $\mathbf{Y}_t = (y_t, \mathbf{Z}_t)'$ under H_c, $c \geq 0$); $r_t = 0$ under H_s, and $r_t = \hat{\pi} y_{t-1}$ under H_c, $c \geq 0$. From $T^{-1} \sum_{t=1}^{T} \mathbf{Z}_{t-1} r_t = o_P(1)$ (Lemma A.2(d) and $\hat{\pi} = o_P(1)$ under H_c), and from Lemma A.2(a,b), it follows that $(\hat{\Xi} - \Xi)' = F_T^{-1} G_T + o_P(1)$. Thus, $\hat{\Xi} - \Xi = o_P(1)$ if and only if

$$G_T = \sum_{t=1}^{T} (\sum_{i=\tau}^{t-1} \Pi^{i-1} \mathbf{i} \delta_{t-i} \eta_{t-i})(\delta_t \eta_t) = o_P(1), \tag{A.11}$$

where the subscript T of τ is subsumed. If $\Xi = 0$, then $\Pi \mathbf{i} = 0$, $G_T = 0$, and consistency of $\hat{\Xi}$ for Ξ is trivial. On the other hand, if $\tau \xrightarrow{P} \infty$, then

$$\|G_T\| \leq (\max_{t:\delta_t = 1} \eta_t)^2 \sum_{t=1}^{T} \sum_{i=\tau}^{t-1} \|\Pi^{i-1}\| \delta_{t-i} \delta_t \leq (\max_{t:\delta_t=1} \eta_t)^2 N_T \sum_{i=\tau}^{\infty} \|\Pi^{i-1}\| \xrightarrow{P} 0$$

since $\max_{t:\delta_t=1} \eta_t = O_P(1)$, $N_T = O_P(1)$ and $\sum_{i=0}^{\infty} \|\Pi^i\| < \infty$. This proves the sufficiency part of the proposition.

We argue next that if $\Xi \neq 0$ and if the probability for exactly two outliers to occur (event E_2, say) is bounded away from zero, then the divergence $\tau \xrightarrow{P} \infty$ conditional on E_2 is necessary for $G_T = o_P(1)$, and hence, for consistency of $\hat{\Xi}$. Indeed, conditionally on E_2,

$$\|G_T\| = \|\Pi^{\tau-1} \mathbf{i} \sum_{t=1}^{T} \delta_{t-\tau} \delta_t \eta_t \eta_{t-\tau}\| \geq \|\Pi^{\tau-1} \mathbf{i}\| (\min_{t:\delta_t=1} \eta_t)^2,$$

and since $(\min_{t:\delta_t=1} \eta_t)^2$ is bounded away from zero in probability (also conditional on E_2, since E_2 has non-vanishing probability), if $G_T = o_P(1)$ (again also conditionally on E_2), it follows that $\|\Pi^{\tau-1} \mathbf{i}\| \xrightarrow{P} 0$ conditionally on E_2, and further, that $\tau \xrightarrow{P} \infty$ conditionally on E_2. The latter because (possibly upon substitution of Π by one of its leading submatrices, and of $\mathbf{i}$ by a matching subvector) we can write $\|\Pi^{\tau-1} \mathbf{i}\| \xrightarrow{P} 0$ together with $\lambda_{\min}(\Pi) > 0$ (because $\Xi \neq 0$), and then, if $\Pi = V^{-1} J V$ is the Jordan decomposition of Π,

$$\mathbf{i}'(\Pi^{\tau-1})' \Pi^{\tau-1} \mathbf{i} \geq (\mathbf{i}'\mathbf{i}) \lambda_{\min}((\Pi^{\tau-1})' \Pi^{\tau-1}) \geq c[\lambda_{\min}(J'J)]^{\tau-1},$$

where $c := \lambda_{\min}(V'V) \lambda_{\min}\left((V^{-1})' V^{-1}\right) > 0$, and $\lambda_{\min}(J'J) > 0$ since $\lambda_{\min}(\Pi) > 0$.

Alternatively, let us condition on the occurrence of at least two outliers (event E_+). Let $\bar{t} := \min\{t \in \{2, ..., T\} : \delta_t \delta_{t-\tau} = 1 \text{ and } \delta_t \delta_{t-i} = 0,\ i < \tau\}$. Then $G_T = G_{T,1} \eta_{\bar{t}} + G_{T,2}$, where $G_{T,1}$ and $G_{T,2}$ depend on $\{(\delta_t, \eta_t) : t \neq \bar{t}\}$. We argue first that if $G_T = o_P(1)$ conditionally on E_+, then also $G_{T,1} = o_P(1)$ conditionally on E_+. Indeed, if $G_{T,1}$ would be bounded away from zero along a subsequence of sample sizes, we would have that $\eta_{\bar{t}} + G_{T,2}/G_{T,1} = o_P(1)$, conditionally on E_+, along that subsequence (we write as if it is the entire sequence). Since the distribution of η_1 is non-degenerate by hypothesis, there exist $a > 0$ and disjoint closed sets of real numbers $\mathcal{F}_{T,1}$ and $\mathcal{F}_{T,2}$ such that $P\left(\eta_{\bar{t}} \in \mathcal{F}_{T,i}|E_+\right) = P\left(\eta_1 \in \mathcal{F}_{T,i}\right) > a$, $i = 1, 2$. Let $\mathcal{U}_{T,1}$ and $\mathcal{U}_{T,1}$ be disjoint open sets such that $\mathcal{F}_{T,i} \subset \mathcal{U}_{T,i}$, $i = 1, 2$. Then, since $\eta_{\bar{t}} + G_{T,2}/G_{T,1} = o_P(1)$ conditionally on E_+, it should hold that $P(G_{T,2}/G_{T,1} \in \mathcal{U}_{T,1}|E_+, \eta_{\bar{t}} \in \mathcal{F}_{T,1}) \to 1$ and $P(G_{T,2}/G_{T,1} \in \mathcal{U}_{T,1}|E_+, \eta_{\bar{t}} \in \mathcal{F}_{T,2}) \to 0$, contradicting the joint independence of $\{\eta_t\}$ (recall also Assumption $\mathcal{S}$(c)). Therefore, $G_{T,1} = o_P\left(1\right)$ conditionally on E_+. But $G_{T,1} = \Pi^{\tau-1}\mathbf{i}\eta_{\bar{t}-\tau} + \sum_{i=\tau+1}^{\bar{t}-1} \Pi^{i-1}\mathbf{i}\delta_{\bar{t}-i}\eta_{\bar{t}-i}$, and by a similar independence argument, $\Pi^{\tau-1}\mathbf{i} = o_P\left(1\right)$ conditionally on E_+, and $\tau_T \to \infty$ conditionally on E_+, as argued earlier for E_2. ■

Proof of Proposition 2. The expression for ADF_α in (a) follows from (A.10). Note that $\varkappa_{0,T} = o_P(1)$ if and only if $G_T = o_P(1)$, which in the proof of Proposition 1 was shown to be necessary and sufficient for the consistent OLS estimation of γ ($= \Xi$ under H_c, $c \geq 0$).

Besides $M_{1,T}$ and $M_{2,T}$ introduced earlier, let $T^{-1}M_{3,T} := S_{ee} - S_{ze}'S_{zz}^{-1}S_{ze} = S_{ee} - G_T'F_T^{-1}G_T + o_P(1)$, the last equality by Lemma A.2(a,b). As $M_{1,T}/T = O_P\left(1\right)$ was shown to hold, and $M_{3,T}/T$ is bounded away from 0 in probability (by A.2(f) and the inequality $Q_T - G_T'F_T^{-1}G_T \geq 0$), we find that

$$\begin{aligned} ADF_t &= M_{1,T}/T(M_{2,T}M_{3,T}/T^3 - M_{1,T}^2/T^3)^{-1/2} = T\hat{\pi}(M_{2,T}/T^2)^{1/2}(M_{3,T}/T)^{-1/2} + o_P\left(1\right) \\ &= (\textstyle\int H_{c,T}dH_{c,T} + \varkappa_{0,T})(\varkappa_{1,T}\int H_{c,T}^2)^{-1/2} + o_P(1) \end{aligned}$$

as asserted in (a), with

$$\varkappa_{1,T} := 1 + \sigma_\varepsilon^{-2}(Q_T - G_T'F_T^{-1}G_T). \tag{A.12}$$

The expressions in (b) obtain by inserting $G_T = o_P(1)$ and $\hat{\gamma} = \gamma + o_P(1)$ into those of (a). ■

A.3 Dummy-based approach

We start from the counterpart of Lemma A.2. A key difference is item (b), where convergence to zero ensures consistent estimation of the coefficients to the stable regressors.

Lemma A.3 *Let Assumptions $\mathcal{M}$ and $\mathcal{S}$ be satisfied. Then, as $T \to \infty$, the following representations hold under H_c ($c \geq 0$) and H_s, unconditionally and conditionally on the occurrence of at least one outlier:*

a. $S_{zz}^{1-\delta} := T^{-1}\sum_{t=1}^T(1-\delta_t)\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}' = F_T^{1-\delta} + o_P(1)$, *where* $\lambda_{\min}(F_T^{1-\delta})$ *is bounded away from 0 in probability and* $F_T^{1-\delta} := F_T - \sum_{t=1}^{T-1}\delta_t(\sum_{i=0}^{t-1}\Pi^i\mathbf{i}\delta_{t-i}\eta_{t-i})(\sum_{i=0}^{t-1}\Pi^i\mathbf{i}\delta_{t-i}\eta_{t-i})'$.

b. $S_{ze}^{1-\delta} := T^{-1}\sum_{t=1}^T(1-\delta_t)\mathbf{Z}_{t-1}e_t = o_P(1)$.

Further, the following representations hold under H_c ($c \geq 0$):

c. $S_{yy}^{1-\delta} := T^{-2}\sum_{t=1}^T(1-\delta_t)y_{t-1}^2 = \sigma^2\int H_{c,T}^2 + o_P(1)$.

d. $S_{zy}^{1-\delta} := T^{-1}\sum_{t=1}^T(1-\delta_t)\mathbf{Z}_{t-1}y_{t-1} = O_P(1)$.

e. $S_{ye}^{1-\delta} := T^{-1}\sum_{t=1}^T(1-\delta_t)y_{t-1}e_t = \sigma^2\bar{\Gamma}(1)[\int H_{T,c}dB_T - c\int H_{T,c}^2] + o_P(1)$.

f. $S_{ee}^{1-\delta} := T^{-1}\sum_{t=1}^T(1-\delta_t)e_t^2 = \sigma_\varepsilon^2 + o_P\left(1\right)$.

Proof. The proof is similar to that of Lemma A.2 and we omit the details. We only note that $T^{-1}\sum_{t=1}^T\delta_t\mathbf{Z}_{t-1}e_t = G_T + o_P\left(1\right)$, $T^{-1}\sum_{t=1}^T\delta_te_t^2 = Q_T' + o_P(1)$ and $T^{-1}\sum_{t=1}^T\delta_ty_{t-1}(\varepsilon_t + \theta_t) = \sigma\int H_{c,T}dC_T - \gamma'(\mathbf{I} - \Pi)^{-1}G_T + o_P(1)$, which together with Lemma A.2(b,f) and (A.5) gives items (b), (f) above, and the relation $T^{-1}\sum_{t=1}^{T}(1-\delta_t)y_{t-1}\varepsilon_t = \sigma^2\bar{\Gamma}(1)\int H_{c,T}dB_T + o_P(1)$. Thus,

$$\begin{aligned}\sum_{t=1}^{T}(1-\delta_t)y_{t-1}e_t &= \sum_{t=1}^{T}(1-\delta_t)y_{t-1}\varepsilon_t - (c/T)\bar{\Gamma}(1)\sum_{t=1}^{T}(1-\delta_t)y_{t-1}^2 + o_P(T) \\ &= T\sigma^2\bar{\Gamma}(1)[\textstyle\int H_{T,c}dB_T - c\int H_{T,c}^2] + o_P(T),\end{aligned}$$

as asserted in (e). ■

Proof of Proposition 3. We follow the steps from the proofs of Propositions 1 and 2. Under H_c ($c \geq 0$), we have $\tilde{\pi} = \tilde{M}_{1,T}/\tilde{M}_{2,T}$, with

$$T^{-1}\tilde{M}_{1,T} := S_{ye}^{1-\delta} - (S_{zy}^{1-\delta})'(S_{zz}^{1-\delta})^{-1}S_{ze}^{1-\delta} \quad \text{and} \quad T^{-2}\tilde{M}_{2,T} := S_{yy}^{1-\delta} - (S_{zy}^{1-\delta})'(S_{zz}^{1-\delta})^{-1}S_{zy}^{1-\delta},$$

and by Lemma A.3,

$$T^{-1}\tilde{M}_{1,T} = \sigma^2\bar{\Gamma}(1)[\textstyle\int H_{T,c}dB_T - c\int H_{T,c}^2] + o_P(1) \quad \text{and} \quad T^{-2}\tilde{M}_{2,T} = \sigma^2\int H_{c,T}^2 + o_P(1).$$

Inserting the above expressions for $\tilde{M}_{1,T}$ and $\tilde{M}_{2,T}$ into that for $\tilde{\pi}$, we conclude that

$$T\tilde{\pi} = \bar{\Gamma}(1)[(\textstyle\int H_{c,T}dB_T)(\int H_{c,T}^2)^{-1} - c] + o_P(1), \tag{A.13}$$

since $T^{-2}\sum_{t=1}^{T}y_{t-1}^2$ is bounded away from zero in probability. Hence, $\tilde{\pi} = O_P(T^{-1})$.

Let $\tilde{\Xi}$ denote the dummy-based estimator of Ξ (the coefficient vector associated to $\mathbf{Z}_{t-1}$) from the regression of Δy_t on $\mathbf{Y}_{t-1}$ ($\mathbf{Y}_t = \mathbf{Z}_t$ under H_s and $\mathbf{Y}_t = (y_t, \mathbf{Z}_t)'$ under H_c). With $\tilde{r}_t = 0$ under H_s and $r_t = \tilde{\pi}y_{t-1}$ under H_c, we have that

$$(\tilde{\Xi} - \Xi)' = (S_{zz}^{1-\delta})^{-1}(S_{ze}^{1-\delta} - T^{-1}\textstyle\sum_{t=1}^{T}(1-\delta_t)\mathbf{Z}_{t-1}\tilde{r}_t).$$

As $T^{-1}\sum_{t=1}^{T}(1-\delta_t)\mathbf{Z}_{t-1}\tilde{r}_t = o_P(T^{-1/2})$ (Lemma A.3(d) and $\tilde{\pi} = O_P\left(T^{-1}\right)$ under H_c), from Lemma A.3(a,b) we obtain that $\tilde{\Xi} - \Xi = o_P(1)$. Furthermore, notice for reference later that

$$T^{1/2}(\tilde{\Xi} - \Xi)' = (S_{zz}^{1-\delta})^{-1}T^{-1/2}\sum_{t=1}^{T}(1-\delta_t)\mathbf{Z}_{t-1}\varepsilon_t + o_P(1). \tag{A.14}$$

The expression for ADF_α^D in (b) follows from (A.13) and the fact that $\bar{\Gamma}(1)$ is estimated consistently. Further, let $T^{-1}\tilde{M}_{3,T} := S_{ee} - S_{ze}'(S_{zz})^{-1}S_{ze}$. As $\tilde{M}_{1,T}/T = O_P(1)$ and $\tilde{M}_{3,T}/T \xrightarrow{P} \sigma_\varepsilon^2$ by Lemma A.3(a,b,f), as for the ADF_t statistic, we find that

$$\begin{aligned}ADF_t^D &= T\tilde{\pi}(\tilde{M}_{2,T}/T^2)^{1/2}(\tilde{M}_{3,T}/T)^{-1/2} + o_P(1) \\ &= (\textstyle\int H_{c,T}dB_T)(\int H_{c,T}^2)^{-1/2} - c(\int H_{c,T}^2)^{1/2} + o_P(1)\end{aligned}$$

as asserted in (b).

From the conclusion that $\tilde{\pi}$ is consistent under H_s, it follows that $T(\tilde{\pi} - 1) \xrightarrow{P} -\infty$. Further, $|\tilde{\Gamma}(1)| = O_P(1)$ since $\tilde{\gamma}$ has a finite probability limit, while $s(\tilde{\pi}) = O_P(1)$ since (i) the $(1,1)$ element of $(S_{zz}^{1-\delta})^{-1}$ is $O_P(1)$ by Lemma A.3(a), and (ii) $\tilde{\sigma}_\varepsilon^2 = O_P(1)$ by its consistency for σ_ε^2, implied by the discussion of the coefficient estimators. ■

Next, we present the derivations underlying the third column of Table 1. Upon substitution of C by hC, we find

$$\begin{aligned}
\textstyle\int H_c^2 &= h^2 \textstyle\int C_c^2 + 2h \int C_c B_c + \int B_c^2 = h^2 \int C_c^2 + O_P(h), \\
\textstyle\int H_c dH &= h^2 \textstyle\int C_c dC + h(\int C_c dB + \int B_c dC) + \int B_c dB = h^2 \int C_c dC + O_P(h), \\
\textstyle\int H_c dB &= h \textstyle\int C_c dB + \int B_c dB = h \int C_c dB + O_P(1).
\end{aligned}$$

Substituting also $[C]$ by $[hC] = h^2[C]$, accounting for the fact that $[C] > 0$ a.s. conditionally on the occurrence of at least one jump, and letting $h \to \infty$ gives directly the limit in the OLS case. In the dummy-variable case, for $c = 0$ the limit of ADF_t^D is $(\int C^2)^{-1/2} \int CdB$, which by the independence of C and B is standard Gaussian. For $c > 0$, its limit is formally

$$(-c)\infty + (\textstyle\int C_c^2)^{-1/2} \int C_c dB = -\infty + O_P(1) = -\infty.$$

The limits of the coefficient statistic follow similarly.

A.4 QML approach

Let $\rho \in (0, 1/4)$ be arbitrary, but fixed in the sequel. Let $\mathbb{A}_T := \mathbb{A}_T^\Gamma \times \mathbb{A}_T^\varepsilon \times \mathbb{A}_T^\eta \times \mathbb{A}_T^\lambda$, with $\mathbb{A}_T^\Gamma := \{\Gamma \in \mathbb{R}^{k+1} : \|T^{1/2} D_T^{-1}(\Gamma - \Gamma_0)\| \leq (\ln T)^{1/4}\}$, $\mathbb{A}_T^\varepsilon := [\sigma_{\varepsilon 0}^2/(1 + \frac{\rho}{2}), 2\sigma_{\varepsilon 0}^2]$, $\mathbb{A}_T^\eta := [1/2, 2]$ and $\mathbb{A}_T^\lambda := [-1/2, 2]$. Define on $\mathbb{A}_T$ the random function ω by

$$\omega(\Gamma', \sigma_\varepsilon^2, x^\eta, x^\lambda) := (\Gamma', \sigma_\varepsilon^2, x^\eta Q_T, x^\lambda + N_T)'.$$

Note that ω is a.s. invertible conditionally on the occurrence of at least one outlier.

To streamline the exposition, the proofs in this section are presented under the hypotheses H_0 and H_s. The extension to H_c ($c > 0$) requires to incorporate the term $-(c/T)\bar{\Gamma}(L)y_{t-1}$ into the error e_t, see (A.1), which poses no conceptual difficulties.

We start from the following crucial Lemma, where $\sup_{\mathbb{A}_T} f(\omega) := \sup_{x \in \mathbb{A}_T} f(\omega(x))$ for any matching f.

Lemma A.4 *Let Assumptions $\mathcal{M}$ and $\mathcal{S}$ hold. If P denotes probability conditional on the occurrence of at least one outlier, the following relations hold as $T \to \infty$.*

a. $\sup_{\mathbb{A}_T} \sum_{t=1}^T |d_t(\omega) - \delta_t| = O_P(T^{\rho - 1/2})$ *and* $\sup_{\mathbb{A}_T} \sum_{t=1}^T \delta_t |d_t(\omega) - \delta_t| \xrightarrow{P} 0$ *faster-than-algebraically.*

b. $\sup_{\mathbb{A}_T} \|(\Phi^\Gamma(\omega) - \tilde{\Gamma}') D_T^{-1} T^{1/2}\| = o_P(1)$.

c. $\sup_{\mathbb{A}_T} \|(\Phi^\varepsilon(\omega) - \sigma_{\varepsilon 0}^2, \Phi^\eta(\omega) - Q_T, \Phi^\lambda(\omega) - N_T)\| = o_P(1)$.

Proof. We write ω^η and ω^λ for $x^\eta Q_T$ and $x^\lambda + N_T$. It holds that $\sum_{t=1}^T |d_t(\omega) - \delta_t| = \sum_{t=1}^T (1 - \delta_t) d_t(\omega) + \sum_{t=1}^T \delta_t (1 - d_t(\omega))$, and we start from the first sum. It satisfies

$$\begin{aligned}
\sum_{t=1}^T (1 - \delta_t) d_t(\omega) &\leq \frac{\omega^\lambda / T}{1 - \omega^\lambda / T} \sum_{t=1}^T (1 - \delta_t) \frac{l_t(1, \omega)}{l_t(0, \omega)} \\
&= \frac{\omega^\lambda / T}{1 - \omega^\lambda / T} \frac{\sigma_\varepsilon}{(\sigma_\varepsilon^2 + T\omega^\eta)^{1/2}} \sum_{t=1}^T (1 - \delta_t) \exp\Big(\frac{(\Delta y_t - \Gamma' \mathbf{Y}_{t-1})^2}{2} \Big(\frac{1}{\sigma_\varepsilon^2} - \frac{1}{\sigma_\varepsilon^2 + T\omega^\eta}\Big)\Big) \\
&< \frac{N_T + 2}{T - N_T + 1/2} \frac{2\sigma_{\varepsilon 0}}{(\sigma_{\varepsilon 0}^2/(2 + \rho) + TQ_T)^{1/2}} \sum_{t=1}^T (1 - \delta_t) \exp\Big(\frac{(\Delta y_t - \Gamma' \mathbf{Y}_{t-1})^2}{2\sigma_\varepsilon^2}\Big)
\end{aligned}$$

at every point in $\mathbb{A}_T$. As $N_T = O_P(1)$ and Q_T is bounded away from 0 in P-probability, the term in front of the summation above is $O_P(T^{-3/2})$. Further, as $(1-\delta_t)(\Delta y_t - \Gamma'\mathbf{Y}_{t-1}) = (1-\delta_t)(\varepsilon_t + (\Gamma_0 - \Gamma)'\mathbf{Y}_{t-1})$, on $\mathbb{A}_T$ the summation itself does not exceed

$$\sum_{t=1}^T \exp\Big(\frac{(\varepsilon_t + (\Gamma_0-\Gamma)'\mathbf{Y}_{t-1})^2}{2\sigma_\varepsilon^2}\Big) \le \exp\Big(\frac{a_T}{2\sigma_\varepsilon^2}(1+\frac{\rho}{2})\Big) \sum_{t=1}^T \exp\Big(\frac{\varepsilon_t^2}{2\sigma_{\varepsilon 0}^2}(1+\frac{\rho}{2})\Big),$$

where (i) a_T is defined in the first line below:

$$\begin{aligned} &\|D_T^{-1}(\Gamma-\Gamma_0)\|^2 \max_{t\le T}\|D_T\mathbf{Y}_{t-1}\|^2 + 2\sup_{\mathbb{A}_T}\|D_T^{-1}(\Gamma-\Gamma_0)\| \max_{t\le T}\|D_T\mathbf{Y}_{t-1}\| \max_{t\le T}|\varepsilon_t| \\ &\le (T^{-1}\ln T)O_P(T) + (T^{-1}\ln T)^{1/2}O_P(T^{1/2})O_P((\ln T)^{1/2}) = O_P(T^{\rho/4}), \end{aligned}$$

and (ii) $\sum_{t=1}^T \exp(\varepsilon_t^2(1+\rho/2)/(2\sigma_{\varepsilon 0}^2)) = O_P(T^{1+3\rho/4})$, both using the Gaussianity of ε_t, and (ii) using also Lemma 7(a) in Georgiev (2005). Thus,

$$\sup_{\mathbb{A}_T}\sum_{t=1}^T (1-\delta_t)d_t(\omega) \le O_P(T^{-3/2})O_P(T^{\rho/4})O_P(T^{1+3\rho/4}) = O_P(T^{\rho-1/2}). \tag{A.15}$$

As $1-d_t(\omega) \le [(1-\omega^\lambda/T)/(\omega^\lambda/T)]l_t(0,\omega)/l_t(1,\omega)$, we find that

$$\begin{aligned} \sum_{t=1}^T \delta_t(1-d_t(\omega)) &= (1+T\frac{\omega^\eta}{\sigma_\varepsilon^2})^{1/2}\frac{1-\omega^\lambda/T}{\omega^\lambda/T}\sum_{t=1}^T \delta_t \exp\Big(\frac{(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{2}\Big(\frac{1}{\sigma_\varepsilon^2 + T\omega^\eta} - \frac{1}{\sigma_\varepsilon^2}\Big)\Big) \\ &\le O_P(T^{3/2})\exp\Big(\frac{\max_{t\le T}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{2(\sigma_\varepsilon^2 + T\omega^\eta)}\Big)\exp\Big(-\frac{\min_{t:\delta_t=1}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{2\sigma_\varepsilon^2}\Big) \end{aligned}$$

uniformly on $\mathbb{A}_T$, since $\sup_{\mathbb{A}_T}\omega^\eta = O_P(1)$, whereas $\inf_{\mathbb{A}_T}\omega^\lambda$ and $\inf_{\mathbb{A}_T}\sigma_\varepsilon^2$ are bounded away from 0 in P-probability. Further, as $\Delta y_t - \Gamma'\mathbf{Y}_{t-1} = T^{1/2}\delta_t\eta_t + \varepsilon_t + (\Gamma_0-\Gamma)'\mathbf{Y}_{t-1}$,

$$\begin{aligned} \max_{t\le T}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2 &\le 3T\max_{t:\delta_t=1}\eta_t^2 + 3\max_{t\le T}\varepsilon_t^2 + 3\sup_{\mathbb{A}_T}\|T^{1/2}(\Gamma-\Gamma_0)\|^2\max_{t\le T}\|T^{-1/2}\mathbf{Y}_t\|^2 \\ &= O_P(T) + O_P(\ln T) + O_P(\ln T) = O_P(T) \end{aligned} \tag{A.16}$$

uniformly on $\mathbb{A}_T$, so that

$$\sup_{\mathbb{A}_T}\exp\Big(\frac{\max_{t\le T}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{2(\sigma_\varepsilon^2 + T\omega^\eta)}\Big) \le \exp\Big(\frac{O_P(T)}{TQ_T}\Big) = \exp(O_P(1)) = O_P(1),$$

since Q_T is bounded away from zero in P-probability. Finally,

$$\begin{aligned} \min_{t:\delta_t=1}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2 &\ge T\min_{t:\delta_t=1}\eta_t^2 - 2T^{1/2}\max_{t:\delta_t=1}|\eta_t|\big(\max_{t\le T}|\varepsilon_t| + \sup_{\mathbb{A}_T}\|T^{1/2}(\Gamma-\Gamma_0)\|\max_{t\le T}\|T^{-1/2}\mathbf{Y}_t\|\big) \\ &= T\min_{t:\delta_t=1}\eta_t^2 + o_P(T^{3/4}). \end{aligned}$$

It follows that $\inf_{\mathbb{A}_T}\min_{t:\delta_t=1}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2 \xrightarrow{P} \infty$ at a linear rate, since $\min_{t:\delta_t=1}\eta_t^2$ is bounded away from 0 in P-probability, and hence,

$$\sup_{\mathbb{A}_T}\exp\Big(-\frac{\min_{t:\delta_t=1}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{2\sigma_\varepsilon^2}\Big) \le \exp\Big(-\frac{\inf_{\mathbb{A}_T}\min_{t:\delta_t=1}(\Delta y_t - \Gamma'\mathbf{Y}_{t-1})^2}{4\sigma_{\varepsilon 0}^2}\Big) \xrightarrow{P} 0$$

faster-than-algebraically. By combining the above magnitude orders, we can conclude that $\sup_{\mathbb{A}_T} \sum_{t=1}^T \delta_t(1-d_t(\omega)) \xrightarrow{P} 0$ faster-than-algebraically, which is the second relation in (a). Combining it with (A.15) yields the first relation there.

Given part (a), the remaining conclusions of the lemma follow naturally. We proceed with part (b). Let $w_t^\delta(\omega) := (1-\delta_t)/\sigma_\varepsilon^2$, so that $w_t(\omega) - w_t^\delta = K_1(\omega)(d_t(\omega)-\delta_t) + K_2(\omega)\delta_t$, with

$$\sup_{\mathbb{A}_T} |K_1(\omega)| = \sup_{\mathbb{A}_T} \Big|\frac{1}{\sigma_\varepsilon^2} - \frac{1}{\sigma_\varepsilon^2 + T\omega^\eta}\Big| = O_P(1) \quad \text{and} \quad \sup_{\mathbb{A}_T} |K_2(\omega)| = \sup_{\mathbb{A}_T} \frac{1}{\sigma_\varepsilon^2 + T\omega^\eta} = O_P(T^{-1}).$$

We show that if $w_t(\omega)$ are replaced by w_t^δ in the expression for Φ^Γ, the effect is asymptotically negligible. Specifically,

$$(\Phi^\Gamma(\omega) - \Gamma_0')D_T^{-1} = \sum_{t=1}^T w_t(\omega)(\varepsilon_t + \delta_t\theta_t)(D_T\mathbf{Y}_{t-1})' \Big[\sum_{t=1}^T w_t(\omega) D_T\mathbf{Y}_{t-1}(D_T\mathbf{Y}_{t-1})'\Big]^{-1}, \tag{A.17}$$

where, first, $\| \sum_{t=1}^T (w_t(\omega) - w_t^\delta)(\varepsilon_t + \delta_t\theta_t)(D_T\mathbf{Y}_{t-1})'\|$ is bounded by

$$\| \sum_{t=1}^T \delta_t \left[K_1(\omega)(d_t(\omega)-1) + K_2(\omega)\right](\varepsilon_t + \theta_t)(D_T\mathbf{Y}_{t-1})'\| + |K_1(\omega)| \| \sum_{t=1}^T (1-\delta_t)d_t(\omega)\varepsilon_t(D_T\mathbf{Y}_{t-1})'\|.$$

The two norms are evaluated separately. The first of them does not exceed

$$\begin{aligned} &(\max_{t\le T} |\varepsilon_t| + T^{1/2} \max_{t:\delta_t=1} |\eta_t|) \max_{t\le T} \|D_T\mathbf{Y}_{t-1}\| \Big[|K_1(\omega)| \sum_{t=1}^T \delta_t(1-d_t(\omega)) + |K_2(\omega)| N_T\Big] \\ = \;& O_P(T)\Big[\sum_{t=1}^T \delta_t(1-d_t(\omega)) + O_P(T^{-1})\Big] = O_P(1), \end{aligned}$$

see part (a). The second norm is bounded by

$$\max_{t\le T} |\varepsilon_t| \max_{t\le T} \|D_T\mathbf{Y}_{t-1}\| \sum_{t=1}^T (1-\delta_t)d_t(\omega) = O_P(T^{2\rho})$$

uniformly on $\mathbb{A}_T$, by (A.15) and the Gaussianity of ε_t. We conclude that, also uniformly,

$$\sum_{t=1}^T w_t(\omega)(\varepsilon_t + \delta_t\theta_t)(D_T\mathbf{Y}_{t-1})' = \frac{1}{\sigma_\varepsilon^2} \sum_{t=1}^T (1-\delta_t)(\varepsilon_t + \delta_t\theta_t)(D_T\mathbf{Y}_{t-1})' + O_P(T^{2\rho}). \tag{A.18}$$

Further, similarly, $\| \sum_{t=1}^T (w_t(\omega) - w_t^\delta)(D_T\mathbf{Y}_{t-1})(D_T\mathbf{Y}_{t-1})'\|$ is bounded by

$$\begin{aligned} &\| \sum_{t=1}^T \left[K_1(\omega)(d_t(\omega)-\delta_t) + K_2(\omega)\delta_t\right](D_T\mathbf{Y}_{t-1})(D_T\mathbf{Y}_{t-1})'\| \\ \le \;& \max_{t\le T} \|D_T\mathbf{Y}_{t-1}\|^2 \Big[|K_1(\omega)| \sum_{t=1}^T |d_t(\omega)-\delta_t| + |K_2(\omega)| N_T\Big] = O_P(T^{\rho+1/2}), \end{aligned}$$

uniformly on $\mathbb{A}_T$, so that, also uniformly,

$$\sum_{t=1}^{T} w_t(\omega)(D_T\mathbf{Y}_{t-1})(D_T\mathbf{Y}_{t-1})' = \frac{1}{\sigma_\varepsilon^2}\sum_{t=1}^{T}(1-\delta_t)(D_T\mathbf{Y}_{t-1})(D_T\mathbf{Y}_{t-1})' + O_P(T^{\rho+1/2}). \quad \text{(A.19)}$$

Inserting this and (A.18) into (A.17), we see that $(\Phi^\Gamma(\omega)-\Gamma_0)D_T^{-1}T^{1/2}$ equals

$$T^{-1/2}\sum_{t=1}^{T}(1-\delta_t)(\varepsilon_t+\delta_t\theta_t)(D_T\mathbf{Y}_{t-1})'\Big[T^{-1}\sum_{t=1}^{T}(1-\delta_t)D_T\mathbf{Y}_{t-1}(D_T\mathbf{Y}_{t-1})'\Big]^{-1} + o_P(1)$$

uniformly on $\mathbb{A}_T$, since the matrix in brackets convergence to a positive definite limit, see Lemma A.3. The main term in the above display is $(\tilde{\Gamma}-\Gamma_0)'D_T^{-1}T^{1/2}$, which proves (b).

Consider next part (c). We have $\sup_{\mathbb{A}_T}|\Phi^\lambda(\omega)-N_T| \le \sup_{\mathbb{A}_T}\sum_{t=1}^T|d_t(\omega)-\delta_t| = o_P(1)$ by (a). From here,

$$\begin{aligned}\alpha_T &:= \sup_{\mathbb{A}_T}\|(\Phi^\varepsilon(\omega),\Phi^\eta(\omega)) - T^{-1}\sum_{t=1}^{T}(1-d_t(\omega),N_T^{-1}d_t(\omega))(\Delta y_t-\Gamma'\mathbf{Y}_{t-1})^2\| \\ &= \sup_{\mathbb{A}_T}\|(\Phi^\varepsilon(\omega),\Phi^\eta(\omega)+T^{-1}\sigma_\varepsilon^2)\,\mathrm{diag}(\frac{\Phi^\lambda(\omega)}{T},\frac{N_T-\Phi^\lambda(\omega)}{N_T}) - (0,T^{-1}\sigma_\varepsilon^2)\| = o_P(1).\end{aligned}$$

Next, from the triangle inequality,

$$\sup_{\mathbb{A}_T}\|(\Phi^\varepsilon(\omega),\Phi^\eta(\omega)) - T^{-1}\sum_{t=1}^{T}(1-\delta_t,N_T^{-1}\delta_t)(\varepsilon_t+\delta_t\theta_t)^2\| \le \alpha_T+\beta_T+\gamma_T, \quad \text{(A.20)}$$

with α_T defined and evaluated above, and with

$$\begin{aligned}\beta_T &:= T^{-1}\sup_{\mathbb{A}_T}\|\sum_{t=1}^{T}[(1-d_t(\omega),N_T^{-1}d_t(\omega))-(1-\delta_t,N_T^{-1}\delta_t)](\Delta y_t-\Gamma'\mathbf{Y}_{t-1})^2\| \\ &\le T^{-1}(1+N_T^{-2})^{1/2}\max_{t\le T}(\Delta y_t-\Gamma'\mathbf{Y}_{t-1})^2\|\sum_{t=1}^{T}|\delta_t-d_t(\omega)| = O_P(T^{\rho-1/2})\end{aligned}$$

using the upper bound from (A.16) for $\max_{t\le T}(\Delta y_t-\Gamma'\mathbf{Y}_{t-1})^2$, and

$$\gamma_T := T^{-1}\sup_{\mathbb{A}_T}\|\sum_{t=1}^{T}(1-\delta_t,N_T^{-1}\delta_t)(\Delta y_t-\Gamma'\mathbf{Y}_{t-1})^2 - \sum_{t=1}^{T}(1-\delta_t,N_T^{-1}\delta_t)(\varepsilon_t+\delta_t\theta_t)^2\|.$$

As $\Delta y_t-\Gamma'\mathbf{Y}_{t-1} = \varepsilon_t+\delta_t\theta_t+(\Gamma-\Gamma_0)'\mathbf{Y}_{t-1}$, we have for $v\in\{\delta,1-\delta\}$ that

$$T^{-1}\sum_{t=1}^{T}v_t((\Delta y_t-\Gamma'\mathbf{Y}_{t-1})^2-(\varepsilon_t+\delta_t\theta_t)^2) = (\Gamma-\Gamma_0)'D_T^{-1}S_{11}^vD_T^{-1}(\Gamma-\Gamma_0)+2(\Gamma-\Gamma_0)'D_T^{-1}S_{10}^v$$

with $\sup_{\mathbb{A}_T}\|D_T^{-1}(\Gamma-\Gamma_0)\| = o(1)$ and

$$S_{11}^v := T^{-1}\sum_{t=1}^{T}v_t(D_T\mathbf{Y}_{t-1})(D_T\mathbf{Y}_{t-1})' = O_P(1), \quad S_{10}^v := T^{-1}\sum_{t=1}^{T}v_t(D_T\mathbf{Y}_{t-1})(\varepsilon_t+v_t\delta_t\theta_t) = O_P(1),$$

the display as a consequence of Lemmas A.2 and A.3. Hence, also $\gamma_T = o_P(1)$, and so is $\alpha_T + \beta_T + \gamma_T$ in (A.20). As $T^{-1}\sum_{t=1}^{T}(1-\delta_t, N_T^{-1}\delta_t)(\varepsilon_t + \delta_t\theta_t)^2 = (T^{-1}\sum_{t=1}^{T}\varepsilon_t^2, Q_T) + O_P(T^{-1/2})$, the proof is completed. ■

We are now ready to prove Theorem 1.

Proof of Theorem 1. We define $\check{\theta}$, whose existence is asserted in (a), as $\omega(\check{\zeta})$, where $\check{\zeta} = (\check{\Gamma}', \check{\sigma}^2, \check{\zeta}^{\eta}, \check{\zeta}^{\lambda})'$ is a measurable global maximizer of $\Lambda \circ \omega$ on $\mathbb{A}_T$. The existence of $\check{\zeta}$ follows, e.g., from Property 24.1 in Gourieroux and Monfort (1995). To show that $\check{\theta}$ is a local maximizer of Λ w.p.a.1, we check that $\check{\zeta}$ is interior for $\mathbb{A}_T$ w.p.a.1. Specifically, we give the details for interiority of $\check{\Gamma}$ for $\mathbb{A}_T^{\Gamma}$ and omit the rest, which is similar.

Since the function $\|T^{1/2}D_T^{-1}((\cdot) - \Gamma_0)\|$ is differentiable at all points different from Γ_0, and Γ_0 is interior for $\mathbb{A}_T^{\Gamma}$, it follows that $\check{\Gamma}$ satisfies the first-order condition

$$\frac{\partial(\Lambda \circ \omega)}{\partial \Gamma'}|_{\check{\zeta}} - \mu \frac{T^{1/2}D_T^{-1}(\check{\Gamma} - \Gamma_0)}{(\ln T)^{1/4}} = 0, \tag{A.21}$$

where $\mu \geq 0$ is a Lagrange multiplier such that $\mu(\|T^{1/2}D_T^{-1}(\check{\Gamma} - \Gamma_0)\| - (\ln T)^{1/4}) = 0$. Inserting the expression for the derivative yields

$$\sum_{t=1}^{T} w_t(\check{\theta})\mathbf{Y}_{t-1}(\Delta y_t - \mathbf{Y}_{t-1}'\check{\Gamma}) = \mu T^{1/2}(\ln T)^{-1/4}D_T^{-1}(\check{\Gamma} - \Gamma_0),$$

and further, since $\Delta y_t = \mathbf{Y}_{t-1}'\Gamma_0 + \varepsilon_t + \delta_t\theta_t$,

$$\sum_{t=1}^{T} w_t(\check{\theta})D_T\mathbf{Y}_{t-1}(\varepsilon_t + \delta_t\theta_t + (D_T\mathbf{Y}_{t-1})'D_T^{-1}(\Gamma_0 - \check{\Gamma})) = \mu T^{1/2}(\ln T)^{-1/4}(\check{\Gamma} - \Gamma_0).$$

Using (A.18) and introducing $S_{11}^{1-\delta}(\check{\theta}) := T^{-1}\sum_{t=1}^{T} w_t(\check{\theta})D_T\mathbf{Y}_{t-1}(D_T\mathbf{Y}_{t-1})'$, we find next that

$$TS_{10}^{1-\delta} + TS_{11}^{1-\delta}(\check{\theta})D_T^{-1}(\Gamma_0 - \check{\Gamma}) + O_P(T^{1/2}) = \mu\check{\sigma}^2 T^{1/2}(\ln T)^{-1/4}(\check{\Gamma} - \Gamma_0),$$

where $S_{10}^{1-\delta} = T^{-1}\sum_{t=1}^{T}(1-\delta_t)D_T\mathbf{Y}_{t-1}\varepsilon_t = O_P(T^{-1/2})$. Premultiplication by $(\check{\Gamma} - \Gamma_0)'D_T^{-1}$ gives

$$(\check{\Gamma} - \Gamma_0)'D_T^{-1}[TS_{11}^{1-\delta}(\check{\theta})]D_T^{-1}(\Gamma_0 - \check{\Gamma}) + (\check{\Gamma} - \Gamma_0)'D_T^{-1}O_P(T^{1/2}) = \mu\check{\sigma}^2\frac{(\check{\Gamma} - \Gamma_0)'D_T^{-1}(\check{\Gamma} - \Gamma_0)}{T^{-1/2}(\ln T)^{1/4}}.$$

Finally, by majorizing the left side, for outcomes such that $\mu > 0$ (and hence, $\check{\Gamma} \neq \Gamma_0$), it follows that

$$-\|T^{1/2}D_T^{-1}(\check{\Gamma} - \Gamma_0)\|^2\lambda_{\min}(S_{11}^{1-\delta}(\check{\theta})) + (\check{\Gamma} - \Gamma_0)'D_T^{-1}O_P(T^{1/2}) > 0.$$

However, for such outcomes the defining constraint of $\mathbb{A}_T^{\Gamma}$ constraint is binding, so that

$$-(\ln T)^{1/2}\lambda_{\min}(S_{11}^{1-\delta}(\check{\theta})) + O_P((\ln T)^{1/4}) > 0.$$

As $\lambda_{\min}(S_{11}^{1-\delta}(\check{\theta})) = \sigma_{\varepsilon}^{-2}\lambda_{\min}(S_{11}^{1-\delta}) + o_P(1)$ by (A.19), and $\lambda_{\min}(S_{11}^{1-\delta})$ is bounded away from zero in P-probability by Lemma A.3, the inequality in the above display can only hold with

P-probability approaching zero. Consequently, $P(\{\mu > 0\}) \to 0$, meaning that $\check{\Gamma}$ w.p.a.1 satisfies the first-order condition (A.21) in the form $(\partial(\Lambda \circ \omega)/\partial\Gamma')|_{\check{\zeta}} = 0$, or equivalently, $\check{\Gamma}' = \Phi_\Gamma(\check{\theta})$. From Lemma A.4(b) and the fact that $T^{1/2}D_T^{-1}(\tilde{\Gamma} - \Gamma_0) = O_P(1)$ it follows that $T^{1/2}D_T^{-1}(\check{\Gamma} - \Gamma_0) = O_P(1)$, and from the definition of $\mathbb{A}_T^\Gamma$, $\check{\Gamma}$ is interior for $\mathbb{A}_T^\Gamma$ w.p.a.1. A similar argument for the other components of $\check{\zeta}$ lets us conclude that $\check{\theta}$ is a local maximizer of $\Lambda(\theta)$ w.p.a.1.

The remaining asserted properties of $\check{\theta}$ are straightforward from $\check{\zeta} \in \mathbb{A}_T$ and Lemma A.4. ■

Proof of Corollary 1. Consistency in part (b) and the statement about ADF_α^Q follow from Theorem 1(c) and Proposition 3(a), whereas the statement about ADF_t^Q follows from Theorem 1(c) and (A.19), with w_t evaluated at $\check{\theta}$. For asymptotic normality, note that by Theorem 1(c) it is enough to establish it for the dummy variables estimator. From (A.14) and the representation $\mathbf{Z}_t = \mathbf{U}_t - (c/T)\iota(L)y_t$ (see the proof of Lemma A.2 for notation), we have that $T^{1/2}(\tilde{\Xi} - \Xi)'$ equals

$$\left(Var(\mathbf{U}_t^\varepsilon) + T^{-1}\sum_{t=1}^{T-1}(1-\delta_t)\mathbf{U}_t^\theta(\mathbf{U}_t^\theta)'\right)^{-1} T^{-1/2}\sum_{t=1}^{T}(1-\delta_t)(\mathbf{U}_{t-1}^\varepsilon + \mathbf{U}_{t-1}^\theta)\varepsilon_t + o_P(1).$$

By the assumed independence of $\{\mathbf{U}_t^\varepsilon\}$ and $\{\mathbf{U}_t^\theta\}$, the main term above converges weakly to $N(0,1)$ conditionally on $\{\mathbf{U}_t^\theta\}$, and hence, also unconditionally. ■

References

Abadir K. & Lucas A. (2004) A comparison of minimum MSE and maximum power for the nearly integrated non-Gaussian model, *Journal of Econometrics* 119, 45–71.

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

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

Balke N.S. & T.B. Fomby (1994) Large shocks, small shocks and economic fluctuations: outliers in macroeconomic time series, *Journal of Applied Econometrics* 9, 181–200.

Berk K.N. (1974) Consistent autoregressive spectral estimates, *Annals of Statistics* 2, 489–502.

Billingsley P. (1968) *Convergence of Probability Measures*, Wiley: New York.

Bohn-Nielsen H. (2004) Cointegration analysis in the presence of outliers, *Econometrics Journal* 7, 249–271.

Boswijk H.P. & Lucas A. (2002) Semi-nonparametric cointegration testing, *Journal of Econometrics* 108, 253–280.

Boswijk H.P. (2005) Adaptive testing for a unit root with nonstationary volatility, Working paper, University of Amsterdam.

Box G.E.P. & G.C. Tiao (1975) Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical Association* 70, 70–79.

Burridge P. & A.M.R. Taylor (2006) Additive outlier detection via extreme-value theory, *Journal of Time Series Analysis* 27, 685–701.

Chang I., G.C. Tiao & C. Chen (1988) Estimation of Time Series Parameters in the Presence of Outliers, *Technometrics* 30, 193–204.

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

Cox D.D. & I. Llatas (1991) Maximum likelihood type estimation for nearly nonstationary autoregressive time series, *Annals of Statistics* 19,1109–1128.

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

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

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

Franses, P.H., T. Kloek & A. Lucas (1999) Outlier robust analysis of long-run marketing effects for weekly scanning data, *Journal of Econometrics* 89, 293–315.

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

Georgiev I. (2005) A mixture distribution factor model for multivariate outliers, manuscript, Universidade Nova de Lisboa, http://docentes.fe.unl.pt/~igrg/mdfm.pdf

Georgiev I. (2006) Asymptotics for cointegrated processes with infrequent stochastic level shifts and outliers, *Econometric Theory*, forthcoming.

Gourieroux C., A. Monfort (1995), *Statistics and Econometric Models*, New York: Cambridge University Press.

Hendry D. F. & K. Juselius (2001) Explaining cointegration analysis: Part II, *Energy Journal* 22, 75–120.

Hodgson D. (1998a) Adaptive estimation of cointegrating regressions with ARMA errors, *Journal of Econometrics* 85, 231–268.

Hodgson D. (1998b) Adaptive estimation of error correction models, *Econometric Theory* 14, 44–69.

Johansen S. (1996) *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford: Oxford University Press.

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

Lanne M., H. Lütkepohl & P. Saikkonen (2002) Unit root tests in the presence of innovational outliers, in Klein I. and S. Mittnik, (eds.) *Contributions to Modern Econometrics*, Dordrecht: Kluwer Academic Publishers, 151–167.

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

Lucas A. (1995a) Unit root tests based on M estimators. *Econometric Theory* 11, 331–346.

Lucas A. (1995b) An outlier robust unit root test with an application to the extended Nelson–Plosser data, *Journal of Econometrics* 66, 153–173.

Lucas A. (1996) *Outlier Robust Unit Root Analysis.* Amsterdam: Thesis Publishers, http://staff.feweb.vu.nl/alucas/thesis/default.htm

Lucas A. (1997). Cointegration testing using pseudo likelihood ratio tests. Econometric Theory 13(2): 149-169

Lucas A. (1998) Inference on cointegrating ranks using LR and LM tests based on pseudo-likelihoods, *Econometric Reviews* 17, 185–214

Lütkepohl H., C. Müller & P. Saikkonen (2001) Unit root tests for time series with a structural break when the break point is known, in C. Hsiao, K. Morimune & J. Powell (eds.), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, Cambridge: Cambridge University Press, 327–348.

Martin R.D. & J. Jong (1977) Asymptotic properties of robust generalized M-estimates for the first order autoregressive parameter, Bell Laboratories Memorandum, Murray Hill, N.J.

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

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

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

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

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

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

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

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

Phillips P.C.B. & Z. Xiao (1998) A primer on unit root testing, *Journal of Economic Surveys* 12, 423–470.

Rothenberg T.J. & Stock J.H. (1997) Inference in a nearly integrated autoregressive model with nonnormal innovations, *Journal of Econometrics* 80, 269–286.

Shin D. W., S. Sarkar & J. H. Lee (1996) Unit Root Tests for Time Series with Outliers, *Statistics and Probability Letters* 30, 189–97.

Stock J. (1994) Unit roots, structural breaks and trends, in Engle R. and D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, New York: North Holland, 2740–2841.

Tsay R. (1988) Outliers, level shifts, and variance changes in time series, *Journal of Forecasting* 7, 1–20.

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

Xiao Z. & L.R. Lima (2004) Purchasing power parity and unit root tests: a robust analysis, Fundação Getulio Vargas, Ensaios Econômicos 552.

Xiao Z. & Peter C.B. Phillips (1998) An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy, *Econometrics Journal* 1, 27–43.

Table 1: Asymptotic distributions of the ADF tests in the presence of large innovational outliers.

	t - based test		
	no outliers	Assumption $\mathcal{S}'$	Assumption $\mathcal{S}'$, at least one IO of large size
ADF	$-c\left(\int B_c^2\right)^{1/2}+\frac{\int B_c dB}{(\int B_c^2)^{1/2}}$	$\frac{1}{\sqrt{1+[C]/\sigma_\varepsilon^2}}\Big[-c\left(\int H_c^2\right)^{1/2}+\frac{\int H_c dH}{(\int H_c^2)^{1/2}}\Big]$	$\frac{1}{\sqrt{[C]}}\Big[-c\left(\int C_c^2\right)^{1/2}+\frac{\int C_c dC}{(\int C_c^2)^{1/2}}\Big]$
ADF^D	$-c\left(\int B_c^2\right)^{1/2}+\frac{\int B_c dB}{(\int B_c^2)^{1/2}}$	$-c\left(\int H_c^2\right)^{1/2}+\frac{\int H_c dB}{(\int H_c^2)^{1/2}}$	$\begin{cases} N(0,1), & \text{if } c=0 \\ -\infty, & \text{if } c>0 \end{cases}$

	coefficient test		
	no outliers	Assumption $\mathcal{S}'$	Assumption $\mathcal{S}'$, at least one IO of large size
ADF	$-c+\frac{\int B_c dB}{\int B_c^2}$	$-c+\frac{\int H_c dH}{\int H_c^2}$	$-c+\frac{\int C_c dC}{\int C_c^2}$
ADF^D	$-c+\frac{\int B_c dB}{\int B_c^2}$	$-c+\frac{\int H_c dB}{\int H_c^2}$	$\begin{cases} 0, & \text{if } c=0 \\ -c, & \text{if } c>0 \end{cases}$

Notes: In the table, 'ADF' denotes tests based on the standard ADF regression, while 'ADF^D' refers to tests based on an ADF regression augumented by the inclusion of impulse dummies (one for each IO). The first column refers to the case of no outliers ($[C]=0$). The second column refers to the case of outliers under Assumption $\mathcal{S}'$. The third column refers to thecase of where $[C]>0$ (at least one IO) and the limiting process C of Assumption $\mathcal{S}$ is replaced hC, with $h\to\infty$. The $N(0,1)$ limit under $\mathcal{S}'$ obtains under the extra assumption that C and B are independent.

TABLE 2: EMPIRICAL SIZE, SIZE ADJUSTED POWER AND EMPIRICAL REJECTION FREQUENCIES OF STANDARD (ADF) AND DUMMY-BASED (ADF^D) ADF TESTS. RAW DATA.

Size		Model S_0		Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α	ADF_t	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D
100	−0.5	5.1	5.1	4.9	5.0	3.1	5.1	4.8	4.7	1.9	5.1	5.2	5.1	2.1	6.1	1.7	1.8	1.0	5.4
	0	5.2	5.1	5.0	5.0	3.1	5.1	5.4	5.4	2.0	5.6	5.2	5.1	1.9	6.1	3.4	3.6	0.8	5.4
	0.5	5.5	5.4	5.2	4.9	3.1	5.1	5.0	5.1	2.0	5.2	5.4	5.1	2.0	6.0	6.6	6.5	1.0	5.5
200	−0.5	5.5	5.4	5.2	5.2	3.0	4.9	4.8	4.9	1.9	5.3	4.9	4.8	1.3	5.2	1.5	1.7	0.7	4.8
	0	4.8	5.0	5.0	5.0	3.1	4.8	4.9	5.0	1.8	4.8	5.0	4.8	1.6	5.2	3.2	3.4	0.8	5.0
	0.5	5.0	5.0	5.1	5.0	3.1	5.0	5.0	5.0	1.8	4.8	5.2	5.0	1.7	5.1	5.8	5.9	0.9	5.4
400	−0.5	5.1	5.1	4.7	5.0	3.0	5.0	4.6	4.6	1.8	4.9	4.7	4.5	1.3	4.7	1.5	1.7	0.6	5.2
	0	4.9	4.9	4.9	5.0	2.9	4.7	4.6	4.6	1.4	4.6	5.1	4.9	1.5	5.0	3.0	3.2	0.6	4.8
	0.5	5.0	4.9	5.0	5.1	2.9	4.5	4.6	4.6	1.7	4.8	5.1	5.1	1.4	5.1	5.9	6.0	0.8	5.3
Power		Model S_0		Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α	ADF_t	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D
100	−0.5	50.9	51.4	50.6	51.4	64.3	66.0	52.4	54.1	75.3	75.4	49.1	50.6	76.7	75.6	57.0	59.2	91.7	87.0
	0	48.0	49.3	48.9	50.1	62.2	64.8	46.2	47.1	70.0	72.2	48.4	48.8	75.2	73.2	52.6	54.4	90.6	86.7
	0.5	42.4	43.2	44.5	46.8	58.4	60.1	43.7	44.8	67.4	69.3	42.8	43.9	70.0	67.8	44.4	47.1	85.0	83.2
200	−0.5	45.7	45.8	49.9	50.7	62.9	67.2	49.3	50.8	74.1	74.3	50.1	51.1	80.2	77.5	56.7	58.2	92.5	89.3
	0	50.3	50.1	49.3	50.5	63.6	65.9	48.6	50.0	75.1	76.7	48.7	49.8	78.3	77.4	52.9	55.4	92.2	88.7
	0.5	46.4	46.4	46.3	48.1	57.9	61.5	45.1	46.6	71.2	73.3	46.1	46.8	75.1	75.1	52.9	54.1	88.5	86.0
400	−0.5	47.8	48.4	49.9	50.1	63.4	65.1	51.2	52.1	76.5	75.6	51.4	52.3	83.5	80.8	55.4	56.6	92.3	88.5
	0	50.5	50.3	49.4	50.6	63.8	66.5	49.8	51.2	76.5	77.4	47.5	48.9	78.8	78.3	55.6	57.2	93.5	89.1
	0.5	48.0	48.6	48.3	48.0	62.9	65.7	49.5	50.6	74.7	74.9	46.9	47.4	78.3	76.7	53.9	55.8	90.7	87.0
Empirical power		Model S_0		Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α	ADF_t	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D
100	−0.5	51.3	52.1	49.7	51.4	47.9	66.9	49.6	52.1	46.5	75.8	51.0	51.5	46.5	78.8	25.0	27.7	45.0	87.9
	0	49.2	50.0	49.4	50.0	47.0	65.3	48.4	49.9	45.1	74.2	49.5	49.4	43.8	76.5	39.4	41.8	42.3	87.4
	0.5	45.2	45.3	45.6	46.3	42.8	60.4	44.1	45.8	39.5	69.8	45.3	44.6	38.4	71.7	53.5	55.4	36.4	84.7
200	−0.5	48.8	49.0	50.8	52.2	47.8	66.5	48.2	50.2	44.5	75.3	49.3	49.6	43.9	78.2	24.4	26.7	42.7	88.9
	0	49.3	49.6	49.1	50.5	46.9	64.8	48.2	49.8	44.8	75.7	48.5	48.8	42.9	78.1	39.6	42.1	41.8	88.8
	0.5	46.2	46.2	46.9	47.8	44.0	61.5	45.0	46.2	41.5	72.7	46.9	46.8	40.8	75.7	57.4	59.7	38.7	87.0
400	−0.5	48.1	48.9	48.4	49.9	46.0	65.3	48.0	49.6	44.2	75.5	48.6	49.3	43.6	79.9	23.5	25.3	41.5	88.8
	0	49.0	49.4	48.6	49.8	46.2	64.3	47.2	48.9	44.0	75.8	48.3	48.4	43.6	78.6	38.9	41.4	40.9	88.8
	0.5	47.9	48.0	48.2	49.2	45.5	63.4	46.9	48.4	43.0	74.4	47.4	47.9	41.3	77.1	59.6	61.9	40.2	88.0

TABLE 3: EMPIRICAL SIZE, SIZE ADJUSTED POWER AND EMPIRICAL REJECTION FREQUENCIES OF STANDARD (ADF) AND DUMMY-BASED (ADF^D) ADF TESTS. STUDENT t INNOVATIONS, RAW DATA.

Size		Model S_0		Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α	ADF_t	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D
100	−0.5	4.6	4.7	5.1	5.1	2.9	5.2	4.8	4.8	1.7	5.1	5.0	4.9	1.5	5.6	1.7	1.8	0.8	5.4
	0	5.5	5.6	5.4	5.4	3.4	5.4	4.8	4.7	1.7	5.0	5.2	5.3	1.9	5.9	3.3	3.4	0.8	5.1
	0.5	5.5	5.2	5.0	4.9	3.1	4.8	4.6	4.6	1.8	5.2	5.5	5.2	2.1	5.7	6.2	6.2	0.8	5.1
200	−0.5	4.8	5.0	5.1	5.2	3.2	5.2	4.7	4.9	2.0	5.0	4.6	4.4	1.5	4.8	1.5	1.6	0.8	4.9
	0	5.0	5.1	4.6	4.7	3.0	4.8	4.6	4.7	1.9	5.3	4.6	4.6	1.4	5.2	3.2	3.5	0.9	4.9
	0.5	5.1	4.9	5.0	5.0	2.9	5.1	4.9	4.8	1.9	5.2	5.2	5.2	1.6	5.1	6.2	6.2	0.7	4.7
400	−0.5	4.7	4.7	4.7	4.7	2.9	4.8	4.6	4.8	1.8	4.9	4.8	4.9	1.6	5.0	1.4	1.5	0.7	5.1
	0	4.5	4.5	5.1	5.2	3.3	5.3	4.4	4.5	1.5	4.4	5.3	5.3	1.7	5.3	2.8	3.0	0.6	4.8
	0.5	5.1	5.0	5.2	5.1	2.9	4.9	4.6	4.6	1.7	4.7	4.9	4.8	1.4	4.5	6.2	6.5	0.7	4.8
Power		Model S_0		Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α	ADF_t	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D
100	−0.5	55.3	54.8	50.0	51.4	65.2	65.8	51.3	53.3	77.3	75.9	51.4	52.6	80.4	77.2	55.0	57.5	92.0	87.4
	0	45.5	46.2	46.1	47.4	60.7	62.9	49.1	51.0	75.1	74.5	46.2	46.7	75.3	73.9	53.7	55.1	91.3	87.9
	0.5	42.4	44.0	45.2	46.6	57.7	60.9	45.8	47.3	68.0	69.8	42.2	43.1	68.0	70.7	48.8	50.7	87.7	85.2
200	−0.5	50.9	50.4	49.4	50.8	64.1	65.9	50.6	52.1	76.5	76.3	52.1	52.9	83.1	80.4	56.6	58.3	92.7	89.0
	0	49.0	49.4	51.5	52.9	65.0	67.4	50.8	52.2	74.3	75.3	51.4	52.2	79.8	78.4	54.0	55.5	92.6	88.7
	0.5	45.6	47.5	46.1	47.4	58.6	61.0	45.9	47.0	70.2	72.3	44.9	45.0	76.7	75.4	51.3	52.7	90.8	88.2
400	−0.5	50.3	51.2	50.9	51.9	64.8	67.2	50.0	50.9	76.0	76.4	50.1	50.1	81.8	78.9	59.1	60.4	92.9	88.6
	0	51.5	51.4	48.1	49.2	59.7	63.2	51.1	52.2	78.7	78.2	47.0	47.8	78.2	77.8	57.0	59.3	92.9	88.9
	0.5	47.3	47.4	46.8	48.1	61.9	64.5	50.8	50.8	74.8	75.8	48.4	48.1	81.4	79.7	52.7	53.3	89.6	88.3
Empirical power		Model S_0		Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α	ADF_t	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α	ADF_t	ADF_α^D	ADF_t^D
100	−0.5	51.3	52.7	50.3	52.2	48.2	66.8	49.9	51.9	45.9	76.5	51.4	51.9	46.3	79.5	25.1	28.0	44.2	88.6
	0	48.7	49.1	48.7	50.0	46.7	65.0	47.9	49.5	45.0	74.5	48.5	48.6	44.3	77.3	40.1	42.4	42.8	88.4
	0.5	44.9	45.2	45.4	46.2	42.2	60.4	43.6	44.8	39.1	70.5	45.1	44.9	38.7	73.1	54.5	56.5	36.8	85.6
200	−0.5	49.5	50.4	50.3	51.8	48.1	67.0	48.9	50.6	44.6	76.3	48.9	49.9	43.9	79.4	23.9	26.2	42.9	88.9
	0	49.2	49.6	49.5	50.3	46.8	65.8	48.0	49.7	45.2	76.5	49.1	49.1	43.3	79.1	40.2	42.8	41.6	88.5
	0.5	46.5	46.9	46.4	47.3	43.6	61.7	45.0	46.5	41.5	72.9	46.0	46.2	40.6	75.7	58.0	60.6	39.0	87.4
400	−0.5	48.7	49.3	48.9	50.2	46.7	65.6	47.7	49.2	44.2	75.8	48.4	49.2	43.5	79.0	23.8	25.8	41.9	88.8
	0	48.7	49.2	48.8	49.9	45.2	64.5	47.7	49.4	44.3	75.9	49.2	49.5	43.4	79.2	39.8	42.5	41.9	88.6
	0.5	47.4	47.5	47.5	48.8	44.7	63.9	47.1	48.4	43.6	74.5	47.1	47.1	40.9	77.8	59.9	62.4	40.2	87.8

TABLE 4: EMPIRICAL SIZE, SIZE ADJUSTED POWER AND EMPIRICAL REJECTION FREQUENCIES OF ROBUST QML (ADF^Q) AND ROBUST M (ADF^L) ADF TESTS. GAUSSIAN INNOVATIONS, RAW DATA.

Size		Model S_0				Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L
100	−0.5	5.2	5.9	5.4	6.6	3.8	6.1	3.6	3.8	2.7	6.5	2.5	2.5	2.9	6.8	2.7	2.4	1.4	4.2	1.2	1.4
	0.0	5.2	5.8	5.4	6.2	3.7	6.1	3.8	3.8	3.1	7.0	2.8	2.4	2.8	6.4	2.7	2.5	2.0	4.7	1.6	2.3
	0.5	5.6	6.2	5.9	6.4	3.8	6.3	3.8	3.8	3.0	6.7	2.8	2.4	2.9	6.5	2.8	2.4	3.6	6.6	2.6	4.0
200	−0.5	5.5	6.0	5.6	5.8	3.1	5.6	3.1	3.5	2.2	5.9	2.2	2.6	1.8	5.7	1.7	2.1	0.8	4.7	0.8	2.2
	0.0	4.9	5.4	4.9	5.1	3.2	5.6	3.2	3.8	2.0	5.3	2.0	2.4	2.1	5.7	2.0	2.2	1.0	5.0	1.0	2.4
	0.5	5.0	5.7	5.4	5.6	3.4	5.7	3.5	3.8	2.0	5.4	2.0	2.5	2.1	5.4	2.1	2.0	1.3	5.7	1.1	3.9
400	−0.5	5.1	5.7	5.3	5.2	3.1	5.4	3.0	3.9	1.8	5.2	1.8	3.2	1.5	4.9	1.6	2.3	0.6	5.1	0.7	3.2
	0.0	4.9	5.5	4.9	5.0	3.0	5.1	3.2	3.9	1.5	4.9	1.5	2.9	1.8	5.5	1.8	2.6	0.7	5.0	0.7	3.4
	0.5	4.9	5.5	5.2	4.9	3.0	5.3	3.1	3.7	1.7	5.1	1.8	3.2	1.9	5.5	1.8	2.1	0.9	5.6	0.9	4.1
Power		Model S_0				Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L
100	−0.5	50.7	49.8	49.1	41.8	59.0	60.3	60.1	54.6	66.6	68.3	68.8	63.2	67.1	68.5	68.7	61.9	80.2	75.4	84.0	64.0
	0.0	48.1	48.2	47.2	43.6	55.2	59.6	57.0	52.6	61.3	63.9	64.6	61.8	65.6	68.7	67.1	60.2	71.7	75.0	78.7	59.5
	0.5	42.2	42.6	40.1	38.4	52.6	55.4	52.3	49.1	58.3	60.3	59.6	56.7	59.4	62.4	60.8	54.5	58.6	71.2	67.0	61.0
200	−0.5	45.8	45.7	44.8	42.6	61.8	64.0	62.8	59.4	71.4	72.3	70.6	67.9	74.7	74.1	75.7	68.3	90.3	87.7	90.1	78.7
	0.0	50.6	50.0	49.9	46.2	61.8	63.7	61.1	56.5	72.7	75.2	73.0	69.8	72.5	74.2	73.1	66.7	89.2	87.1	90.1	77.5
	0.5	46.1	46.4	44.2	40.8	56.7	59.1	57.2	53.8	69.5	71.6	69.5	65.7	70.3	72.7	71.7	65.7	84.3	85.2	84.9	75.2
400	−0.5	47.5	47.4	46.8	44.3	63.0	65.8	62.8	61.2	75.4	75.3	75.0	71.6	80.9	80.3	80.2	75.0	91.8	88.6	91.3	83.0
	0.0	49.3	49.8	49.1	44.7	63.1	65.8	62.4	60.0	75.9	77.1	75.4	71.3	76.9	77.4	76.2	72.7	92.7	88.9	92.0	83.4
	0.5	48.2	48.1	46.3	44.9	61.8	63.7	61.7	59.9	74.1	74.9	72.8	69.5	76.0	75.6	75.8	73.0	89.3	87.3	89.0	82.1
Empirical power		Model S_0				Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L	ADF_α^Q	ADF_t^Q	ADF_α^L	ADF_t^L
100	−0.5	51.4	54.5	51.5	50.3	48.8	66.7	48.7	47.0	46.9	74.6	46.8	45.5	47.7	76.1	47.1	44.7	35.8	71.0	38.4	28.2
	0.0	49.3	52.9	49.3	49.1	47.8	65.2	47.7	46.3	46.0	72.8	45.6	43.9	46.3	74.8	46.2	43.3	41.1	73.9	41.5	37.8
	0.5	45.2	48.4	45.7	45.0	43.8	61.2	43.8	41.5	40.5	67.8	40.7	39.3	41.1	69.5	40.6	38.7	45.4	77.8	43.2	54.1
200	−0.5	48.7	51.9	48.8	46.8	48.6	68.5	48.1	50.0	44.9	76.0	45.0	52.4	45.6	77.7	45.0	49.9	39.5	86.9	41.8	55.0
	0.0	49.6	52.6	49.6	46.6	47.4	66.7	46.9	49.0	45.1	76.1	44.9	52.3	44.1	77.0	43.7	48.9	41.6	87.1	42.0	58.5
	0.5	46.2	49.9	46.2	43.6	44.3	63.4	44.1	45.8	41.6	73.5	41.7	48.5	41.9	74.6	41.4	46.9	42.2	87.4	40.2	67.8
400	−0.5	48.3	52.0	48.1	45.7	46.0	67.7	46.0	53.9	44.3	76.3	44.2	60.5	44.1	79.9	44.0	58.6	40.3	88.9	41.9	74.7
	0.0	49.0	52.7	48.8	44.8	46.5	66.8	46.6	53.4	44.4	76.5	44.0	59.8	44.0	78.9	44.2	57.8	40.6	88.9	41.0	75.7
	0.5	47.8	52.2	47.4	44.2	45.6	65.7	45.2	52.0	43.0	75.5	42.9	58.7	42.3	77.4	42.2	56.3	41.4	88.6	40.4	78.4

TABLE 5: EMPIRICAL SIZE, SIZE ADJUSTED POWER AND EMPIRICAL REJECTION FREQUENCIES OF ROBUST QML (ADF^Q) AND ROBUST M (ADF^L) ADF TESTS. STUDENT t INNOVATIONS, RAW DATA.

Size		Model S_0				Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t
100	−0.5	3.5	6.5	3.1	2.4	2.6	6.9	2.3	1.7	1.4	6.6	1.4	0.9	1.5	7.1	1.5	0.8	0.8	5.4	0.7	0.6
	0.0	4.2	7.0	3.8	2.6	2.7	7.3	2.5	1.6	1.5	7.1	1.4	1.0	2.0	7.4	1.9	1.2	1.0	6.0	0.9	1.2
	0.5	4.3	7.1	4.1	2.7	2.5	6.6	2.2	1.4	1.6	6.9	1.5	0.8	2.1	6.7	2.1	1.1	1.6	6.6	1.4	1.7
200	−0.5	3.6	6.5	3.3	1.6	2.3	6.0	2.0	1.3	1.3	5.9	1.1	0.8	1.3	5.8	1.2	0.7	0.5	5.2	0.4	0.7
	0.0	3.7	6.6	3.5	2.1	2.2	5.8	2.0	1.5	1.3	5.9	1.1	0.9	1.1	5.8	1.0	0.8	0.5	5.3	0.5	0.9
	0.5	4.0	6.8	3.7	2.3	2.1	6.1	1.9	1.3	1.5	6.3	1.3	1.0	1.4	5.9	1.2	0.7	0.5	5.3	0.5	1.3
400	−0.5	3.5	6.5	3.1	1.4	2.0	5.5	1.7	1.1	1.3	5.6	1.1	1.0	1.1	5.3	0.9	0.7	0.5	5.4	0.4	1.0
	0.0	3.3	6.2	2.9	1.3	2.4	6.2	2.0	1.2	1.1	5.3	0.9	1.0	1.2	5.7	1.0	0.6	0.4	5.0	0.3	1.0
	0.5	3.7	6.3	3.3	1.4	2.1	5.9	1.9	1.1	1.1	5.5	0.9	1.1	1.1	5.0	0.9	0.7	0.5	5.3	0.4	1.2

Power		Model S_0				Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t
100	−0.5	61.3	61.1	63.6	56.0	69.1	70.6	71.9	64.1	83.9	80.4	84.3	75.3	81.1	78.6	82.7	71.1	92.2	87.2	94.0	77.0
	0.0	53.4	56.8	55.9	52.1	66.3	68.5	68.6	63.3	79.9	78.0	80.9	73.5	74.9	75.2	77.0	67.8	90.6	85.9	92.3	73.2
	0.5	48.9	51.6	50.4	47.4	63.6	65.1	66.2	60.1	73.8	72.4	75.0	67.8	68.2	72.6	70.6	64.1	82.6	85.2	85.4	72.5
200	−0.5	58.6	60.3	60.8	55.7	72.0	73.7	74.6	69.8	83.6	81.5	86.3	77.0	85.9	83.5	88.2	79.6	95.8	92.1	96.8	85.8
	0.0	55.9	58.9	58.9	53.1	72.7	73.8	74.8	69.6	82.9	82.3	85.2	78.8	84.3	82.0	86.0	78.1	95.2	92.0	96.0	84.7
	0.5	52.5	53.5	54.6	49.6	68.0	69.7	71.2	65.6	76.2	77.1	79.8	74.1	82.2	79.9	84.7	76.5	93.9	91.8	94.7	84.1
400	−0.5	58.5	59.7	61.3	58.8	74.8	74.5	77.6	72.3	85.1	82.2	87.7	79.7	86.9	84.4	89.7	82.9	96.2	92.4	97.1	90.4
	0.0	59.7	61.6	62.5	59.6	68.9	71.1	73.1	70.1	85.2	83.3	87.2	81.7	85.4	83.1	88.5	82.1	96.3	93.5	96.7	89.9
	0.5	54.5	59.0	57.8	58.0	69.3	71.5	72.9	71.5	81.4	81.3	85.2	79.6	85.4	84.0	88.1	82.2	94.8	92.3	95.9	88.6

Empirical power		Model S_0				Model S_2				Model S_4				Model S_r				Model S_c			
T	γ	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t	ADF^Q_α	ADF^Q_t	ADF^L_α	ADF^L_t
100	−0.5	50.1	68.6	50.0	39.1	47.3	78.7	46.8	38.1	45.1	85.6	44.8	37.5	46.1	84.4	46.1	36.7	37.6	88.1	39.4	29.7
	0.0	47.1	65.4	46.9	36.5	45.6	76.8	44.8	36.4	43.4	83.5	43.2	37.2	43.9	83.0	43.5	36.3	40.0	88.3	40.1	35.1
	0.5	43.4	61.5	42.8	33.5	40.9	72.2	40.3	32.7	37.3	79.0	37.5	32.0	38.9	78.8	38.1	32.1	40.4	88.9	38.6	48.1
200	−0.5	47.7	67.3	47.0	32.5	46.2	77.9	45.8	39.1	43.5	84.3	42.8	43.5	43.3	85.7	42.9	42.1	39.0	92.5	40.7	51.5
	0.0	47.1	66.4	46.5	32.6	45.2	77.6	44.4	38.4	43.2	84.5	42.9	43.7	42.7	84.4	42.1	40.9	39.4	92.6	39.3	54.1
	0.5	44.8	63.7	44.2	29.8	42.3	74.3	41.7	35.7	39.8	81.5	39.4	40.5	40.0	82.5	39.7	38.8	38.9	92.5	36.5	59.2
400	−0.5	47.1	67.7	46.5	29.7	45.3	76.9	44.3	40.0	42.8	84.2	42.0	47.8	41.9	85.3	41.1	47.9	37.8	93.0	39.2	62.2
	0.0	47.3	67.9	46.7	29.7	44.4	77.1	43.6	39.0	42.8	84.0	42.1	48.1	41.6	85.3	40.8	47.4	40.4	93.4	40.1	64.8
	0.5	45.8	65.9	44.8	28.7	44.0	76.1	42.7	39.0	41.7	82.9	40.7	47.2	40.2	84.1	39.6	46.4	39.2	92.9	37.2	66.2

Table 6: Empirical size, size adjusted power and empirical rejection frequencies of standard (ADF), dummy-based (ADF^D) and robust QML (ADF^Q) ADF tests. Gaussian errors, trended data.

Size		Model S_0				Model S_4					
T	γ	ADF_α	ADF_t	ADF_α^Q	ADF_t^Q	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α^Q	ADF_t^Q
100	-0.5	4.5	4.5	4.5	4.6	6.2	6.4	1.6	4.3	3.8	6.8
	0	4.4	4.4	4.4	4.6	6.6	6.6	1.9	4.5	4.2	7.2
	0.5	4.6	4.4	4.6	4.5	6.6	6.4	2.1	4.3	4.3	7.0
200	-0.5	5.5	5.6	5.4	5.7	7.8	8.0	2.1	5.3	2.5	6.0
	0	5.0	5.1	5.0	5.3	7.3	7.6	2.0	4.6	2.3	5.3
	0.5	5.4	5.4	5.4	5.5	7.4	7.6	2.4	5.1	2.7	5.7
400	-0.5	4.9	5.2	4.9	5.3	7.1	7.5	2.0	4.6	2.1	4.8
	0	5.0	5.1	4.9	5.2	6.9	7.2	1.8	4.5	1.9	4.7
	0.5	4.9	5.0	4.9	5.2	6.9	7.2	1.8	4.6	1.9	4.7
Power		Model S_0				Model S_4					
T	γ	ADF_α	ADF_t	ADF_α^Q	ADF_t^Q	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α^Q	ADF_t^Q
100	-0.5	49.6	50.5	49.6	50.7	49.9	49.9	69.2	72.2	55.4	60.9
	0	47.4	47.9	47.4	46.9	44.8	44.6	64.1	68.0	51.1	56.8
	0.5	38.2	38.9	38.1	38.8	37.2	37.3	50.3	58.6	41.1	47.3
200	-0.5	49.7	50.4	49.9	50.3	48.6	48.4	69.3	72.3	67.3	70.8
	0	51.2	51.2	51.1	51.2	48.8	49.4	70.0	73.3	67.7	71.8
	0.5	43.4	43.7	43.5	44.0	43.2	43.1	59.8	65.2	58.4	63.6
400	-0.5	50.7	50.7	50.9	50.7	49.8	49.8	71.0	75.0	70.5	74.9
	0	49.9	50.3	49.8	50.3	50.0	49.4	72.7	75.3	72.4	74.5
	0.5	47.3	48.1	47.1	47.4	46.6	46.6	67.0	70.9	66.5	70.8
Empirical power		Model S_0				Model S_4					
T	γ	ADF_α	ADF_t	ADF_α^Q	ADF_t^Q	ADF_α	ADF_t	ADF_α^D	ADF_t^D	ADF_α^Q	ADF_t^Q
100	-0.5	47.0	48.0	47.0	48.6	55.6	56.4	41.7	68.6	47.8	68.9
	0	44.9	45.2	44.8	45.6	53.0	52.9	39.0	65.4	45.1	65.2
	0.5	36.2	36.0	36.2	36.5	44.5	43.9	30.9	54.7	37.4	55.3
200	-0.5	52.1	53.6	52.0	53.9	60.8	62.1	48.0	73.9	48.8	74.9
	0	51.2	52.2	51.1	52.4	59.2	60.2	45.8	71.9	47.1	73.3
	0.5	45.5	46.3	45.6	46.7	54.0	54.8	40.9	65.7	42.1	67.0
400	-0.5	49.9	51.6	49.9	51.8	58.8	60.4	45.9	73.3	46.2	74.0
	0	49.7	51.1	49.5	51.4	58.1	59.7	45.9	73.0	46.0	73.2
	0.5	46.7	48.3	46.6	48.6	55.6	56.9	41.6	69.0	42.0	69.7