Journal article

ON PARTIAL-SUM PROCESSES OF ARMAX RESIDUALS

By Steffen Grønneberg* and Benjamin Holcblat†

*BI Norwegian Business School and †Universite du Luxembourg

This is an Accepted Manuscript of a journal article published in *Annals of statistics,* 2018.
https://www.imstat.org/journals-and-publications/annals-of-statistics/

Submitted to the Annals of Statistics

ON PARTIAL-SUM PROCESSES OF ARMAX RESIDUALS

By Steffen Grønneberg[*], Benjamin Holcblat[†]

BI Norwegian Business School[*] and Université du Luxembourg, LSF[†]

We establish general and versatile results regarding the limit behavior of the partial-sum process of ARMAX residuals. Illustrations include ARMA with seasonal dummies, misspecified ARMAX models with autocorrelated errors, nonlinear ARMAX models, ARMA with a structural break, a wide range of ARMAX models with infinite-variance errors, weak GARCH models and the consistency of kernel estimation of the density of ARMAX errors. Our results identify the limit distributions, and provide a general algorithm to obtain pivot statistics for CUSUM tests.

1. Introduction. Autoregressive moving-average models with covariates (ARMAX) is one of the most common model classes for at least three reasons. Firstly, it nests and combines the widely-used linear regression model and ARMA models, the backbone of traditional time series analysis [e.g., 13]. Secondly, VARMAX models, which have met a renewed interest with the emergence of "big data" through factor models [e.g., 18], can be written as a system of ARMAX models [e.g., 9, sec. 7.2.2]. Thirdly, a large class of nonlinear models [e.g., 23, sec. 2] and state-space models [28, Theorem 1.2.1] have an ARMAX representation from which they can be studied.

In many situations, estimation and inference in ARMAX models require the use of residuals instead of the error terms, as the latter are *unobservable*. The present paper provides weak assumptions that relate the partial-sum processes of error terms and the partial-sum process of residuals, i.e., $\mathfrak{U}_T(s) := \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} u_t$ and $\hat{\mathfrak{U}}_T(s) := \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} \hat{u}_{T,t}$ for $s \in [0,1]$, and where (u_t) and $(\hat{u}_t)$ are respectively error terms and residuals originating from fitting a univariate ARMAX model to a time-series of length T. Our main application is the identification of the limit behavior of CUSUM tests for structural breaks, i.e., statistical functionals, such as the supremum of the absolute value of a function, applied to $\hat{\sigma}_u^{-1}\sqrt{T}\hat{\mathfrak{U}}_T$ or transformations thereof, where $\hat{\sigma}_u$ is a consistent estimator of the residuals' standard deviation. At least

MSC 2010 subject classifications: Primary 62M10, 60F17, 62J20; Secondary 91B84.

Keywords and phrases: ARMAX residuals, Change-point problems, CUSUM test, Misspecified ARMAX, Autocorrelated errors, Infinite-variance errors, Heteroscedastic errors, Linear and nonlinear regression, Seasonal dummy, Kernel density estimation of ARMAX errors, Nonlinear ARMAX, Weak GARCH.

since Brown, Durbin and Evans [14], CUSUM tests have become standard diagnostic tools in different areas, such as medical statistics, econometrics and signal processing.

The core results of the present paper are of the form $\sup_{s\in[0,1]} |\sqrt{T}[\hat{\mathfrak{U}}_T(s) - \mathfrak{U}_T(s)] + \zeta_T(s)| = o_P(1)$, where ζ_T corresponds to the asymptotic gap between the scaled partial-sum process of the residuals $\sqrt{T}\hat{\mathfrak{U}}_T$ and the scaled partial-sum process of the errors $\sqrt{T}\mathfrak{U}_T$. When $\zeta_T = 0$, the CUSUM tests based on $\sqrt{T}\hat{\mathfrak{U}}_T$ have the same critical values as when observing the error terms directly. This is shown to hold in Bai [2] in the case of an ARMA with a known zero-mean parameter, but it will not hold most of the time: Even the inclusion of a mean parameter in the ARMA model is sufficient to affect the behavior of $\sqrt{T}\hat{\mathfrak{U}}_T$, i.e., to make ζ_T non-zero. While known for some time in the basic linear regression model, this was noticed by Lee [36] in the AR case, and it was later generalized to the ARMA case in Yu [55] and in Ghoudi and Rémillard [26]. A simple illustration of this effect is the linear regression model with only a constant as regressor, i.e., $Y_t = \mu + u_t$. Then, the OLS estimator is the average, i.e., $\hat{\mu} = \frac{1}{T}\sum_{t=1}^T Y_t$, so that $\hat{u}_t - u_t = (Y_t - \hat{\mu}) - u_t = \mu - \hat{\mu}$, which, in turn, implies that $\zeta_T(1) = \sqrt{T}[\hat{\mathfrak{U}}_T(1) - \mathfrak{U}_T(1)] = \sqrt{T}(\mu - \hat{\mu})$. By the central limit theorem, this means that for $s = 1$, and thus a fortiori for the supremum over s, the asymptotic gap ζ_T is a random element that does not go to zero asymptotically.

As the above illustration suggests, the identification of ζ_T, and the related question of taking ζ_T into account when suggesting statistical tests, are practically important: Ignoring ζ_T will result in erroneous critical values for CUSUM tests and most of the other inference procedures based on residuals. Our paper develops a flexible calculus for identifying ζ_T in a large class of cases encountered in practice, from simple cases such as ARMA with seasonal dummy variables, to more complicated models. Once ζ_T is calculated, attention is given to identifying a transformation $\tilde{\Delta}$ of $\hat{\sigma}_u^{-1}\sqrt{T}\hat{\mathfrak{U}}_T$ which is s.t. (such that) $\sup_{s\in[0,1]} |\tilde{\Delta}[\zeta_T](s)| = o_P(1)$, yielding $\sup_{s\in[0,1]} |\sqrt{T}[\tilde{\Delta}[\hat{\mathfrak{U}}_T](s) - \tilde{\Delta}[\mathfrak{U}_T](s)]| = o_P(1)$. Under weak conditions we therefore have process convergence $\hat{\sigma}_u^{-1}\sqrt{T}\tilde{\Delta}[\hat{\mathfrak{U}}_T](s) \xrightarrow[T\to\infty]{\mathscr{L}} \tilde{\Delta}[B](s)$ where B is a Brownian motion, i.e., a pivot process, hence enabling the statistician to apply CUSUM tests for structural stability. We define the pivot transformation $\tilde{\Delta}$, which corresponds to a bounded linear operator, through an algorithm. A special case of the transformation $\tilde{\Delta}$ yields the scaled partial-sum process $\hat{\mathfrak{V}}_T(s) = \frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor} \hat{v}_{T,t}$ where $\hat{v}_{T,t} := \hat{u}_t - \frac{1}{T}\sum_{j=1}^T \hat{u}_j$, not of residuals but of average-corrected residuals. This transformation and its pivot properties were identified by Yu [55] in the case of an ARMA, and it suffices when $\sup_{s\in[0,1]}|\zeta_T(s)-s\zeta_T(1)|=o(1)$, which is not the case in more general ARMAX models.

In the present paper, we work with ARMAX models in a wide sense. In particular, the error terms need not be weak white noise, i.e., zero-mean and constant finite variance with zero autocorrelation. In this way, our results have direct implications for a large class of models, such as ARMA-GARCH models where the error term is not IID (independent and identically distributed), as well as nonlinear models and state space models with ARMAX representation, where errors terms are often *not* IID *nor* martingale-difference processes, [e.g., 23, sec. 2]. Following Bai [2, 4], Yu [55], and Ghoudi and Rémillard [26], we typically make no assumption on the estimators of the ARMAX parameters other than they are $O_P(T^{-1/2})$ away from their targets. Hence, almost all of our results hold irrespectively of the estimation method chosen. In addition, motivated by practical consideration and the latest development of estimation theory, our assumptions allow for heteroskedasticity [e.g., 9, chap. 8, Assumption 8.1.1.], autocorrelation in the errors [e.g., 33] and higher-order dependence [e.g., 23], feedback effect between the covariates and the dependent variable [e.g., 9, chap. 8, p. 155], nonlinear components in the covariates, infinite-variance errors [e.g., 39], seasonal dummies, and several wide classes of covariates (e.g., integrable stationary and ergodic [e.g., 10, p. 494], or fractional ARMA [13, Definition 13.2.2]). Finally, we are also able to analyze cases where the model is misspecified, and where we are estimating least-false parameters. Our reliance on elementary but general inequalities in the crux of our proofs rather than on probabilistic sophistication explains the generality and versatility of our results.

1.1. *Technical setup.* We consider a univariate ARMAX processes $(Y_{T,t})$ s.t.

$$\Phi(B)(Y_{T,t}-\mu)=\boldsymbol{\lambda}'\boldsymbol{X}_{T,t-1}+\Theta(B)u_t, \tag{1.1}$$

where B denotes the lag operator, $\Phi(z) := 1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p$, $\Theta(z) := 1+\theta_1 z+\theta_2 z^2+\cdots+\theta_q z^q$, $\boldsymbol{\lambda} := (\lambda_1,\lambda_2,\ldots,\lambda_{d_{\boldsymbol{\lambda}}})'\in\mathbf{R}^{d_{\boldsymbol{\lambda}}}$ and $\boldsymbol{X}_{T,t-1} := (X_{T,t-1,1},\ldots,X_{T,t-1,d_{\boldsymbol{\lambda}}})'$ is a triangular vector array of covariates. We assume that, based on observations $(Y_{T,t},\boldsymbol{X}_{T,t-1})_{t=1}^T$, there exist $O_P(T^{-\frac{1}{2}})$-consistent estimators $\hat\mu$, $\hat{\boldsymbol{\lambda}}$, $(\hat\phi_i)_{i=0}^p$ and $(\hat\theta_j)_{j=1}^q$ of μ, $\boldsymbol{\lambda}$, $(\phi_i)_{i=0}^p$ and $(\theta_j)_{j=1}^q$, respectively. When (u_t) has finite and constant variance, we denote it by σ_u^2.

Notice that since the covariates in $\boldsymbol{X}_{T,t-1}$ are allowed to depend on T, this dependence is transferred to $Y_{T,t}$, which therefore also depends on T.

This allows us to study covariates such as the dummy variable $I\{t \leqslant T\mathfrak{p}\}$, i.e., a change point at the first $\mathfrak{p}$-th fraction of the sample. This dependence is transferred to $Y_{T,t}$, but we do not assume that u_t depends on the sample-size for mathematical convenience, although it should be possible to extend our results in that direction.

The ARMAX residuals $(\hat{u}_t)_{t=-q+1}^{T}$ are then defined as follows. For $t \in [\![1, T]\!]$,

$$\hat{u}_t := (Y_{T,t} - \hat{\mu}) - \sum_{i=1}^{p} \hat{\phi}_i (Y_{T,t-i} - \hat{\mu}) - \sum_{j=1}^{q} \hat{\theta}_j \hat{u}_{t-j} - \hat{\boldsymbol{\lambda}}' \boldsymbol{X}_{T,t-1}, \text{ and} \tag{1.2}$$

for $t \in \mathbf{Z}_-$, $\hat{u}_t = 0$,[1] where, for all $(a, b) \in \mathbf{R}^2$, $[\![a, b]\!] := [a, b] \cap \mathbf{Z}$, and $\mathbf{Z}_- :=]\!] -\infty, 0]\!]$. We also use the average-corrected error and the average-corrected residuals

$$v_{t,T} := u_t - \frac{1}{T}\sum_{j=1}^{T} u_j, \qquad \text{and} \qquad \hat{v}_{t,T} \;:=\; \hat{u}_t - \frac{1}{T}\sum_{j=1}^{T} \hat{u}_j. \tag{1.3}$$

Our main focus is on the following partial-sum processes

$$\hat{\mathfrak{U}}_T(s) := \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} \hat{u}_t, \qquad \mathfrak{U}_T(s) := \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} u_t,$$

$$\hat{\mathfrak{V}}_T(s) := \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} \hat{v}_{t,T}, \qquad \text{and} \qquad \mathfrak{V}_T(s) := \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} v_{t,T}, \tag{1.4}$$

where for all $a \in \mathbf{R}$, $\lfloor a \rfloor := \max\{n \in \mathbf{Z} : n \leqslant a\}$. The present paper establishes general limit theorems for partial-sums process of ARMAX residuals and transformations thereof.

1.2. *Related literature.* While, to the best of our knowledge, the present paper is the first to study partial-sum processes of residuals from full-blown ARMAX models, it is related to many existing papers in addition to the already cited papers. In particular, the present paper complements Andreou and Werker [1], Ghoudi and Rémillard [26], and an extensive literature on partial-sum processes and empirical processes of residuals of regression models [e.g., 50, sec. 4.6, and references therein]. Andreou and Werker [1] rely

[1] Following the existing literature [e.g., 2], residuals with negative indexes are put to zero because they cannot be deduced from the observable data (i.e., the Y_ts and X_ts with positive indexes) as in equation (1.2). However, this does not imply that the Y_ts and X_ts with negative indexes are assumed to be zero.

on Le Cam's theory to develop an elegant and general framework to analyze residual-based statistics. However, their high-level assumptions cannot always be checked (many statistical models cannot be expressed in terms of likelihood as the ULAN assumption requires) and satisfied [e.g., 39, 49, for counter examples to the assumption on asymptotic normality]. Ghoudi and Rémillard [26] derive the asymptotic limit of empirical processes of ARMA residuals, which is a more general object than the partial-sum process of residuals. However, Ghoudi and Rémillard [26] do not allow for covariates, and they require IID square-integrable errors with a probability distribution absolutely continuous w.r.t.(with respect to) the Lebesgue measure. Another advantage of the approach developed in the present paper w.r.t. [1] and [26] is that the crux of our proofs is based on elementary inequalities, thereby gaining a high degree of generality and versatility.

The literature on partial-sum processes and empirical processes of residuals from regression models include results for autoregressive processes [e.g., 12, 34, 7] with long-memory errors [e.g., 15] and time trends [e.g., 54, 40, 7]. Our main contribution with respect to this literature is to allow for a MA (moving average) component in the process in eq. (1.1). Inspection of the proofs shows that tackling the MA component is one the main technical challenges of the present paper. Unlike this literature [e.g., 34, 20, 53, 54], the present paper does not tackle unit roots and polynomial time trend. However, as a follow-up paper shows, the framework of the present paper can readily be extended to tackle these cases. The present paper is also indirectly related to papers that derives consistency results for some specific functions of ARMAX residuals [e.g. 51, 22], and to the literature on CUSUM tests. Following MacNeill [38], Ploberger and Krämer [44, 45] and others, but unlike a part of the literature on CUSUM tests [e.g., 14, 35], our CUSUM tests statistics are not based on recursive residuals, but on the standard residuals from the whole sample. From a practical point of view, the rationale for using standard residuals instead of recursive residuals is that the former are readily computed, while the latter requires repetitive computations, which can become numerically unstable, especially given the nonlinear objective functions of full-blown ARMA and full-blown ARMAX models. From a theoretical point of view, neither of the two types of residuals has been shown to yield uniformly superior tests.

1.3. *Organization of the paper.* Besides this introductory section, our paper has two sections: Core results, found in Section 2, which encompasses most standard cases of interest, and extensions, found in Section 3. Section 3 also introduce a general algorithm to obtain pivot statistics for CUSUM tests. All asymptotic statements of the present paper are understood as $T \to \infty$, so the latter qualification is omitted from the main text.

All proofs, and a list of abbreviations with their meaning are found in the supplementary material [27], which consists of several appendices. Page numbers in the supplementary material are prefixed by "S" so that "S1" is the first page in [27]. Because our assumptions are very weak, most proofs consist of long calculations to reach the point where low-level techniques, such as, say, the use of sub-additivity of probability measures, can be applied. We therefore build up a library of lemmas to simplify these calculations, and we also provide very detailed proofs. The high level of detail in the proofs is motivated firstly by the desire for transparency, and secondly in order to make our techniques easy to apply in further research.

2. Core results and immediate applications.

2.1. *Core assumptions and expansion of the ARMA part.* As mentioned in the introduction, we consider ARMAX processes in a wide sense, i.e., we consider a process (Y_t) to be an ARMAX processes if it is a solution to an equation of the form of eq. (1.1) on p. 3. Thus, our assumptions, which are extensions or weakening of the assumptions in the related papers [2, 4, 55, 26], allow us to consider processes that are outside the traditional ARMAX framework where error terms are usually IID or martingale differences. Assumption 1, which is standard, requires the roots of $\Phi(.)$ and $\Theta(.)$ to be outside the unit circle.

ASSUMPTION 1 (Invertibility of lag polynomial). *Let $\Phi(.)$ and $\Theta(.)$ be the AR and MA polynomials of the ARMAX process (1.1) on p. 3.* **(a)** *All roots of $\Phi(z)$ lie outside the unit circle of the complex plane.* **(b)** *All roots of $\Theta(z)$ lie outside the unit circle of the complex plane.*

For ARMA processes with IID errors, Assumption 1(a) and (b) respectively correspond to causality and invertibility [e.g., 13, pp. 83–89]. Assumption 1(a) allows us to solve eq. (1.1) for $Y_{T,t}$. Assumption 1(a) is not a binding assumption, as we can always incorporate the autoregressive part with roots inside the unit circle of the complex plane among the covariates' part. In a follow-up paper, we tackle unit roots in this way. In contrast to Assumption 1(a), Assumption 1(b) is crucial and binding. Assumption 1(b) ensures that errors (or alternatively residuals) can be expressed in terms of observables (see Lemma 12 in Appendix B.1, p. S11 of [27]). Under Assumption 1 (b), we note that the process $\check{Y}_t := \Phi(B)^{-1}\Theta(B)u_t$ is a so-called weak

ARMA if (u_t) is weak white noise. We will sometimes work with $(\check{Y}_t)$ in the proofs, which, in contrast to the ARMAX process (Y_t), is unobservable.

The following Assumption 2 is mild: It requires the difference between the estimators of the ARMA parameters and the population ARMA parameters to be $O_P(T^{-\frac{1}{2}})$.

Assumption 2 ($O_P(T^{-\frac{1}{2}})$-consistency of ARMA parameters). *Define the stacked parameter* $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_p)'$ *and* $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_q)'$. *Let* $\hat{\mu}$, $\hat{\boldsymbol{\phi}}$, *and* $\hat{\boldsymbol{\theta}}$ *be the respective estimators of* μ, $\boldsymbol{\phi}$, *and* $\boldsymbol{\theta}$ *s.t.* **(a)** $\sqrt{T}(\hat{\mu} - \mu) = O_P(1)$; **(b)** $\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) = O_P(1)$; *and* **(c)** $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_P(1)$.

Assumption 2 allows us to determine which terms survive asymptotically once we multiply the difference between the partial-sum process of the residuals and the partial-sum process of the errors by $\sqrt{T}$. Assumption 2 is weaker than $\sqrt{T}$-asymptotic normality, which has been proved for ARMAX [e.g., 29, 9]. Assumption 2 allows for faster rates of convergence such as the one that has been established for ARMA with infinite-variance errors [e.g., 39, p. 310, Theorem 2.2]: See Section 3.1 on p. 17. Note that Assumption 2 rules out any identification problem, because identifiability is a necessary condition for the consistency of an estimator.

Remark 1. For ARMAX models with unknown lag order, Hannan and Deistler [28, Chap. 5] give conditions under which certain model-selection criteria are consistent. As noticed by Hannan and Quinn [30, p 191] and further discussed in Pötscher [46, see especially his Lemma 1] with the caveat pointed out in [37], this means that all asymptotic results based on the assumption that the true model is known also hold when using consistent model-selection procedures. This observation, of course, also applies to our results. ◇

The following Assumption 3 is also mild, as we explain below.

Assumption 3 (Error term u_t). **(a)** *For a constant* $\epsilon_u > 0$, $\sup_{t \in \mathbf{Z}} \mathbb{E}|u_t|^{1+\epsilon_u} < \infty$. **(b)** $\sup_{s \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} u_t \right| = o_P(1)$.

Assumption 3(a) ensures the existence of certain power series in B applied to the error u_t (Lemma 7(i) on p. S4 of [27]). In the present paper, except when indicated otherwise, we understand power series in B applied to a process in terms of almost sure convergence: As we do not always require the existence of the second moment, using the standard convergence in L^2 (i.e., the space of square-integrable random variables) is not possible. Assumption 3(a) also allows us to apply the Phillips-Solo device [42]. The latter is a technique based on the Beveridge-Nelson decomposition [8] that allows us to asymptotically reduce the study of partial-sums of linear filters of a process to simply the partial-sums of the process, i.e., it allows us to factor out linear filters. See Lemma 9 on p. S6 of [27] for a precise statement of the versions of the Phillips-Solo device used in the present paper. By the Phillips-Solo device [42], Assumption 3(b) ensures that the partial-sum *average* process of power series of the error vanishes asymptotically (Corollary 3 on p. S8 of [27]). Assumption 3(b) is weaker than the standard assumptions that $(u_t)_{t\in\mathbf{Z}}$ is a square-integrable zero-mean IID process, or that it is at least a L^p-bounded martingale difference with $p > 2$ [e.g., 2, 55]. Appendix B.6 (p. S19) of [27] provides a catalogue of sufficient conditions. In particular, Assumption 3(b) allows for (unconditional and conditional) heteroscedasticity and for autocorrelation or higher forms of time-dependence. Heteroscedasticity and time-dependence in the errors are likely to occur. Many financial and economic time series appear unconditionally heteroscedastic [e.g., 48, and references therein]. Similarly, autocorrelation of the errors is often difficult to rule out [e.g., 33, and references therein], and ARMAX representations of nonlinear models often yield errors that are neither IID nor martingales differences [e.g., 23]. Thus, as further illustrated below in some examples, the generality of Assumption 3(b) is useful.

Lemma 1(i) shows that the "ARMA part" of the partial-sum processes defined in eq. (1.4) can be characterized without assumptions on the covariates. It is therefore the core lemma we use in all upcoming results, under various assumptions on the covariates.

LEMMA 1 (Fundamental Lemma: Expansion for ARMA part). *Under Assumptions 1, 2 and 3, w.p.a.1 as $T \to \infty$,*

$$\text{(i)} \quad \sup_{s\in[0,1]} \left| \sqrt{T}[\hat{\mathfrak{U}}_T(s) - \mathfrak{U}_T(s)] + s\frac{\Phi(1)}{\Theta(1)}\sqrt{T}(\hat{\mu} - \mu) - \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \Xi_{t,T} \right| = o_P(1);$$

$$\text{(ii)} \quad \sup_{s\in[0,1]} \left| \sqrt{T}[\hat{\mathfrak{V}}_T(s) - \mathfrak{V}_T(s)] - \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} [\Xi_{t,T} - \bar{\Xi}_T] \right| = o_P(1),$$

where, denoting the lag coefficients in the inverse of the polynomial $1 +$

$\sum_{j=1}^{q} \hat{\theta}_j z^j$ *with* $(\psi_j(\hat{\boldsymbol{\theta}}))_{j=0}^{\infty}$ *(i.e.,* $\left[\sum_{j=0}^{\infty} \psi_j(\hat{\boldsymbol{\theta}}) z^j\right] \left[1 + \sum_{j=1}^{q} \hat{\theta}_j z^j\right] = 1$*),*

(2.1)

$$\Xi_{t,T} := -\sum_{j=0}^{t-1} \psi_j(\hat{\boldsymbol{\theta}}) \left\{ (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})' \boldsymbol{X}_{T,t-1-j} + \sum_{i=1}^{p} (\hat{\phi}_i - \phi_i) \Phi(B)^{-1} \boldsymbol{\lambda}' \boldsymbol{X}_{T,t-1-i-j} \right\}.$$

PROOF. See Appendix C.2 on p. S28 of [27]. □

The inverse of the polynomial $1+\sum_{j=1}^{q} \hat{\theta}_j z^j$ (i.e., the power series $\sum_{j=0}^{\infty} \psi_j(\hat{\boldsymbol{\theta}}) z^j$) exists w.p.a.1 (with probability approaching one) as $T \to \infty$, so that $\Xi_{T,t}$ is well-defined (Proposition 9 on p. S12 in Appendix B.4 of [27]). Hereafter, we drop the qualification "w.p.a.1 as $T \to \infty$" because all upcoming results rest on Lemma 1. If there is no MA part in the model (i.e., $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} = 0$), then $\psi_0(\hat{\boldsymbol{\theta}}) = 1$ and $\psi_j(\hat{\boldsymbol{\theta}}) = 0$ for $j \in [\![1, \infty[\![$, and $\Xi_{t,T}$ simplifies to $\Xi_{t,T} = -(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})' \boldsymbol{X}_{T,t-1} - \sum_{i=1}^{p} (\hat{\phi}_i - \phi_i) \Phi(B)^{-1} \boldsymbol{\lambda}' \boldsymbol{X}_{T,t-1-i}$. If there is no covariate (i.e., $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}} = 0$ and $\boldsymbol{X}_{t-1} = 0$ for all $t \in \mathbf{Z}$), then, for all $(t,T) \in \mathbf{N}^2$, $\Xi_{t,T} = 0$, and thus Lemma 1(i) and Lemma 1(ii) respectively implies Theorem 1 and Corollary 1 in [55] for $k = 1$ because $\sup_{s \in [0,1]} |s\Phi(1)/\Theta(1) - \lfloor Ts \rfloor (1 - \sum_{i=1}^{p} \phi_i)/[T(1 + \sum_{j=1}^{q} \theta_j)]\,| = o(1)$. If, for all $(t,T) \in \mathbf{N}^2$, $\Xi_{t,T} = 0$ and $\hat{\mu} = \mu = 0$ (i.e., no covariates and no intercept), Lemma 1(i) implies the main result in Bai [2, Theorem 1]. We therefore generalize the core results of these papers by showing that they also hold under weaker conditions on the error term. See the discussion of Assumption 3 and the discussion that follows it. This generalization is practically relevant, as illustrated in the following examples.

EXAMPLE 1. To illustrate that heteroscedasticity and autocorrelation in the error terms can easily arise within our assumptions, let us revisit the simple model $Y_t = \mu + u_t$ from the introduction. Assume (u_t) is a linear process $u_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ where (ε_t) is a zero-mean process with uncorrelated elements, but where $\operatorname{Var} \varepsilon_t$ may depend on t. Moreover, assume $\sup_{t \in \mathbf{Z}} \mathbb{E}\varepsilon_t^2 < \infty$ and that there exist $M > 0$ and $\rho \in]0,1[$ s.t., for all $j \in [\![0, \infty[\![$, we have $|\alpha_j| < M\rho^j$. Then, for the representation $Y_t = \mu + u_t$ estimated with the OLS estimator $\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} Y_t$, Assumptions 1-3 hold, as shown in Appendix C.3 on p. S33 of [27]. ◇

EXAMPLE 2. An important class of models are ARMA models where (u_t) has time varying volatility. A prominent example is ARMA-GARCH models, where the error term follows a GARCH model. By definition, integrable GARCH processes are martingale difference sequences [e.g., 25, Definition

2.1.(i)]. Now, in Proposition 11 of [27] (p. S23), we show that martingale differences satisfy Assumption 3(b) under the standard assumption that they are uniformly L_r-bounded with $r > 1$. Thus, Lemma 1 covers most ARMA-GARCH models. In this way, Lemma 1 greatly generalizes Bai [2, Theorem 1], which applies to ARMA-GARCH models only when the intercept is assumed to be zero and hence not estimated – not even indirectly through first subtracting the average of the observations, and which requires (u_t) to be at least uniformly L_r-bounded martingale differences with $r > 2$ [2, assumption a.1']. The upcoming Theorem 1 further generalizes this result to a large class of ARMAX-GARCH models, which, in particular, nests the ARCH *regression* model of Engle [19, sec. 5]. $\diamond$

EXAMPLE 3. A weak GARCH process is any process (ε_t) s.t. (ε_t^2) is an ARMA process of the form $\Phi(B)(Y_t - \mu) = \Theta(B)u_t$ where (u_t) is at least assumed to be weak white noise (see Francq and Zakoïan [24, Section 2] where additional technical assumptions are made). Weak GARCH processes generalize the class of standard GARCH models and span several other interesting volatility models, including Markov-switching GARCH processes, stochastic volatility models and aggregated GARCH processes [25, Section 4.2]. For several of these representation results to hold, it is essential that (u_t) is not restricted to be IID or even martingale difference sequences. Previously known theory on partial-sum processes of residuals therefore does not apply in this setting. Estimation theory for weak GARCH processes is developed in Francq and Zakoïan [24], providing $\sqrt{T}$-consistent estimators for the parameters. The analysis of partial-sum processes of residuals and average-corrected residuals from weak GARCH models is a consequence of Lemma 1, as long as the ARMA representation of (ε_t^2) fulfils Assumptions 1, 2 and 3. Assumptions 1 and 2 follow from [24]. Because (u_t) is assumed to be weak white noise, Assumption 3 also holds (Lemma 18 (a), Appendix B.6 on p. S19 of [27]). $\diamond$

The upcoming Section 2.2.2 shows that, under weak assumptions, we have $\sup_{s\in[0,1]} |\hat{\sigma}_u^{-1}\hat{\mathfrak{V}}_T(s)| \xrightarrow[T\to\infty]{\mathscr{L}} \sup_{s\in[0,1]} |B^\circ(s)|$ where B° is a Brownian bridge process and $\hat{\sigma}_u$ is the empirical standard deviation of the residuals. For weak GARCH processes treated in Example 3, this appears to induce new tests for structural stability.

2.2. *Expansions for generic covariates.* We here study the contribution of covariates to expansions of the partial-sum process in the most common cases encountered in practice, namely settings where $\hat{\boldsymbol{\lambda}}$ is $\sqrt{T}$-consistent.

2.2.1. *Assumption and theorem.* We now analyze the process $s \mapsto \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \Xi_{t,T}$. Lemma 1 (p. 8) shows that this process is central for understanding ARMAX residuals. We make the following assumptions.

ASSUMPTION 4 (Covariates $\boldsymbol{X}_{T,t}$). **(a)** $\sqrt{T}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) = O_P(1)$. **(b)** *For a constant $\epsilon_X > 0$, for all $l \in [\![1, d_{\boldsymbol{\lambda}}]\!]$, $\sup_{(T,t)\in\mathbf{N}\times\mathbf{Z}} \mathbb{E}|X_{T,t,l}|^{1+\epsilon_X} < \infty$.* **(c)** $\sup_{s\in[0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} (\boldsymbol{X}_{T,t-1} - \mathbb{E}\boldsymbol{X}_{T,t-1}) \right| = o_P(1)$. **(d)** *For all $i \in [\![1, p]\!]$,* $\sup_{s\in[0,1]} \left| \left[\frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} \mathbb{E}\boldsymbol{X}_{T,t-1-i} \right] - s\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\boldsymbol{X}_{T,t-1-i} \right| = o(1)$.

Assumption 4 (a) requires $O_P(T^{-\frac{1}{2}})$-consistency. As noted for the ARMA parameters, this allows for faster rate of convergence. Assumption 4 (b) ensures the a.s. (almost sure) finiteness of the infinite series $\Phi(B)^{-1}\boldsymbol{X}_t$, $\Theta(B)^{-1}\boldsymbol{X}_{T,t-1}$, and $\Theta(B)^{-1}\Phi(B)^{-1}\boldsymbol{X}_{T,t-1}$, and their expectation by an extended Minkowski inequality (Lemma 7 on p. S4 in [27]). It also allows the application of the Phillips-Solo device [42] on partial-sums of $\Theta(B)^{-1}\boldsymbol{X}_{T,t-1}$ and $\Theta(B)^{-1}\Phi(B)^{-1}\boldsymbol{X}_{T,t-1}$ (Lemma 9 on p. S6 in [27]). Assumption 4 (b) can be weakened into $\sup_{(T,t)\in\mathbf{N}\times\mathbf{Z}} \mathbb{E}|X_{T,t,l}| < \infty$, but Assumptions 4 (c) and (b) would then need to be modified and extended.

Assumptions 4 (a)-(c) reduce the study of the process $s \mapsto \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \Xi_{t,T}$ into the study of the deterministic processes

$$
\begin{aligned}
L_{1,T}(s) &:= \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} \Theta(B)^{-1}\mathbb{E}\boldsymbol{X}_{T,t-1}, \text{ and} \\
L_{2,i,T}(s) &:= \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} \Theta(B)^{-1}\Phi(B)^{-1}\mathbb{E}\boldsymbol{X}_{T,t-1-i},
\end{aligned}
\tag{2.2}
$$

where $i \in [\![1, p]\!]$. Taken together Assumptions 4 (c) and (d) are the counterparts of Assumption 3 (b) for the covariates: Assumption 4 (d) places restrictions on the term subtracted in Assumption 4 (c). Appendices B.6 (p. S19) and C.5 (p. S38) in [27] show that Assumption 4 (c) and (d) hold for most of the processes considered in the time-series literature. By the Phillips-Solo device [42], Assumptions 4 (c) and (d) allow a further reduction of the processes in eq. (2.2).

THEOREM 1 (Expansion for generic covariates). *Under Assumptions 1, 2, 3, and 4(a)-(c),*

(i) $\sup_{s\in[0,1]} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \Xi_{t,T} - \sqrt{T}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})' L_{1,T}(s) - \sum_{i=1}^{p} \sqrt{T}(\hat{\phi}_i - \phi_i)\boldsymbol{\lambda}' L_{2,i,T}(s) \right| =$

$o_P(1)$ and hence $\sup_{s\in[0,1]}\left|\sqrt{T}[\hat{\mathfrak{U}}_T(s)-\mathfrak{U}_T(s)]+s\frac{\Phi(1)}{\Theta(1)}\sqrt{T}(\hat{\mu}-\mu)-\sqrt{T}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})'L_{1,T}(s)-\sum_{i=1}^{p}\sqrt{T}(\hat{\phi}_i-\phi_i)\boldsymbol{\lambda}'L_{2,i,T}(s)\right|=o_P(1)$*;*

(ii) Under the additional Assumption 4(d),

(ii.a) $\sup_{s\in[0,1]}\left|L_{1,T}(s)-s\Theta(1)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\boldsymbol{X}_{T,t-1}\right|=o(1)$ *and, for all* $i\in[\![1,p]\!]$, $\sup_{s\in[0,1]}\left|L_{2,i,T}(s)-s\Theta(1)^{-1}\Phi(1)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\boldsymbol{X}_{T,t-1-i}\right|=o(1)$*; and*

(ii.b) $\sqrt{T}\sup_{s\in[0,1]}\left|\hat{\mathfrak{V}}_T(s)-\mathfrak{V}_T(s)\right|=o_P(1)$.

PROOF. See Appendix C.4 on p. S35 of [27]. □

Theorem 1(i), which does not assume Assumption 4 (d), characterizes the asymptotic difference ζ_T between the scaled residual partial-sum process and the scaled errors partial-sum process. Note that while the expansion does not depend on $\hat{\boldsymbol{\theta}}$, the actual MA terms $\boldsymbol{\theta}$ influence the functions $L_{1,T}$ and $L_{2,i,T}$, see eq. (2.2). Theorem 1(ii) shows that Assumption 4(d) yields a simplification in the analysis of the partial-sum process of average-corrected residuals. More precisely, Theorem 1(ii.a), which relies on the Phillips-Solo device, simplifies the expression of $L_{1,T}(.)$ and $L_{2,i,T}(.)$. Theorem 1(ii.b) provides an easy way to reach a pivot statistic for CUSUM-type tests even when the covariates $(\boldsymbol{X}_{T,t-1})$ have non-zero-mean. In order to actually reach a pivot statistic in standard cases, σ_u needs to be consistently estimated. This is treated in Section 2.2.2. We also generalize Theorem 1 (ii.b) in Section 3.4.

We end this section with an example verifying Theorem 1 (i) in an elementary case where —in contrast to the large classes of models we consider in the upcoming subsections— Assumption 4 (d) does not hold. The example also anticipates various extensions of Theorem 1 given later in the paper, and will be revisited in Sections 3.4 and 3.5.2.

EXAMPLE 4. Consider the model $Y_t=\lambda I\{t\leqslant \mathfrak{p}T\}+u_t$ where $0<\mathfrak{p}<1$ is known and (u_t) is zero-mean IID with finite variance. Note that $\boldsymbol{X}_{T,t}$ depends on T. Note also that Assumption 4 (c) trivially holds since the covariate is deterministic. However, Assumption 4 (d) does not hold. Indeed, $\frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor}I\{t\leqslant\mathfrak{p}T\}=\frac{1}{T}\sum_{t=1}^{\min(\lfloor Ts\rfloor,\lfloor\mathfrak{p}T\rfloor)}1=\frac{1}{T}\min(\lfloor Ts\rfloor,\lfloor\mathfrak{p}T\rfloor)=\min(s,\mathfrak{p})+o(1)$, and $s\frac{1}{T}\sum_{t=1}^{T}I\{t\leqslant\mathfrak{p}T\}=s\mathfrak{p}+o(1)$ w.r.t. the uniform norm, hence their difference does not go to zero.

By definition, we have $\hat{u}_t = Y_t - \hat{\lambda} I\{t \leqslant \mathfrak{p}T\} = (\lambda - \hat{\lambda}) I\{t \leqslant \mathfrak{p}T\} + u_t$, so $\sqrt{T}\hat{\mathfrak{U}}_T(s) = \sqrt{T}(\lambda - \hat{\lambda})\frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} I\{t \leqslant \mathfrak{p}T\} + \sqrt{T}\mathfrak{U}_T(s)$, which agrees with Theorem 1 (i) since $L_{1,T}(s) = \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} I\{t \leqslant \mathfrak{p}T\}$. Following the above calculation, we see that if $\sqrt{T}(\lambda - \hat{\lambda}) = O_P(1)$, then $\sqrt{T}\hat{\mathfrak{U}}_T(s) = \min(s, \mathfrak{p})\sqrt{T}(\lambda - \hat{\lambda}) + \sqrt{T}\mathfrak{U}_T(s) + o_P(1)$ uniformly. We have $\sqrt{T}\hat{\mathfrak{V}}_T(s) = \sqrt{T}\hat{\mathfrak{U}}_T(s) - s\sqrt{T}\hat{\mathfrak{U}}_T(1) + o_P(1) = \sqrt{T}\mathfrak{U}_T(s) - s\sqrt{T}\mathfrak{V}_T(1) + \min(s, \mathfrak{p})\sqrt{T}(\lambda - \hat{\lambda}) - s\min(1, \mathfrak{p})\sqrt{T}(\lambda - \hat{\lambda}) + o_P(1)$ uniformly. Hence, average-correcting residuals does not lead to an asymptotic pivot, i.e., the nuisance term related to $\min(s, \mathfrak{p})\sqrt{T}(\lambda - \hat{\lambda})$ is not removed. In this special case, we could study $\sqrt{T}(\lambda - \hat{\lambda})$ and derive the joint process limit of $\sqrt{T}\hat{\mathfrak{U}}_T$. However, this limit would in general depend on further nuisance parameters, and the joint process limit can be more challenging to derive in more complex settings. As a special case of a general technique described in Section 3.4, we see that $\sqrt{T}\hat{\mathfrak{U}}_T(s) - [\min(s, \mathfrak{p})/\mathfrak{p}]\sqrt{T}\hat{\mathfrak{U}}_T(1) = \min(s, \mathfrak{p})\sqrt{T}(\lambda - \hat{\lambda}) + \sqrt{T}\mathfrak{U}_T(s) - [\min(s, \mathfrak{p})/\min(1, \mathfrak{p})][\min(1, \mathfrak{p})\sqrt{T}(\lambda - \hat{\lambda}) + \sqrt{T}\mathfrak{U}_T(1)] + o_P(1) = \sqrt{T}\mathfrak{U}_T(s) - [\min(s, \mathfrak{p})/\min(1, \mathfrak{p})]\sqrt{T}\mathfrak{U}_T(1) + o_P(1)$. Thus, if $\hat{\sigma}_u \xrightarrow{P} \sigma_u > 0$, the functional central limit theorem and the continuous mapping theorem imply that $\hat{\sigma}_u^{-1}\sqrt{T}\sup_{s\in[0,1]} |\hat{\mathfrak{U}}_T(s) - [\min(s, \mathfrak{p})/\mathfrak{p}]\hat{\mathfrak{U}}_T(1)| \xrightarrow{\mathscr{L}} \sup_{s\in[0,1]} |B(s) - [\min(s, \mathfrak{p})/\min(1, \mathfrak{p})]B(1)|$, where B is a Brownian motion. This CUSUM test, which to our knowledge is new, has critical values easily found via simulation. $\diamond$

2.2.2. *Consistency of empirical residual-based variance, and obtaining pivot statistics.* We here derive asymptotic pivot statistics from partial-sum processes. This requires estimating σ_u, which is the first topic of this section. Because our proofs immediately generalize to the multivariate case, we will here consider a system of ARMAX models, and treat ARMAX residuals as a special case. We are minimalistic in the introduced notation, as we will only work with the multivariate case in the present section.

PROPOSITION 1 (Consistency of empirical variance). *Suppose given a system of d ARMAX models, each fulfilling Assumptions 1, 2, 3, 4 (a). Denote the i'th element of the covariates in the j'th ARMAX model with $X_{t-1,i,j}$, and the error terms of the j'th ARMAX model with $u_{t,j}$. If, for all $j \in [\![1, d]\!]$, $\sup_{t\in\mathbf{Z}} \mathbb{E}|u_{t,j}|^2 < \infty$ and, for all $i \in [\![1, d_{\boldsymbol{\lambda}}]\!]$, $\sup_{t\in\mathbf{Z}} \mathbb{E}X_{t-1,i,j}^2 < \infty$, then*

$$\hat{\Sigma}_{u,T} = \Sigma_{u,T} + o_P(1),$$

where $\Sigma_{u,T} := \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{u}_t\boldsymbol{u}_t' - \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t\right)'$ *denotes the empirical covariance matrix of the error* $\boldsymbol{u}_t := (u_{1,t}, \ldots, u_{1,t})'$ *and* $\hat{\Sigma}_{u,T} :=$

$\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t' - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\right)'$ *the empirical covariance matrix of the residuals* $\hat{\boldsymbol{u}}_t := (\hat{u}_{t,1}, \ldots, \hat{u}_{t,d})'$.

PROOF. See Appendix C.6.1 on p. S40 of [27]. □

Proposition 1 combined with Theorem 1(ii) provides pivot CUSUM statistics when used in conjunction with a statistical functional, such as the supremum.

COROLLARY 1 (Pivot statistic). *Assume that the following conditions hold.*

(a) $\Sigma_{u,T} = \Sigma_u + o_P(1)$, *where* Σ_u *is the covariance matrix of* $\boldsymbol{u}_t$.

(b) We have process convergence $\sqrt{T}(\mathfrak{U}_{1,T}(s), \ldots, \mathfrak{U}_{d,T}(s))' \xrightarrow[T\to\infty]{\mathscr{L}} \Sigma_u^{1/2}(B_1(s), \ldots, B_d(s))'$, *where* $B_1, \ldots, B_d$ *are independent Brownian motion processes on* $[0,1]$, *and* $\Sigma_u^{1/2}$ *is the lower-triangular invertible Cholesky matrix.*

Then, under the assumptions of Proposition 1 and Assumptions 4(b)-(d) for each ARMAX model of the d-dimensional system, we have process convergence

$$\hat{\Sigma}_{u,T}^{-1/2}\sqrt{T}(\hat{\mathfrak{V}}_{1,T}(s), \ldots, \hat{\mathfrak{V}}_{d,T}(s))' \xrightarrow[T\to\infty]{\mathscr{L}} (B_1^\circ(s), \ldots, B_d^\circ(s))',$$

where $B_1^\circ, \ldots, B_d^\circ$ *are independent Brownian bridge processes on* $[0,1]$.

PROOF. See Appendix C.6.2 on p. S45 of [27]. □

Corollary 1 immediately implies the asymptotic distribution of various multivariate CUSUM-statistics, whose limit distribution is easily identifiable due to the independence of the above Brownian bridge processes. For brevity we do not discuss this further. Condition (a) corresponds to a multivariate functional central limit theorem, which has been proved in various settings [e.g 42, 17, Section 27.7] when (u_t) has no autocorrelation. Condition (b) just requires the usual empirical covariances to converge to the covariance. Such results have been proved under mild assumptions [e.g 42, Theorems 3.7 and 3.16 and Remark 3.9].

2.3. *Examples and immediate applications.* For simplicity, we consider different types of covariates separately. However, it is clear from the formula of $\Xi_{T,t}$, as well as Lemma 1 and Theorem 1 and their proofs, that, by the triangle inequality, we can jointly consider them (e.g., $\boldsymbol{\lambda}X_{t-1} = \boldsymbol{\lambda}_1X_{t-1}^{(1)} + \boldsymbol{\lambda}_2X_{t-1}^{(2)}$ with $(X_{t-1}^{(1)})$ an L^1 ergodic stationary process and where $(X_{t-1}^{(2)})$ contains seasonal dummy variables).

2.3.1. *ARMA with seasonal dummies.* Seasonality is present in many time series. One way to model seasonality is to introduce seasonal dummies. The following proposition shows how seasonal dummies affect the partial-sum processes of residuals. This is an example of practical importance, since many time-series are analyzed after being seasonally adjusted in this way. We see that the seasonal dummies induce extra terms in the expansion for the partial sum of the residuals, which vanish for average-corrected residuals.

PROPOSITION 2 (Seasonal dummy). *Let* $d = d_{\boldsymbol{\lambda}}+1$ *and* $\boldsymbol{X}_{T,t-1} = (I\{t \equiv 1 \pmod d\}, \ldots, I\{t \equiv d-1 \pmod d\})'$, *so that* $\boldsymbol{\lambda}'\boldsymbol{X}_{T,t-1} = \sum_{k=1}^{d-1} \lambda_k I\{t \equiv k \pmod d\}$ *models a seasonal component. Then, under Assumptions 1, 2, 3, and 4(a),*

(i) for all $l \in [\![1, d_{\boldsymbol{\lambda}}]\!]$, $\sup_{s\in[0,1]} |L_{1,T,l}(s) - s\frac{\Theta(1)^{-1}}{d}| = o(1)$ $\sup_{s\in[0,1]} |L_{2,i,T,l}(s) - s\frac{\Theta(1)^{-1}\Phi(1)^{-1}}{d}| = o(1)$ *for all* $i \in [\![1, p]\!]$, *where* $L_{1,T}(s) =: (L_{1,T,1}(s), L_{1,T,2}(s), \ldots, L_{1,T,d_{\boldsymbol{\lambda}}}(s))'$ *and* $L_{2,i,T}(s) =: (L_{2,i,T,1}(s), L_{2,i,T,2}(s), \ldots, L_{2,i,T,d_{\boldsymbol{\lambda}}}(s))'$*; and*

(ii) the conclusions of Theorem 1 hold (i.e., Assumptions 4(b)-(d) hold).

PROOF. See Appendix C.6.3 on p. S45 of [27]. □

2.3.2. *ARMA with covariates whose expectations are constant.* The following Proposition 3 provides conditions that ensure the assumptions of Theorem 1 regarding the covariates $(\boldsymbol{X}_{T,t-1})$ under common conditions.

PROPOSITION 3 (Covariates with constant expectations). *Assume that* $(\boldsymbol{X}_{T,t-1})_{t\in\mathbf{Z}}$ *does not depend on* T *(i.e., for all* $(T,t) \in \mathbf{Z}^2$, $\boldsymbol{X}_{T,t-1} = \boldsymbol{X}_{t-1}$*), and satisfies one of the following conditions.*

(a) $(\boldsymbol{X}_{t-1})_{t\in\mathbf{Z}}$ *is a strictly stationary and ergodic process.*

(a') For all $t \in \mathbf{Z}$, $\mathbb{E}\boldsymbol{X}_t = \mathbb{E}\boldsymbol{X}_0$, *and, for each* $k \in [\![1, d_{\boldsymbol{\lambda}}]\!]$, *there exist a* $\beta \in]0,1[$, *so that* $\sup_{k\in\mathbf{N}} \left[(1+k)^{\beta} \sup_{(i,j)\in[\![1,\infty[\![^2 : |i-j|=k} |\operatorname{Cov}(X_{i,k}, X_{j,k})|\right] < \infty$.

Then, under Assumptions 1, 2, 3, and 4(a)(b), Theorem 1 holds, (i.e., Assumptions 4(c)-(d) hold).

PROOF. See Appendix C.6.4 on p. S46 of [27]. □

Conditions (a) and (a') are weaker than the usual assumption in econometrics: See Appendix B.6 in [27], in which Propositions 10 and 11 provide a catalogue of sufficient conditions.

2.3.3. *Simplifications of the limit processes for zero-mean covariates.* In several particular cases, a stream of results, which go back at least to Ploberger and Krämer [44, Theorem 1], has shown that the scaled partial-sum process of residuals asymptotically behaves as the scaled partial-sum process of average-corrected errors. The following Proposition 4 provides general assumptions under which such results can be extended to full-blown ARMAX models.

PROPOSITION 4 (Equivalence between residuals and average-corrected errors). *Define the polynomial estimator* $\hat{\Phi}(z) := 1 - \hat{\phi}_1 z - \hat{\phi}_2 z^2 - \cdots - \hat{\phi}_p z^p$. *Assume that*

(a) $\mathbb{E}\boldsymbol{X}_{T,t} = 0$, *for all* $(t,T) \in \mathbf{Z} \times \mathbf{N}$;
(b) $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Y_t - \frac{\hat{\boldsymbol{\lambda}}'}{\hat{\Phi}(1)}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{T,t} + o_P(T^{-1/2})$.

Then, under Assumptions 1, 2, 3, 4 (a)-(c),

$$\sup_{s\in[0,1]} \sqrt{T}|\hat{\mathfrak{U}}_T(s) - \mathfrak{V}_T(s)| = o_P(1).$$

PROOF. See Appendix C.6.5 on p. S46 of [27]. □

The proposition identifies conditions that imply that the limit processes of residuals and average-corrected error terms are asymptotically equivalent. Assumption (a) of Proposition 4 can be fulfilled through a reparameterization [e.g., 44, Assumption A.2] when $\mathbb{E}\boldsymbol{X}_{T,t}$ does not depend on the sample size T. Assumption (b) of Proposition 4 is by the Phillips-Solo device [42] connected to the empirical average of the reduced form equation $Y_t - \mu = \Phi(B)^{-1}\boldsymbol{\lambda}\boldsymbol{X}_{T,t-1} + \Phi(B)^{-1}\Theta(B)u_t$ where the unobservable ARMA part $\Phi(B)^{-1}\Theta(B)u_t$ is left out, and where the unknown parameters are estimated. When there is no covariate (i.e., $\boldsymbol{X}_{T,t} = 0$, for all $(t,T) \in \mathbf{Z} \times \mathbf{N}$), assumption (b) requires that $\hat{\mu}$ corresponds to the average of (Y_t) modulo $o_P(T^{-\frac{1}{2}})$. In this case the common practice of average adjusting data prior to analysis trivially implies assumption (b).

REMARK 2. Because this equivalence is somewhat counter-intuitive, let us check Proposition 4 in the linear regression model with only a constant as regressor, i.e., $Y_t = \mu + u_t$. As recalled in the introduction, in this case, the OLS estimator is the average, i.e., $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Y_t$. Then $\hat{u}_t = Y_t - \frac{1}{T}\sum_{t=1}^{T} Y_t = (\mu + u_t) - \frac{1}{T}\sum_{t=1}^{T}(\mu + u_t)$, so that $\hat{\mathfrak{U}}_T(s) - \mathfrak{V}_T(s)$ is not only $o_P(1)$ as stated by Proposition 4, but is equal to zero. ⋄

The conclusion of Proposition 4 is contrary to what is expected from Bai [2], where partial-sum processes of residuals are asymptotically first order equivalent to partial-sum processes of error terms. The inclusion of the estimation of a mean parameter is seen to change the behavior of the partial-sum process of residuals in an abrupt manner. Proposition 4 extends this observation from the pure ARMA case treated in Yu [55] and Ghoudi and Rémillard [26], the AR case treated in Lee [36], and the linear regression case in [43, 44]. Ploberger and Krämer [44] derive the weak process limit of partial-sums of residuals when using the OLS in a linear regression problem under standard econometric assumptions. In their Theorem 1 [see eq. (12) and (13) of 44] it is shown that $\sup_{s\in[0,1]}\sqrt{T}|\hat{\mathfrak{U}}_T(s)-\mathfrak{V}_T(s)| = o_P(1)$ (in our notation). The stated conclusion of their Theorem 1 is a direct implication of this uniform approximation. Hence their Theorem 1 is a special case of Proposition 4.

3. Extensions and further applications.

3.1. *ARMA with infinite-variance errors.* Lemma 1 only assumes bounded $(1+\epsilon)$-moment, and specifically does not assume finite variance. This means that our results have implications for a wide range of ARMA models with infinite-variance errors.

There are not many results for ARMAX models with infinite-variance errors, but we mention Mikosch et al. [39], which works with ARMA models without intercept nor covariates, and Klüppelberg and Mikosch [32] which extends Mikosch et al. [39] to allow for model misspecification (the topic of Section 3.2) in the infinite variance case. In these papers, it is shown under IID conditions and under Assumption 1, that there are estimators $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}$ such that $(T/\log T)^{1/\alpha}(\hat{\boldsymbol{\phi}}' - \boldsymbol{\phi}', \hat{\boldsymbol{\theta}}' - \boldsymbol{\theta}')' = O_P(1)$. Since $\alpha < 2$, we have faster than $\sqrt{T}$-convergence, so Assumption 2 holds. Note that Mikosch et al. [39] allows for $0 < \alpha < 2$, but that $\alpha \leqslant 1$ is incompatible with Assumption 3(a) which our results require.

Assumption (a) of the following result assumes no covariates, but we allow for an intercept term. While Mikosch et al. [39] assumes $\mu = 0$, and their setting can be re-gained by setting $\hat{\mu} = \mu = 0$, there seems to be no inference theory for the case when μ is estimated. Even so, we include it and assume that it can be estimated at the $O_P(T^{-\frac{1}{2}})$ rate, since we see that the upcoming expansion of the partial-sum process of ARMA residuals is here not influenced by nuisance terms originating from the estimation of μ. Proposition 5 shows that the scaled difference between the residual partial-sum process and the error partial-sum process even goes to zero. This is a rare case where having an intercept term, or not, does not affect our expansions, which is counter to the intuition built up by Proposition 4 in the finite variance case. In the following proposition, all process convergences are understood in $D[0,1]$ using the J_1 Skorokhod metric where $D[0,1]$ denotes the space of càdlàg functions on $[0,1]$.

PROPOSITION 5 (ARMA with infinite-variance errors). *Assume that the following two conditions hold.*

(a) For all $(T,t) \in \mathbf{N} \times \mathbf{Z}$, $\boldsymbol{X}_{T,t-1} = 0$.
(b) There is an $\alpha \in]1,2[$ *such that* $T^{-1/\alpha} \sum_{t=1}^{\lfloor Ts \rfloor} u_t \xrightarrow[T\to\infty]{\mathscr{L}} Y(s)$, *where* Y *is non-degenerate.*

Under Assumptions 1, 2, and 3,

$$\frac{1}{T^{\frac{1}{\alpha}}} \sum_{t=1}^{\lfloor Ts \rfloor} \hat{u}_t = \frac{1}{T^{\frac{1}{\alpha}}} \sum_{t=1}^{\lfloor Ts \rfloor} u_t + o_P(1) \xrightarrow[T\to\infty]{\mathscr{L}} Y(s).$$

PROOF. See Appendix D.1 on p. S48 of [27]. □

Assumption (a) of Proposition 5 rules out covariates. Assumption (b) of Proposition 5 corresponds to a functional version of eq. (2.2) in [39]. Unlike the Brownian motion, the limit process Y is not continuous, and we therefore use the J_1 Skorokhod topology, see Section 12 in Billingsley [11].

3.2. *Misspecification.* In this section, we investigate the situation in which one fits an ARMAX model to the observations $(Y_t, \boldsymbol{X}_{t-1})_{t=1}^T$, although the process $(Y_t, \boldsymbol{X}_{t-1})_{t=1}^T$ does not need to solve eq. (1.1) (p. 3) for a process (u_t) that is IID or even weak white noise. Standard estimators then do not have their usual interpretation, but usually converge towards least-false parameters, see e.g. Dahlhaus et al. [16] and Klüppelberg and Mikosch [32].

Our core results also hold in such settings. The present section explores these settings (i) by deriving the formula of the error term (u_t) for a given set of least-false parameters (Lemma 2, p. 19), (ii) by providing assumptions directly on the observables to verify our core assumptions (Proposition 6, p. 19), and (iii) by deducing the limiting behaviour of the partial sum process of average corrected residuals when the data generating process is a linear process (Corollary 2, p. 20). Note that in this section, we do not allow the covariates $(\boldsymbol{X}_t)$ to depend on T because this may lead to a dependence on T also for (u_t), which would require an extension of our main results.

The following lemma shows that, under Assumption 1 (b), for any process $(Y_t, \boldsymbol{X}_{t-1})_{t\in\mathbf{Z}}$ with bounded first absolute moments, there exists a process

$(u_t)_{t\in\mathbf{Z}}$ s.t. $(Y_t, \boldsymbol{X}_{t-1}, u_t)_{t\in\mathbf{Z}}$ is an ARMAX process, which corresponds to equation (1.1) on p. 3. Such a result parallels the fundamental econometric OLS assumption for a linear regression model $Y_t = \boldsymbol{X}_t'\boldsymbol{\lambda} + u_t$ given by $\mathbb{E}[\boldsymbol{X}_t u_t] = 0$, which can either be seen as an assumption on u_t, or as a requirement for $\boldsymbol{\lambda}$ which defines the error term u_t using the observations [see e.g. 31, Section 2.9].

LEMMA 2 (ARMAX representation of arbitrary processes). *Any process* $(Y_t, \boldsymbol{X}_{t-1})_{t\in\mathbf{Z}}$ *s.t.* $\sup_{t\in\mathbf{Z}} \mathbb{E}|Y_t| < \infty$ *and* $\sup_{t\in\mathbf{Z}} \mathbb{E}|\boldsymbol{X}_{t-1}| < \infty$*, where the* Y_t *are random scalars and the* $\boldsymbol{X}_{t-1}$ *random vectors, defines an ARMAX process that corresponds to equation* (1.1) *on p. 3 for*

$$u_t := \Theta(B)^{-1}\left[\Phi(B)(Y_t - \mu) - \boldsymbol{\lambda}'\boldsymbol{X}_{t-1}\right], \quad t \in \mathbf{Z},$$

where $\mu, \boldsymbol{\lambda}, \Theta(B)$ *and* $\Phi(B)$ *are respectively any chosen scalar, vector of the same dimension as* $(\boldsymbol{X}_{t-1})$ *and lag polynomials (of finite order) s.t.* $\Theta(B)$ *is invertible.*

PROOF. See Appendix D.2.1 on p. S49 of [27]. □

In order to apply our results, we make assumptions on the observable processes, which leads to conditions verifying Assumption 3. This is achieved under the assumptions of Proposition 6, which can be checked by the same arguments that lead to Assumption 3. In view of Lemma 18 of [27] (Appendix B.6, p. S19), assumption (b) of Proposition 6 is weak, but it still rules out cases where there is a global misspecification of the expectation structure: For example, if we do not model a trend in the mean of a time-series under consideration, assumption (b) of Proposition 6 typically does not hold.

PROPOSITION 6 (Theorem 1 for misspecified ARMAX). *Assume that*

(a) for a constant $\epsilon_Y > 0$, $\sup_{t\in\mathbf{Z}} \mathbb{E}|Y_t|^{1+\epsilon_Y} < \infty$*; and that*
(b) $\sup_{s\in[0,1]} \left| \frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor} [\Phi(B)(Y_t - \mu) - \boldsymbol{\lambda}'\mathbb{E}(\boldsymbol{X}_{t-1})] \right| = o_P(1)$.

Under Assumption 1(b),

(i) if Assumptions 4(b) and (c) hold, then Assumption 3 holds for (u_t) *as given in Lemma 2; and*
(ii) if Assumption 1(a), Assumption 2, and Assumption 4 hold, then Theorem 1 holds for (u_t) *as given in Lemma 2.*

PROOF. See Appendix D.2, p. S49 of [27]. □

We now provide an illustration of the above results for a class of linear processes.

COROLLARY 2. *Let* $Y_t - \mu = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ *where* (ε_t) *is a zero-mean IID sequence with finite non-zero variance,* $|\alpha_j| < M\rho^j$ *with* $M \in \mathbf{R}$ *and* $\rho \in]0,1[$*, for all* $j \in \mathbf{N}$*, and where there exists a* $j \in \mathbf{N}$ *s.t.* $|\alpha_j| > 0$*. If Assumptions 1 and 2 hold for an ARMA representation* $\Phi(B)(Y_t - \mu) = \Theta(B)u_t$ *where* $\Phi(B)$ *and* $\Theta(B)$ *are chosen finite lag polynomials, then we have process convergence*

$$\hat{\sigma}_u^{-1}\sqrt{T}\hat{\mathfrak{V}}_T(s) \xrightarrow[T\to\infty]{\mathscr{L}} \tau B^{\circ}(s) \text{ with } \tau := \frac{\left(\sum_{j=0}^{\infty} \check{\zeta}_j\right)\left(\sum_{j=0}^{\infty} \alpha_j\right)}{\sqrt{\left(\sum_{j=0}^{\infty} \check{\zeta}_j^2\right)\left(\sum_{j=0}^{\infty} \alpha_j^2\right)}},$$

where B° *is a Brownian Bridge on* $[0,1]$, $\hat{\sigma}_u^2 := (\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2) - (\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t)^2$ *(as in the univariate version of Proposition 1 on p. 13), and where* $(\check{\zeta}_j)$ *are the lag coefficients of* $\Theta(B)^{-1}\Phi(B)$*. Moreover, if the ARMA representation is correctly specified (i.e.,* $\Phi(B)^{-1}\Theta(B) = \sum_{j=0}^{\infty} \alpha_j B^j$*), then* $\tau = 1$*.*

PROOF. See Appendix D.2.3 on p. S51 of [27]. □

Comparison of Corollary 2 with Corollary 1 (p. 14) shows that the misspecification only affects the asymptotic behaviour of CUSUM-test statistics through the factor τ, which therefore can be seen as a robustness measure.

EXAMPLE 5. Let us revisit the basic model $Y_t = \mu + u_t$ from Example 1 (p. 9) under the additional assumptions of Corollary 2. This is an ARMA(0,0) so that $\check{\zeta}_j = I\{j = 0\}$, giving $\tau = \sum_{j=0}^{\infty} \alpha_j / \sum_{j=0}^{\infty} \alpha_j^2$. ⋄

REMARK 3. Proposition 5 (p. 18) combined with Basrak and Krizmanić [6] yields a counterpart of Corollary 2 for potentially misspecified ARMA models with infinite-variance errors as in Klüppelberg and Mikosch [32]. ⋄

3.3. *Nonparametric density estimation of the errors.* This section establishes uniform consistency of nonparametric density estimation of the p.d.f. of ARMAX errors. More precisely, under weak assumptions, the following theorem shows that the p.d.f. of the errors can be estimated using standard kernel estimation with the residuals in place of the unobserved errors.

THEOREM 2 (Kernel estimation). *Let* $f(.)$ *be a p.d.f. and* $f_T(.) := \frac{1}{Th_T}\sum_{t=1}^{T} K\left(\frac{.-u_t}{h_T}\right)$ *be a kernel estimator of* f *s.t.*

(a) *the bandwidth parameters* $(h_T)_{T\in\mathbf{Z}}$ *are a sequence of non-zero real numbers s.t.* $\sqrt{T}h_T^2 \to \infty$, *as* $T \to \infty$;
(b) *the kernel* $K(.)$ *is a Lipschitz-continuous function;*
(c) $\sup_{x\in\mathbf{R}} |f_T(x) - f(x)| = o_P(1)$.

Then, under Assumptions 1, 2, 3 and 4(a)(b),

$$\sup_{x\in\mathbf{R}} \left| \hat{f}_T(x) - f(x) \right| = o_P(1),$$

where $\hat{f}_T(x) := \frac{1}{Th_T} \sum_{t=1}^{T} K\left(\frac{x-\hat{u}_t}{h_T}\right)$.

PROOF. See Appendix D.3 on p. S54 of [27]. □

To the best of our knowledge, Theorem 2 is the first to establish consistency of kernel estimation of errors from full-blown ARMAX models. Theorem 2 generalizes a result in Bai [2] and complements one of the main theorems in [47]. Under stronger assumptions, Bai [2, p. 257, eq. 30] proves the same result for ARMA processes with known zero-mean parameter. Robinson [47] proves the same result for a wide class of zero-mean covariance-stationary processes without covariates. In Robinson [47, Theorem 3], the assumptions are neither weaker (e.g., they require finite second moment) nor stronger (e.g., less stringent conditions on the rate of convergence to zero of the bandwidth parameter), but they are more complicated, as pointed out in Bai [2, p. 257]. In Theorem 2, condition (a) is stronger than the usual bandwidth assumptions, which require $Th_T \to \infty$ or $Th_T^2 \to \infty$ [41, eq. 2.8, and eq. 3.6, respectively], but condition (a) is satisfied by usual "optimal" bandwidths, which are of order $T^{-\frac{1}{5}}$ [41, Lemma 4A, eq. 4.15 for $r = 2$]. Condition (b) is also satisfied by the usual "optimal" kernel,[2] the Epanechnikov kernel [21, sec. b], and other commonplace kernels (e.g., Gaussian kernel): Their derivatives are bounded so that they are Lipschitz-continuous by the mean-value theorem. Condition (c) corresponds to a standard result in kernel estimation, which has been proved under general conditions [41, Theorem 3A, eq. 3.7]. The proof of Theorem 2 easily follows from an intermediary result proved to establish the consistency of the empirical variance (Proposition 1, p. 13). This indicates that the toolbox developed in the present paper is useful beyond CUSUM tests.

[2]We write "optimal" in quotation marks, because the traditional criterion of optimality for bandwidth parameters and kernels is questionable [e.g., 52, chap. 1].

3.4. *Generalized average corrections.* Theorem 1 (ii.b) and Corollary 1 show that under Assumption 4 (d), the partial-sum process of average-corrected residuals divided by $\hat{\sigma}$ is an asymptotic pivot process under weak conditions. This enables the statistician to perform CUSUM type tests in a large set of cases.

There are practically relevant examples where the assumptions of Theorem 1 (i) hold, but Assumption 4 (d) does not, meaning we cannot use the simplification provided by Theorem 1 (ii). As we saw in Example 4, there may then be transformations of the residual process which are asymptotic pivot processes. We here develop a general framework which under weak additional assumptions leads to asymptotic pivots under the conditions of Theorem 1 (i), i.e., we generalize Theorem 1 (ii.b).

For brevity, we only provide a single illustration, extending Example 4. Another illustration is ARMA models with potentially a unit root and a polynomial time trend. The details of this case is lengthy and complex, and is given in a follow-up paper.

EXAMPLE 6 (Continuation of Example 4). Consider an ARMA model with $\boldsymbol{X}_{T,t-1} = I\{t \leqslant \mathfrak{p}T\}$. We here assume that $\mathfrak{p}$ is known. In the upcoming Section 3.5.2 (p. 27), we show that the estimation of $\mathfrak{p}$ does not substantially affect the following discussion due to certain adaptivity properties.

We may use a slight extension of the Phillips-Solo device [42] (see Lemma 9 on p. S6) and the calculations in Example 4 (p. 12) to get $L_{1,T}(s) = \Theta(1)^{-1}\frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor} I\{t \leqslant \mathfrak{p}T\} + o(1) = \Theta(1)^{-1}\frac{1}{T}\sum_{t=1}^{\min(\lfloor Ts\rfloor, \lfloor \mathfrak{p}T\rfloor)} + o(1) = \Theta(1)^{-1}\min(s,\mathfrak{p}) + o(1)$ uniformly. Similarly, $L_{2,i,T}(s) = \Theta(1)^{-1}\Phi(1)^{-1}\min(s,\mathfrak{p}) + o(1)$ uniformly. Under the conditions of Theorem 1 (i), we see that $\sup_{s\in[0,1]}|\sqrt{T}[\hat{\mathfrak{U}}_T(s) - \mathfrak{U}_T(s)] + \zeta_T(s)| = o_P(1)$ where $\zeta_T(s) = s\Phi(1)\Theta(1)^{-1}\sqrt{T}(\hat{\mu}-\mu) - \Theta(1)^{-1}\min(s,\mathfrak{p})\sqrt{T}(\hat{\lambda} - \lambda) - \lambda\Theta(1)^{-1}\Phi(1)^{-1}\min(s,\mathfrak{p})\sum_{i=1}^{p}\sqrt{T}(\hat{\phi}_i - \phi_i)$ is of the form $\zeta_T(s) = \sum_{j=1}^{2} b_{T,j}g_j(s)$, with $g_1(s) = s$ and $g_2(s) = \min(s,\mathfrak{p})$, and where $b_{T,1}, b_{T,2}$ are both $O_P(1)$ random variables. $\diamond$

Under general assumptions, Theorem 1 (i) gives conditions for $\sup_{s\in[0,1]}|\sqrt{T}[\hat{\mathfrak{U}}_T(s) - \mathfrak{U}_T(s)] + \zeta_T(s)| = o_P(1)$ where $\zeta_T(s) = s[\Phi(1)/\Theta(1)]\sqrt{T}(\hat{\mu}-\mu) - \sqrt{T}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})' L_{1,T}(s) - \sum_{i=1}^{p}\sqrt{T}(\hat{\phi}_i - \phi_i)\boldsymbol{\lambda}' L_{2,i,T}(s)$, i.e.,

$$\zeta_T(s) = \sum_{j=1}^{n} b_{T,j}g_j(s), \tag{3.1}$$

in which $(b_{T,j})_{j=1}^{n}$ are $O_P(1)$ random variables that are unknown and not possible to estimate consistently, but where $(g_j)_{j=1}^{n}$ are either known or can be consistently estimated. Note that certain nuisance parameters may be included in $(b_{T,j})_{j=1}^n$, such as $\Phi(1)/\Theta(1)$ in Example 6. However, in that example, $g_2(s) = \min(s, \mathfrak{p})$ includes the nuisance parameter $\mathfrak{p}$, which cannot be absorbed in $(b_{T,j})_{j=1}^n$, but has to be estimated.

The central component of our proposed pivot transformation is the following bounded linear operator. For any bounded function $x : [0,1] \mapsto \mathbf{R}$ and a number $a \in [0,1]$ s.t. $x(a) \neq 0$, define the operator $\tilde{\Delta}[x,a]$ on functions with domain $[0,1]$ by

$$\tilde{\Delta}[x,a]y(s) = y(s) - \frac{x(s)}{x(a)}y(a). \tag{3.2}$$

Notice that if $\bar{u}_T = O_P(1)$ and $x(s) = s$ then

$$\begin{aligned}\sqrt{T}\mathfrak{V}_T(s) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Ts\rfloor}(u_t - \bar{u}_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Ts\rfloor}u_t - \sqrt{T}\frac{\lfloor sT\rfloor}{T}\bar{u}_T \\ &= \sqrt{T}\mathfrak{U}_T(s) - s\sqrt{T}\mathfrak{U}_T(1) + o_P(1) = \tilde{\Delta}[x,1]\sqrt{T}\mathfrak{U}_T(s) + o_P(1),\end{aligned}$$

where the third equality follows from Lemma 25 of [27] (Appendix C.1, p. S28). The linear operator $\tilde{\Delta}[x,a]$ therefore generalizes the average-correction of Theorem 1 (ii.b) on p. 11.

We also define the compounded operator $\tilde{\Delta}[\underline{g_n^{(n)}}, \underline{a}_n]$ with $\underline{g_n^{(n)}} := (g_1^{(1)}, \dots, g_n^{(n)})$, $\underline{a}_n := (a_1, \dots, a_n) \in [0,1]^n$ and $n \in \mathbf{N}$, through

$$\tilde{\Delta}[\underline{g_n^{(n)}}, \underline{a}_n] := \tilde{\Delta}[g_1^{(1)}, a_1] \circ \tilde{\Delta}[g_2^{(2)}, a_2] \circ \cdots \circ \tilde{\Delta}[g_n^{(n)}, a_n], \tag{3.3}$$

where $g_k^{(k)}$ is defined recursively through a starting set of functions $g_1(s), \dots, g_n(s)$. The recursion is

$$g_j^{(n)}(s) = g_j(s), \quad g_j^{(n-k-1)}(s) := g_j^{(n-k)}(s) - \frac{g_{n-k}^{(n-k)}(s)}{g_{n-k}^{(n-k)}(a_{n-k})}g_j^{(n-k)}(a_{n-k}) \tag{3.4}$$

for $k \in [\![0, n-2]\!]$ and $j \in [\![1, n-k]\!]$. The following lemma shows that $\tilde{\Delta}[\underline{g_p^{(p)}}, \underline{a_p}]$ is a transformation that has good properties and that cancels out the asymptotic gap between $\sqrt{T}\hat{\mathfrak{U}}_T$ and $\sqrt{T}\mathfrak{U}_T$.

LEMMA 3 (Algorithm to reach a pivot statistic). *Let $(a_1, a_2, \dots, a_n) \in [0,1]^n$ be real numbers with $n \in \mathbf{N}$. Let $g_1(s), g_2(s) \dots, g_n(s)$ be a set of known bounded real-valued functions with domain $[0,1]$ s.t., for all $j \in [\![1,n]\!]$, $g_j^{(j)}(a_j) \neq 0$ where $g_j^{(j)}$ are defined by recursion (3.4). Then, for any function $f : [0,1] \to \mathbf{R}$ of the form $f(s) = \sum_{j=1}^n b_j g_j(s)$,* *(i)* $\tilde{\Delta}[\underline{g_n^{(n)}}, \underline{a_n}]f(s) = 0$; *and*

(ii) *if, for all* $j \in [\![1, n]\!]$, $\sup_{s\in[0,1]} |g_j(s)| < \infty$, $\tilde{\Delta}[\underline{g_n^{(n)}}, \underline{a_n}]$ *is a linear bounded operator on any linear subspace of the space of real-valued functions with domain* $[0, 1]$, *and thus it is continuous on the same linear subspace.*

PROOF. See Appendix D.4.1 on p. S55 of [27]. □

Lemma 3 requires two assumptions on the functions g_js. Firstly, it requires them to be bounded over $[0, 1]$, which is a condition that is trivially satisfied in our applications. Secondly, it requires that $g_j^{(j)}(a_j) \neq 0$. This can be numerically checked, as the b_js do not enter in the definition of the $g_j^{(j)}$s: See recursion (3.4) on p. 23. In practice, choosing distinct a_j's appears to be sufficient to satisfy this second assumption.

We now apply Lemma 3. We assume uniformly consistent estimators $(\hat{g}_j)_{j=1}^n$ of $(g_j)_{j=1}^n$ to be at hand.

THEOREM 3. *Suppose* $\sup_{s\in[0,1]} |\sqrt{T}[\hat{\mathfrak{U}}_T(s) - \mathfrak{U}_T(s)] + \zeta_T(s)| = o_P(1)$ *in which* $\zeta_T(s) = \sum_{j=1}^n b_{T,j} g_j(s)$ *where* $b_{T,j} = O_P(1)$ *for* $j = 1, 2, \ldots, n$. *Let* $a_1, \ldots, a_n$ *be constants fulfilling the conditions of Lemma 3 w.r.t.* $g_1, \ldots, g_n$. *Suppose there are functions* $(\hat{g}_j)_{j=1}^n$ *computed from data s.t.* $\sup_{s\in[0,1]} |g_j(s) - \hat{g}_j(s)| = o_P(1)$ *for all* $j \in [\![1, n]\!]$. *Letting* $\underline{\hat{g}_n^{(n)}}$ *be defined through the recursion of eq. (3.4) starting with* $\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_n$, *the following holds.*

(i) *If* $\sqrt{T}\mathfrak{U}_T = O_P(1)$ *w.r.t. the uniform norm, we have*

$$\sup_{s\in[0,1]} \left| \tilde{\Delta}[\underline{\hat{g}_n^{(n)}}, \underline{a_n}][\sqrt{T}\hat{\mathfrak{U}}_T](s) - \tilde{\Delta}[\underline{g_n^{(n)}}, \underline{a_n}][\sqrt{T}\mathfrak{U}_T](s) \right| = o_P(1).$$

(ii) *If also* $\hat{\sigma}_u = \sigma_u + o_P(1)$ *with* $\sigma_u > 0$ *and* $\sigma_u^{-1}\sqrt{T}\mathfrak{U}_T \xrightarrow[T\to\infty]{\mathscr{L}} B$ *for some process* B, *we have process convergence*

$$\tilde{\Delta}[\underline{\hat{g}_n^{(n)}}, \underline{a_n}][\hat{\sigma}_u^{-1}\sqrt{T}\hat{\mathfrak{U}}_T](s) \xrightarrow[T\to\infty]{\mathscr{L}} \tilde{\Delta}[\underline{g_n^{(n)}}, \underline{a_n}][B](s).$$

PROOF. See Appendix D.4.2 on p. S58 of [27]. □

EXAMPLE 7. Continuing Example 6 from p. 22, we have $\hat{g}_1(s) := g_1(s) = s$ and $\hat{g}_2(s) := \min(s, \hat{\mathfrak{p}})$. In Appendix D.4.3 on p. S60 of [27] we show that $\sup_{s\in[0,1]} |\hat{g}_2(s) - g_2(s)| = o_P(1)$ as long as $\hat{\mathfrak{p}} = \mathfrak{p} + o_P(1)$, a weak assumption, since we typically have $T(\hat{\mathfrak{p}} - \mathfrak{p}) = O_P(1)$, see e.g. Bai [5, 3]. Hence, Theorem 3 can be applied, yielding a new type of CUSUM test. ⋄

REMARK 4. We note that the estimation of $(g_j)_{j=1}^n$ is a simpler problem than identifying critical values directly from the expansion of Theorem 1 (i). Using this expansion directly, one can typically show that $\hat{\sigma}_u^{-1}\sqrt{T}\hat{\mathfrak{U}}_T$ converges weakly to a zero-mean Gaussian process, whose covariance function depends on the functions $L_{1,T}$ and $L_{2,i,t}$ of eq. (2.2) as well as the asymptotic covariance matrix of $\sqrt{T}(\sigma_u^{-1}\sqrt{T}\mathfrak{U}_T, \hat{\mu}-\mu, \hat{\boldsymbol{\phi}}-\boldsymbol{\phi}, \hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$. Hence, one would need to estimate a greater number of nuisance parameters than when using Theorem 3. $\diamond$

3.5. *Nonlinear ARMAX models.* Consider the nonlinear ARMAX model

$$\Phi(B)(Y_t-\mu) = \boldsymbol{\lambda}' g(Z_{T,t}, \boldsymbol{\gamma}_0) + \Theta(B)u_t, \tag{3.5}$$

where $(Z_{T,t})$ is an observable time-series and $\boldsymbol{\gamma} \mapsto g(., \boldsymbol{\gamma})$ is a given function with domain Ω. In such a model, which is a useful extension of linear ARMAX [e.g., 9, sec. 7.4, pp. 152-153], the definition of the residuals is such that

$$\tilde{u}_t = \begin{cases} (Y_t-\hat{\mu}) - \sum_{i=1}^p \hat{\phi}_1 (Y_{t-i}-\hat{\mu}) - \sum_{j=1}^q \hat{\theta}_1 \hat{u}_{t-j} - \hat{\boldsymbol{\lambda}}' \hat{\boldsymbol{X}}_{T,t-1} & \text{for } t \in [\![1,T]\!] \\ 0 & \text{for } t \in \mathbf{Z}_-, \end{cases} \tag{3.6}$$

where $\hat{\boldsymbol{X}}_{T,t-1} := g(Z_{T,t}, \hat{\gamma})$. The only difference between the definition of $(\hat{u}_t)$ in eq. (1.2) on p. 4 and $(\tilde{u}_t)$ in eq. (3.6) is that $\boldsymbol{X}_{T,t-1}$ is replaced with $\hat{\boldsymbol{X}}_{T,t-1}$. Thus, subtracting eq. (1.2) to eq. (3.6), and then summing over t and multiplying by $\frac{1}{\sqrt{T}}$ yields

$$\sqrt{T}[\tilde{\mathfrak{U}}_T(s) - \hat{\mathfrak{U}}_T(s)] = \hat{\boldsymbol{\lambda}}' \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} [\hat{\boldsymbol{X}}_{T,t-1} - \boldsymbol{X}_{T,t-1}] = \hat{\boldsymbol{\lambda}}' \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} [g(Z_{T,t}, \hat{\gamma}) - g(Z_{T,t}, \boldsymbol{\gamma}_0)], \tag{3.7}$$

where $\tilde{\mathfrak{U}}_T(s) := \frac{1}{T}\sum_{t=1}^{\lfloor Ts \rfloor} \tilde{u}_t$. As the above equation relates $\tilde{\mathfrak{U}}_T(s)$ to $\hat{\mathfrak{U}}_T(s)$ in a simple manner, we can study the partial-sum residual process $\tilde{\mathfrak{U}}_T(s)$ through $\hat{\mathfrak{U}}_T(s)$, for which Lemma 1 and Theorem 1 provide uniform approximations. We here give two examples of this technique. One non-smooth example, where $g(Z_{T,t}, \boldsymbol{\gamma}_0) = I\{t \leqslant \mathfrak{p}T\}$ for $g(x, \boldsymbol{\gamma}_0) = I\{\boldsymbol{\gamma}_0^{-1} \leqslant x\}$ with $Z_{T,t} = T/t$ and $\boldsymbol{\gamma}_0 = \mathfrak{p}$, and one smooth example, where $\boldsymbol{\gamma} \mapsto g(Z_{T,t}, \boldsymbol{\gamma})$ is differentiable and $Z_{T,t}$ does not depend on T.

3.5.1. *Smooth nonlinear ARMAX models.* In this section, we assume $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda} = 1$ and $\boldsymbol{X}_{T,t-1} = g(Z_t, \boldsymbol{\gamma}_0)$, where (Z_t) is an observable time-series that does not depend on T as in Bierens [9, sec. 7.4], and our observations follow equation (3.5). Proposition 7 characterizes ζ_T, i.e., the asymptotic gap between the scaled partial-sum process of nonlinear ARMAX errors and its residual counterpart.

PROPOSITION 7 (Smooth nonlinear ARMAX). *Let (Y_t) and (Z_t) be two processes that satisfy eq. (3.5) where $\boldsymbol{\lambda}$ is known to be equal to one, i.e., $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda} = 1$. Let the function $g(.)$ in eq. (3.5) and $\hat{\boldsymbol{\gamma}}$, which denotes an estimator of $\boldsymbol{\gamma}_0$, be s.t. they satisfy the following assumptions.*

(a) $\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = O_P(1)$.

(b) For each x, the function $\boldsymbol{\gamma} \mapsto g(x, \boldsymbol{\gamma})$ is twice continuously differentiable in a neighborhood $V_{\boldsymbol{\gamma}_0}$ of $\boldsymbol{\gamma}_0$, and, for all $\gamma \in \Omega$, $x \mapsto g(x, \boldsymbol{\gamma})$ is measurable.

(c) There exists a vector G such that $\sup_{s\in[0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} \frac{\partial}{\partial \boldsymbol{\gamma}} g(Z_t, \boldsymbol{\gamma}_0) - sG \right| = o_P(1)$.

(d) $\sup_{t\in\mathbf{N}} \mathbb{E}\left[\sup_{\boldsymbol{\gamma}\in V_{\boldsymbol{\gamma}_0}} \left| \frac{\partial^2}{\partial \boldsymbol{\gamma}' \partial \boldsymbol{\gamma}} g(Z_t, \boldsymbol{\gamma}) \right|\right] < \infty$.

Then,

(i) $\sup_{s\in[0,1]} \left| \sqrt{T}[\tilde{\mathfrak{U}}_T(s) - \hat{\mathfrak{U}}_T(s)] + s\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)G \right| = o_P(1)$*; and*

(ii) under the additional Assumptions 1, 2, 3, and 4 (a)-(c), $\sup_{s\in[0,1]} \Big| \sqrt{T}[\tilde{\mathfrak{U}}_T(s) - \mathfrak{U}_T(s)] + s\frac{\Phi(1)}{\Theta(1)}\sqrt{T}(\hat{\mu} - \mu) - \sum_{i=1}^{p} \sqrt{T}(\hat{\phi}_i - \phi_i) L_{2,i,T}(s) - s\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)G \Big| = o_P(1)$.

PROOF. See Appendix D.5 on p. S60 of [27]. □

Proposition 7 shows that the asymptotic gap ζ_T between the scaled partial-sum process of nonlinear ARMAX $\sqrt{T}\mathfrak{U}_T$ and its residual counterpart $\sqrt{T}\tilde{\mathfrak{U}}_T$ has a structure similar to the linear case (i.e., Theorem 1(i) on p. 11). Writing ζ_T in the form of eq. (3.1) (p. 22), the first term in ζ_T can be written as $b_{T,1}g_1(s)$ where $b_{T,1} = b_{T,1}^{(1)} + b_{T,1}^{(2)} = O_P(1)$, with $g_1(s) = s$, $b_{T,1}^{(1)} = -\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)G$ and $b_{T,1}^{(2)} = (\Phi(1)/\Theta(1))\sqrt{T}(\hat{\mu} - \mu)$. Thus, using Theorem 3 (p. 24), the pivot transformation is the same as if we knew $\boldsymbol{\gamma}_0$ and thus observed $g(Z_t, \boldsymbol{\gamma}_0)$. The assumptions of Proposition 7 are quite general, although they could be weakened at the cost of more complex proofs. Primitive conditions for assumption (a) of Proposition 7 can be found in Bierens [9, p. 164, Theorem 8.2.4]. Assumptions (b) and (d) of Proposition 7 are standard primitive conditions to establish assumption (a) [e.g., 9, p. 166-167, Assumptions 8.2.1(b) and 8.2.4]. Assumption (c) of Proposition 7 is slightly nonstandard, but it is implied by the typical primitive conditions used to establish assumption (a). In particular, if (Z_t) is strictly stationary and ergodic, and $\mathbb{E}|\frac{\partial}{\partial\boldsymbol{\gamma}}g(Z_t,\boldsymbol{\gamma}_0)| < \infty$ then assumption (c) holds with $G = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\gamma}}g(Z_t,\boldsymbol{\gamma}_0)\right]$ by the ergodic theorem and a simple lemma from Nielsen and Sohkanen [40, Lemma 4.2.]. Note also that in the latter case, if $g(.)$ is linear and $(Z_t)_{t\in\mathbf{Z}}$ is zero-mean, $G = 0$, i.e., there is adaptivity in the trivial linear case.

3.5.2. *Change points.* In Examples 4 (p. 12) and 6 (p. 22), we worked with an ARMAX model which included a change-point type covariate. We here justify our claim that the estimation of the change-point does not affect our results under weak conditions. Note that adaptivity with respect to the estimation of the placement of the change-point also holds in related cases, such as in the parameter estimation theory of Bai [5, 3], and so this result is expected.

We here have $\boldsymbol{\lambda}' = \lambda$, which is univariate, and $\boldsymbol{X}_{T,t-1} = I\{t \leqslant \mathfrak{p}T\}$ and $\hat{\boldsymbol{X}}_{T,t-1} = I\{t \leqslant \hat{\mathfrak{p}}T\}$. In the following Proposition 8, we assume that $\lfloor\hat{\mathfrak{p}}T\rfloor = \lfloor\mathfrak{p}T\rfloor + o_P(T^{\frac{1}{2}})$. This assumption is considerably weaker than the expected T-convergence of change point problems found in e.g. Bai [3, 5], i.e., that $T(\hat{\mathfrak{p}} - \mathfrak{p}) = O_P(1)$ which implies that $\lfloor\hat{\mathfrak{p}}T\rfloor = \lfloor(\hat{\mathfrak{p}} - \mathfrak{p})T + \mathfrak{p}T\rfloor = \lfloor O_P(1) + \mathfrak{p}T\rfloor = \lfloor\mathfrak{p}T\rfloor + O_P(1)$.

PROPOSITION 8 (ARMAX with estimated change-point). *Let (Y_t) and $(Z_{T,t})$ be two processes that satisfy eq. (3.5) s.t. $g(Z_{T,t},\boldsymbol{\gamma}_0) = I\{t \leqslant \mathfrak{p}T\}$ for $\boldsymbol{\gamma}_0 = \mathfrak{p}$. Define the residuals s.t. $g(Z_{T,t},\hat{\boldsymbol{\gamma}}) = I\{t \leqslant \hat{\mathfrak{p}}T\}$ for $\hat{\boldsymbol{\gamma}} = \hat{\mathfrak{p}}$. If $\lfloor\hat{\mathfrak{p}}T\rfloor = \lfloor\mathfrak{p}T\rfloor + o_P(\sqrt{T})$, then*

(i) under Assumption 4(a), $\sup_{s\in[0,1]}\left|\sqrt{T}[\tilde{\mathfrak{U}}_T(s) - \hat{\mathfrak{U}}_T(s)]\right| = o_P(1)$*; and*

(ii) under the additional Assumptions 1, 2, and 3, $\sup_{s\in[0,1]}\Big|\sqrt{T}[\tilde{\mathfrak{U}}_T(s) - \mathfrak{U}_T(s)] + s\frac{\Phi(1)}{\Theta(1)}\sqrt{T}(\hat{\mu} - \mu) - \Phi(1)^{-1}\min(s,p)\sqrt{T}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) - \Theta(1)^{-1}\Phi(1)^{-1}\min(s,p)\sum_{i=1}^{p}\sqrt{T}(\hat{\phi}_i - \phi_i)\Big| = o_P(1).$

PROOF. See Appendix D.5 on p. S61 of [27]. □

In the setting of Proposition 8, a pivot transformation was identified in Examples 6 and 7.

Acknowledgements. We thank Bent Nielsen (University of Oxford) for very helpful discussions on the paper and for introducing us to [40]. We also thank Genaro Sucarrat (BI Norwegian Business School) for motivating us to study partial-sum processes of residuals in connection with the paper Sucarrat, Grønneberg and Escribano [51], and seminar/conference participants at the University of Oxford, and CFE-CMStatistics 2016. Finally, we thank three referees for their constructive comments which improved the paper.

References.

[1] Andreou, E. and Werker, B. J. M. (2012). An Alternative Asymptotic Analysis of Residual-Based Statistics. *The Review of Economics and Statistics,* **94** 88-99.

[2] Bai, J. (1993a). On the partial sums of residuals in autoregressive and moving average models. *Journal of Time Series Analysis* **14** 247-260.

[3] Bai, J. (1993b). Estimation of structural change based on Wald-type statistics. Technical Report. Dept. of Economics, Massachusetts Institute of Technology.

[4] Bai, J. (1994a). Weak convergence of the sequential empirical process of residuals in ARMA models. *The Annals of Statistics* **22** 2051-2061.

[5] Bai, J. (1994b). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* **15** 453–472.

[6] Basrak, B. and Krizmanić, D. (2014). A limit theorem for moving averages in the α-stable domain of attraction. *Stochastic Processes and their Applications* **124** 1070–1083.

[7] Berenguer-Rico, V. and Nielsen, B. (2015). Cumulated sum of squares statistics for non-linear and non-stationary regressions Technical Report. Nuffield Discussion Paper 2015-W09, University of Oxford.

[8] Beveridge, S. and Nelson, C. (1981). A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycles. *Journal of Monetary Economics* **7** 151-174.

[9] Bierens, H. J. (1996). *Topics in advanced econometrics: estimation, testing, and specification of cross-section and time series models.* Cambridge University Press, Cambridge.

[10] Billingsley, P. (1995). *Probability and Measure.* John Wiley & Sons, New York. 3rd. Edition.

[11] Billingsley, P. (2013). *Convergence of probability measures.* John Wiley & Sons, New York. 2nd edition.

[12] Boldin, M. (1983). Estimation of the distribution of noise in an autoregression scheme. *Theory of Probability & Its Applications* **27** 866–871.

[13] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer, New York. 2nd. Edition.

[14] Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)* **37** 149–192.

[15] Chan, N. H. and Ling, S. (2008). Residual empirical processes for long and short memory time series. *The Annals of Statistics* **36** 2453–2470.

[16] Dahlhaus, R., Wefelmeyer, W. et al. (1996). Asymptotically optimal estimation in misspecified time series models. *The Annals of Statistics* **24** 952–974.

[17] Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press, Oxford.

[18] Dufour, J.-M. and Stevanović, D. (2013). Factor-augmented VARMA models with macroeconomic applications. *Journal of Business & Economic Statistics* **31** 491–506.

[19] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007.
[20] Engler, E. and Nielsen, B. (2009). The empirical process of autoregressive residuals. *The Econometrics Journal* **12** 367–381.
[21] Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications* **14** 153–158.
[22] Francq, C. and Sucarrat, G. (2017). Equation-by-Equation Estimation of a Multivariate Log-GARCH-X Model of Financial Returns. *Journal of Multivariate Analysis* **153** 16–32.
[23] Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* **68** 145–165.
[24] Francq, C. and Zakoïan, J.-M. (2000). Estimating weak GARCH representations. *Econometric theory* **16** 692–728.
[25] Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models*. John Wiley & Sons, New York.
[26] Ghoudi, K. and Rémillard, B. (2015). Diagnostic tests for innovations of ARMA models using empirical processes of residuals. In *Asymptotic Laws and Methods in Stochastics* 239–282. Springer, Berlin.
[27] Grønneberg, S. and Holcblat, B. (2018). Supplement to "On Partial-Sum Processes of ARMAX Residuals". Submitted.
[28] Hannan, E. J. and Deistler, M. (2012). *The statistical theory of linear systems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Originally published in 1988 by Wiley, New York.
[29] Hannan, E., Dunsmuir, W. and Deistler, M. (1980). Estimation of vector ARMAX models. *Journal of Multivariate Analysis* **10** 275–295.
[30] Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)* **41** 190–195.
[31] Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton.
[32] Klüppelberg, C. and Mikosch, T. (1996). Parameter estimation for a misspecified ARMA model with infinite variance innovations. *Journal of Mathematical Sciences* **78** 60–65.
[33] Koreisha, S. G. and Fang, Y. (2001). Generalized least squares with misspecified serial correlation structures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63** 515–531.
[34] Koul, H. L., Levental, S. et al. (1989). Weak convergence of the residual empirical process in explosive autoregression. *The Annals of Statistics* **17** 1784–1794.
[35] Krämer, W., Ploberger, W. and Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica* **56** 1355–1369.
[36] Lee, S. (1997). A note on the residual empirical process in autoregressive models. *Statistics & probability letters* **32** 405–411.
[37] Leeb, H. and Pötscher, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* **24** 338–376.
[38] MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics* **6** 422–433.
[39] Mikosch, T., Gadrich, T., Kluppelberg, C. and Adler, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics* **23** 305–326.
[40] Nielsen, B. and Sohkanen, J. S. (2011). Asymptotic behavior of the cusum of

squares test under stochastic and deterministic time trends. *Econometric Theory* **27** 913–927.

[41] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of mathematical statistics* **33** 1065–1076.

[42] Phillips, P. C. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics* **20** 971–1001.

[43] Ploberger, W. and Krämer, W. (1987). Mean adjustment and the CUSUM test for structural change. *Economics Letters* **25** 255–258.

[44] Ploberger, W. and Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica* **60** 271–285.

[45] Ploberger, W. and Krämer, W. (1996). A trend-resistant test for structural change based on OLS residuals. *Journal of Econometrics* **70** 175–185.

[46] Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory* **7** 163–185.

[47] Robinson, P. (1987). Time series residuals with application to probability density estimation. *Journal of time series analysis* **8** 329–344.

[48] Sensier, M. and Van Dijk, D. (2004). Testing for volatility changes in US macroeconomic time series. *Review of Economics and Statistics* **86** 833–839.

[49] Shin, D. W. and Lee, J. H. (2000). Consistency of the maximum likelihood estimators for nonstationary ARMA regressions with time trends. *Journal of Statistical Planning and Inference* **87** 55–68.

[50] Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics* **59**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Originally published in 1986 by Wiley, New York.

[51] Sucarrat, G., Grønneberg, S. and Escribano, A. (2016). Estimation and inference in univariate and multivariate log-GARCH-X models when the conditional density is unknown. *Computational Statistics & Data Analysis* **100** 582–594.

[52] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation.* Springer, Berlin.

[53] Wright, J. H. (1993). The CUSUM test based on least squares residuals in regressions with integrated variables. *Economics Letters* **41** 353–358.

[54] Xiao, Z. and Phillips, P. C. (2002). A CUSUM test for cointegration using regression residuals. *Journal of Econometrics* **108** 43–61.

[55] Yu, H. (2007). High moment partial sum processes of residuals in ARMA models and their applications. *Journal of Time Series Analysis* **28** 72-91.

Department of Economics
BI Norwegian Business School
Oslo, Norway 0484
E-mail: steffeng@gmail.com

Université du Luxembourg, LSF
Luxembourg 1246
E-mail: Benjamin.Holcblat@uni.lu