Oxford Bulletin of Economics and Statistics

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Least Trimmed Squares: Cointegration and Outliers

Vanessa Berenguer-Rico[1] | Bent Nielsen[2]

[1]Mansfield College & Department of Economics, University of Oxford, Oxford, UK | [2]Nuffield College & Department of Economics, University of Oxford, Oxford, UK

Correspondence: Bent Nielsen (bent.nielsen@nuffield.ox.ac.uk)

Received: 1 April 2026 | **Accepted:** 8 April 2026

Keywords: asymptotic theory | cointegration | least trimmed squares | outliers

ABSTRACT

When applying the cointegrated autoregressive distributed lag model it is common to include indicator variables for outliers. This is often done in a somewhat ad hoc way. Least Trimmed Squares estimation provides a more systematic approach. This estimator is robust to a large number of outliers of many types. We analyse the estimator in a model that allows a range of contamination and show that it has the same asymptotic properties as the infeasible Ordinary Least Squares estimator applied to a model generated by the good errors.

1 | Introduction

When applying a cointegrated Autoregressive Distributed Lag (ADL) regression, it is common to include indicator variables for outlying errors. This is done out of a concern that inference may be distorted if there are unmodelled outliers and an intuition that standard inference may be valid when outliers are modelled. We investigate this intuition through an asymptotic analysis of the Least Trimmed Squares (LTS) estimator.

A simple approach to outlier detection is to apply Ordinary Least Squares (OLS) to the full sample, remove or dummy out observations with outlying residuals and reestimate the model by OLS. This approach has long been used in econometric analysis of time series. Early examples include indicators and level shifts in UK economic models [1, 2]. For instance, the latter includes a consumption function with dummies relating to the 1968 introduction of purchase tax. These dummies were later adopted by Davidson et al. [3] in their consumption function analysis using an ADL model in equilibrium correction form. This simple approach relies on consistency of the initial OLS estimator. As outliers can bias the initial OLS estimator, this procedure is not robust in general and it has to be used with care [4].

The robustness concern has lead to various algorithms that search for outliers but do not start from full sample OLS estimators. This includes the Forward Search [5] with an implementation to structural time series in [6], and Impulse Indicator Saturation [7] which is aimed at ADL models and implemented in OxMetrics [8], as Gets in R [9] and in the Eviews software. Asymptotic analysis of these methods has focused on the situation without outliers [10–12] and with little emphasis on cointegration. In time series, it is common to distinguish between additive and innovative outliers [13], both of which can be relevant in cointegrated models [14]. Methods based on extreme value theory can detect a finite number of additive outliers in a first order autoregression [15]. Here, the aim is to cover fairly general contamination including a diverging number of additive or innovative outliers. In order to make progress with the theory, we consider the LTS estimator as vehicle for analysis in models with contamination. LTS has not been used much in time series econometrics, although some simulation evidence for a stationary vector autoregression (VAR) is available [16].

The least trimmed squares (LTS) estimator [17] is defined as follows. The investigator specifies that there are h 'good' observations and $T-h$ 'outliers' in a sample of T observations. The set of good observations is estimated by the h-subsample with the smallest residual sum of squares. OLS is then applied to the estimated set of good observations. In particular, if $h = T$ then LTS is simply full sample OLS.

The literature offers two models for outliers with very different inference theories. First, in the Huber [18] contamination model, all regression errors are independent draws from a normal distribution mixed with a contamination distribution. This formulation excludes leverage effects whereby the locations of the outlier errors depend on the regressors. Asymptotic theory has been developed for the case of stationary regressors and symmetric contamination. Even so, inference involves nuisance parameters depending on the contamination distribution and proportion [19]. It is therefore challenging to conduct reliable statistical inference in this model.

Second, LTS has recently been found to be maximum likelihood in a regression model where h good observations have normal errors and $T - h$ outliers have errors that are more extreme than the realized good observations [20]. Under regularity conditions, it can be shown that the estimator is asymptotically bounded in probability with the oracle property that it has the same asymptotic distribution as the OLS estimator applied infeasibly to the actual good observations [21]. In particular, LTS is asymptotically efficient relative to the infeasible OLS estimator with the same asymptotic power properties.

In this paper we check the LTS regularity conditions for an ADL regression for data generated by a vector autoregression with cointegration. We recall that, two variables are cointegrated if they have random walk trends, but a linear combination does not [22, 23].

The general LTS theory balances two types of regularity conditions for the regressors. First, to show boundedness of the LTS estimator, it is assumed that the regressors are not too concentrated. Second, to derive an asymptotic expansion of the LTS estimator, it is assumed that the regressors are not too spread out. These LTS conditions have not been fully analysed in the context of cointegrated processes. We do so here using a cointegrated ADL model. The proofs require a modification of the classic cointegration representations [22, 24, 25] to permit outliers and in format that retains the autoregressive structure.

We find that if the proportion of outliers vanishes, but their number possibly diverges, then the LTS estimator has the oracle property in a cointegrated ADL model. Due to the autoregressive structure, the amount of outliers that the LTS estimator can cope with in the cointegrated ADL case is lower than in cross-sectional models. Notwithstanding, the oracle property holds with a diverging number of outliers.

The practical consequence of the results is that the asymptotic theory known for OLS estimation of ADL models without outliers transfers to LTS estimation of ADL models with outliers. In particular, under weak exogeneity [26, 27], the hypothesis of no cointegration can be tested using Dickey–Fuller-type distributions with a likelihood ratio statistic [28] or various t-statistics [29, 30]; tests on coefficients in the cointegrating vector have standard normal inference [27, 31]; and tests for lag length [32] can be used.

In the analysis of LTS, the number of good observations, h, will be taken as given. Estimation of h will be discussed in the empirical application and in the conclusion.

Outline: Section 2 describes the ADL equation and the LTS estimator. Section 3 presents a vector autoregression describing the system of variables along with the Granger–Johansen representation. Section 4 describes the data generating process including the outliers. Section 5 presents the asymptotic theory for LTS applied to cointegrated ADLs. Section 6 has simulations illustrating the theory. Section 7 gives an empirical illustration using consumption data. Section 8 concludes. An Appendix has the technical derivations.

2 | Regression Equation and Estimation Method

We describe the ADL equation that is to be used in modelling and the LTS estimator.

2.1 | Autoregressive Distributed Lag Equation

We consider an ADL regression in equilibrium correction form for a scalar y_t given a $(p-1)$-dimensional vector z_t. Let $\boldsymbol{x}_t = (y_t, z_t')'$. The regression equation is

$$\Delta y_t = \omega' \Delta z_t + \alpha\left(y_{t-1} - \kappa' z_{t-1} - \nu_c\right) + \sum_{j=1}^{k-1} \gamma_j' \Delta \boldsymbol{x}_{t-j} + \sigma \varepsilon_t \quad \text{for } t = 1, \ldots, T. \tag{2.1}$$

The joint distribution of the contemporaneous regressor z_t and the errors ε_t is described in subsequent Sections. In vector notation, the ADL equation is equivalent to

$$\Delta y_t = x_t' \beta + \sigma \varepsilon_t \qquad \text{for } t = 1, \ldots, T, \tag{2.2}$$

where y_t is as before, the regressor vector is $x_t = (\Delta z_t', y_{t-1}, z_{t-1}', \Delta \boldsymbol{x}_{t-1}', \ldots, \Delta \boldsymbol{x}_{t-k+1}', 1)'$ and the regression parameter is $\beta = (\omega', \alpha, -\alpha\kappa', \gamma_1', \ldots, \gamma_{k-1}', -\alpha\nu_c)'$.

We also consider a model with a linear trend, for which we have

$$\Delta y_t = \omega' \Delta z_t + \alpha\left(y_{t-1} - \kappa' z_{t-1} - \nu_\ell t\right) + \sum_{j=1}^{k-1} \gamma_j' \Delta \boldsymbol{x}_{t-j} + \mu_c + \sigma \varepsilon_t, \tag{2.3}$$

the regressor vector $x_t = (\Delta z_t', y_{t-1}, z_{t-1}', \Delta \boldsymbol{x}_{t-1}', \ldots, \Delta \boldsymbol{x}_{t-k+1}', 1, t)'$ and the regression parameter is $\beta = (\omega', \alpha, -\alpha\kappa', \gamma_1', \ldots, \gamma_{k-1}', \mu_c, -\alpha\nu_\ell)'$.

2.2 | Least Trimmed Squares Estimation

LTS estimation was suggested by Rousseeuw [17]. The estimator divides the data in two groups. There is a given number of h good errors with indices in an unknown h-subset ζ of $1, \ldots, T$. The indices in ζ need not be consecutive. The remaining $T - h$ indices in ζ^c are the outliers. The LTS estimator finds the h-subsample with the smallest residual sum of squares [33]. Thus, the LTS estimator is defined as follows. Given a h-index set ζ, the OLS estimators are

$$\hat{\beta}_\zeta = \left(\sum_{t\in\zeta} x_t x_t'\right)^{-1} \sum_{t\in\zeta} x_t \Delta y_t \quad \text{and} \quad \hat{\sigma}^2_\zeta = h^{-1}\sum_{t\in\zeta}\left(\Delta y_t - x_t'\hat{\beta}_\zeta\right)^2, \tag{2.4}$$

where $\sum_{t\in\zeta} x_t x_t'$ is assumed invertible for any choice of ζ. In passing, we note that this follows from the assumptions below, see Appendix A. The LTS estimators are then

$$\hat{\zeta} = \arg\min_{\zeta} \hat{\sigma}^2_\zeta, \qquad \hat{\beta} = \hat{\beta}_{\hat{\zeta}}, \qquad \hat{\sigma}^2 = \hat{\sigma}^2_{\hat{\zeta}}. \tag{2.5}$$

As the number of h-sets is finite, we need not be concerned about measurability issues. LTS reduces to full-sample OLS when all errors are good, $h = T$.

LTS estimation requires evaluation of all h-subsets of the n observations. The computational order is huge. This computational problem can be approximated by the fast LTS algorithm [33, 34].

3 | The Vector Autoregression

For inference, we set up a joint vector autoregressive model for the variables. Next, we consider the special case of a univariate autoregression to provide intuition for the importance of initial observations of good episodes in outlier analysis. We then elaborate on the Granger–Johansen representation to show how initial observations are transmitted.

3.1 | Definition of the Model

The vector autoregression. The ADL equations involve a p-vector of observations $\boldsymbol{x}_t = (y_t, z_t')'$. We describe the distribution of $\boldsymbol{x}_t$ by an unobserved components formulation with either a constant level or a linear trend as in (2.1) or (2.3). Thus, let

$$\boldsymbol{x}_t = \boldsymbol{x}_t^* + \boldsymbol{\tau}_c \qquad \text{or} \qquad \boldsymbol{x}_t = \boldsymbol{x}_t^* + \boldsymbol{\tau}_c + \boldsymbol{\tau}_\ell t, \tag{3.1}$$

where the vector $\boldsymbol{x}_t^*$ satisfies a vector autoregression (VAR)

$$\Delta \boldsymbol{x}_t^* = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1}^* + \sum_{j=1}^{k-1}\boldsymbol{\Gamma}_j \Delta \boldsymbol{x}_{t-j}^* + \boldsymbol{A}\boldsymbol{\varepsilon}_t \qquad \text{for } t = 1, \ldots, T. \tag{3.2}$$

We describe the distribution of the errors in Section 5. The Granger–Johansen representation manipulates the Equation (3.2), but does not rely on the distribution of the errors. That distribution will involve outliers. Despite the outliers, we apply the terminology of Johansen [25] and refer to, for example, $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}^*$ as the cointegrating relation. The parameters satisfy $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{p\times r}$ and $\boldsymbol{\Gamma}_j, \boldsymbol{\Omega} \in \mathbb{R}^{p\times p}$, such that $\boldsymbol{\Omega}$ is positive definite and

$$\boldsymbol{\Omega} = \boldsymbol{A}\boldsymbol{A}' = \begin{pmatrix} \boldsymbol{\Omega}_{yy} & \boldsymbol{\Omega}_{yz} \\ \boldsymbol{\Omega}_{zy} & \boldsymbol{\Omega}_{zz} \end{pmatrix}.$$

Relation between ADL and VAR parameters and errors. When linking the ADL to the VAR we rely on partial system analysis [27] and the notion of weak exogeneity [26]. For this purpose, we assume a unit cointegrating rank, a unit coefficient for the first element of the cointegrating vector and a weak exogeneity assumption restricting the adjustment to the cointegrating vector, that is

$$r = 1, \qquad \boldsymbol{\beta}' = \left(1\ ,\ -\kappa'\right), \qquad \boldsymbol{\alpha} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}. \tag{3.3}$$

For the derivation of the ADL equation, define the population regression coefficient $\omega' = \boldsymbol{\Omega}_{yz}\boldsymbol{\Omega}_{zz}^{-1}$. The ADL equation is obtained by pre-multiplying $\boldsymbol{x}_t$ by $(1, -\omega)$, exploiting the Equations (3.1), (3.2) and solving for Δy_t. This leads to the ADL Equations (2.1), (2.3) with

$$\gamma_j = (1, -\omega)\boldsymbol{\Gamma}_j, \quad \sigma^2 = \left(1\ ,\ -\omega'\right)\boldsymbol{\Omega}\begin{pmatrix} 1 \\ -\omega \end{pmatrix} = \Omega_{yy} - \Omega_{yz}\Omega_{zz}^{-1}\Omega_{zy}.$$

The ADL errors are defined through

$$\sigma\varepsilon_t = \left(1\ ,\ -\omega'\right)\boldsymbol{A}\boldsymbol{\varepsilon}_t. \tag{3.4}$$

For the deterministic quantities, we define $\boldsymbol{\Psi} = I_p - \sum_{j=1}^{k-1}\boldsymbol{\Gamma}_j$ and get either

$$\nu_c = \boldsymbol{\beta}'\boldsymbol{\tau}_c \qquad \text{or} \qquad \nu_\ell = \boldsymbol{\beta}'\boldsymbol{\tau}_\ell, \quad \mu_c = (1, -\omega')\boldsymbol{\Psi}\boldsymbol{\tau}_\ell - \alpha\boldsymbol{\beta}'\boldsymbol{\tau}_c.$$

The implied triangular system. The VAR Equation (3.2) also implies an equation for the regressor z_t. Consider the case with a constant level, so that $\Delta\boldsymbol{x}_t = \Delta\boldsymbol{x}_t^*$. Pre-multiplying Equation (3.2) by $(0, I_{p-1})$ and using the restriction to $\boldsymbol{\alpha}$ in (3.3) shows that $\Delta z_t = (0, I_{p-1})\Delta\boldsymbol{x}_t$ satisfies

$$\Delta z_t = \sum_{j=1}^{k-1}\left(0\ ,\ I_{p-1}\right)\boldsymbol{\Gamma}_j\Delta\boldsymbol{x}_{t-j} + \left(0\ ,\ I_{p-1}\right)\boldsymbol{A}\boldsymbol{\varepsilon}_t. \tag{3.5}$$

Taken together with the ADL Equation (2.1), we get a triangular system where z_t feeds into the ADL equation for y_t given z_t. When $\boldsymbol{\varepsilon}_t$ is normal, we find that the errors in the ADL Equation (2.1), (3.4) are independent of those in the Equation (3.2) for z_t.

The situation where there is an outlier in the Equation (3.5) for z_t at a particular t, but not in the errors (3.4) of the ADL Equation (2.1), is of special interest. The ADL equation may then have a structural interpretation. This relates to the ideas of super-exogeneity [26, 35] and causal transmission [36]. We will allow for this situation.

3.2 | A Univariate Representation

For the distributional analysis we must turn the autoregressive model Equation (2.1) into a moving average type representation. We will be interested in the role of the initial observation. It is useful to start from the special case of a univariate first order autoregression

$$\Delta y_t = \alpha y_{t-1} + \sigma\varepsilon_t \qquad \text{for } t = 1, \ldots, T, \tag{3.6}$$

with some initial observation y_0. In general, we have the solution

$$y_t = \sum_{s=0}^{t-1}(1+\alpha)^s\sigma\varepsilon_{t-s} + (1+\alpha)^t y_0, \tag{3.7}$$

which applies regardless of any distributional assumptions. We consider the I(0) case of a stationary root, $|1+\alpha| < 1$ and the I(1) case $\alpha = 0$ so that $1+\alpha = 1$ is a unit root.

In the I(0) case, $|1+\alpha| < 1$, the model Equation (3.6) has a stationary solution when the innovations are independent normal. We can express the stationary initial distribution in terms of an infinite past of independent normal innovations, $y_0^\dagger = \sum_{s=0}^{\infty}(1+\alpha)^s \sigma\varepsilon_{-s}$. The stationary solution to (3.6) is then of the form

$$y_t^\dagger = \sum_{s=0}^{t-1}(1+\alpha)^s \sigma\varepsilon_{t-s} + (1+\alpha)^t y_0^\dagger = \sum_{s=0}^{\infty}(1+\alpha)^s \sigma\varepsilon_{t-s}. \tag{3.8}$$

In the asymptotic analysis, we consider sums of data. For instance, we can rewrite the normalized average using the above solutions (3.7) and (3.8) as

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{ y_t^\dagger + \left(y_t - y_t^\dagger\right)\right\} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t^\dagger + \frac{y_0 - y_0^\dagger}{\sqrt{T}}\sum_{t=1}^{T}(1+\alpha)^t. \end{aligned}\tag{3.9}$$

The first term is asymptotically normal for large T [37, theorem 20.1]. For the second term, the geometric sum converges, while $y_0^\dagger$ is bounded in probability. Thus, the entire second term vanishes if also the initial observations y_0 is bounded. If $y_0/\sqrt{T}$ converges, we get a non standard limit distribution. However, if $y_0/\sqrt{T}$ diverges, the normalized average also diverges.

In the I(1) case, $\alpha = 0$, the initial observation has a similar influence. The solution (3.7) reduces to

$$y_t = \sum_{s=1}^{t}\sigma\varepsilon_s + y_0. \tag{3.10}$$

In particular, for any $0 \le u \le 1$ and using the floor function $\lfloor\cdot\rfloor$, we get

$$\frac{1}{\sqrt{T}} y_{\lfloor Tu\rfloor} = \frac{1}{\sqrt{T}}\sum_{s=1}^{\lfloor Tu\rfloor}\sigma\varepsilon_s + \frac{y_0}{\sqrt{T}}. \tag{3.11}$$

If the innovations are independent normal, the first term converges to a Brownian motion as a process in u [37, theorem 16.1]. Again, the second term vanishes if the initial observations y_0 is bounded. A non-standard limit arises when $y_0/\sqrt{T}$ converges. And, when $y_0/\sqrt{T}$ diverges, the normalized average also diverge.

In Section 4, we present a model for good and outlying innovations. Episodes of good innovations will be interspersed by one or more outlying innovations. The outlying innovations combine into the initial value for the following good episode. In analogy with the above derivations, a few large outliers or many smaller outliers may combine into large initial values that can influence the asymptotic theory. Thus, the initialization of the good episodes will be a key ingredient for the asymptotics.

In cointegration analysis, the Granger–Johansen representation of Johansen [25, theorem 4.2] has that I(0) terms start in a stationary distribution and therefore combines the ideas behind (3.8), (3.10). For outlier analysis, we will need a representation of the form (3.7). This is derived in the following subsection.

3.3 | A New Granger–Johansen Representation

We need a Granger–Johansen representation to show how the time series depends on the initial observations. The proof generalizes univariate ideas in [38]. Unless explicitely stated, the results make no assumptions to the innovation distribution.

Recall the unobserved components formulation $\boldsymbol{x}_t = \boldsymbol{x}_t^* + \boldsymbol{\tau}_c$ from (3.1), where the dynamic part $\boldsymbol{x}_t^*$ satisfies the VAR in (3.2). Define the companion vectors

$$\boldsymbol{y}_{t-1}^* = \begin{pmatrix}\boldsymbol{\beta}'\boldsymbol{x}_{t-1}^* \\ \Delta\boldsymbol{x}_{t-1}^* \\ \vdots \\ \Delta\boldsymbol{x}_{t-k+1}^*\end{pmatrix}, \quad \widetilde{\boldsymbol{y}}_{t-1}^* = \begin{pmatrix}\Delta\boldsymbol{x}_t^* \\ \boldsymbol{y}_{t-1}^*\end{pmatrix},$$

$$\overline{\boldsymbol{y}}_{t-1}^* = \begin{pmatrix}\Delta z_t^* \\ \boldsymbol{y}_{t-1}^*\end{pmatrix} = \left(0, I_{\dim\overline{y}}\right)\widetilde{\boldsymbol{y}}_{t-1}^*, \tag{3.12}$$

where $\dim\overline{\boldsymbol{y}}_t^* = \dim\widetilde{\boldsymbol{y}}_t^* - 1 = \dim\boldsymbol{y}_t^* + p - 1 = r + kp - 1$. The model Equation (3.2) implies

$$\boldsymbol{y}_t^* = \boldsymbol{Y}\boldsymbol{y}_{t-1}^* + \boldsymbol{e}_{y^*}\boldsymbol{A}\boldsymbol{\varepsilon}_t, \tag{3.13}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{I}_r + \boldsymbol{\beta}'\boldsymbol{\alpha} & \boldsymbol{\beta}'\boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\beta}'\boldsymbol{\Gamma}_{k-2} & \boldsymbol{\beta}'\boldsymbol{\Gamma}_{k-1} \\ \boldsymbol{\alpha} & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{k-2} & \boldsymbol{\Gamma}_{k-1} \\ 0 & \boldsymbol{I}_p & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \boldsymbol{I}_p & 0 \end{pmatrix}, \qquad \boldsymbol{e}_{y^*} = \begin{pmatrix}\boldsymbol{\beta}' \\ \boldsymbol{I}_p \\ 0 \\ \vdots \\ 0\end{pmatrix}.$$

We assume that $\boldsymbol{Y}$ has eigenvalues with absolute values less than unity. This implies that (3.13) has a stationary solution when the errors are independent normal.

Assumption 1. **Stationarity**. $|\text{eigen}\,(\boldsymbol{Y})| < 1$.

The stationary condition implies, in particular, that the eigenvalues of $\boldsymbol{Y}$ differ from unity. This, in turn, is equivalent to the so-called I(1) condition by [25, theorem 4.2], see [39] for a proof. To express the I(1) condition, suppose $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{p\times r}$ have orthogonal complements $\boldsymbol{\alpha}_\perp, \boldsymbol{\beta}_\perp \in \mathbb{R}^{p\times(p-r)}$ such that, $(\boldsymbol{\beta}, \boldsymbol{\beta}_\perp)$ is invertible and $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = 0$. Let $\boldsymbol{\Psi} = \boldsymbol{I}_p - \sum_{j=1}^{k-1}\boldsymbol{\Gamma}_j$. The I(1) condition is that $\boldsymbol{\alpha}_\perp'\boldsymbol{\Psi}\boldsymbol{\beta}_\perp$ is invertible.

Further, define the common trend impact matrix and a parameter that will be used to describe how the I(1) part of the process $\boldsymbol{x}_t^*$ depends on the I(0) components:

$$\boldsymbol{C} = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}_\perp'\boldsymbol{\Psi}\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}_\perp', \qquad \boldsymbol{\psi}' = -\boldsymbol{\beta}_\perp'\boldsymbol{C}\left(\boldsymbol{\Psi}\overline{\beta}, \boldsymbol{\Gamma}_1, \ldots, \sum_{j=1}^{k-1}\boldsymbol{\Gamma}_j\right). \tag{3.14}$$

The extended series $\widetilde{\boldsymbol{y}}_t^*$ satisfies a VAR with moving average errors (VARMA):

$$\widetilde{\boldsymbol{y}}_t^* = \widetilde{\boldsymbol{Y}}\widetilde{\boldsymbol{y}}_{t-1}^* + \widetilde{\boldsymbol{\varepsilon}}_t^*, \tag{3.15}$$

with

$$\widetilde{\boldsymbol{Y}} = \begin{pmatrix} 0 & \boldsymbol{\omega}' \\ 0 & \boldsymbol{Y} \end{pmatrix}, \qquad \widetilde{\boldsymbol{\varepsilon}}_t^* = \begin{pmatrix} \boldsymbol{A}\boldsymbol{\varepsilon}_t \\ \boldsymbol{e}_{y^*}\boldsymbol{A}\boldsymbol{\varepsilon}_{t-1} \end{pmatrix} \tag{3.16}$$

and where $\boldsymbol{\omega}$ is defined by writing the homogeneous model Equation (3.2) as $\Delta \boldsymbol{x}_t^* = \boldsymbol{\omega}'\boldsymbol{y}_{t-1}^* + \boldsymbol{A}\boldsymbol{\varepsilon}_t$. When $\boldsymbol{Y}$ has absolute eigenvalues less than unity, so does $\widetilde{\boldsymbol{Y}}$.

We now extend the representation of Johansen [25, theorem 4.2]. To support the subsequent analysis, we formulate the representation for indices t satisfying $\underline{t} < t \leq \bar{t}$ with particular attention to the dependence on the initial observation at time $\underline{t}$. This initial observations is a time series aggregation of previous innovations. Later on we will think of the indices $\underline{t} < t \leq \bar{t}$ as good while the observation at time $\underline{t}$ could represent a single outlier or the aggregated effect of a cluster of outliers as discussed in Section 3.2.

Theorem 1. Granger-Johansen representation. *Consider the model Equation (3.2) for t larger than some $\underline{t}$ and suppose Assumption 1.*

a. **I(1) component**. *Define $\boldsymbol{C}, \boldsymbol{v}$ as in (3.14). Then*

$$\boldsymbol{\beta}_\perp'\boldsymbol{x}_t^* = \boldsymbol{\beta}_\perp'\boldsymbol{C}\boldsymbol{A}\sum_{s=\underline{t}+1}^{t}\boldsymbol{\varepsilon}_s + \boldsymbol{\psi}'\boldsymbol{y}_t^* + \boldsymbol{\beta}_\perp'\boldsymbol{x}_{\underline{t}}^* - \boldsymbol{\psi}'\boldsymbol{y}_{\underline{t}}^* \qquad \text{for } \underline{t} < t. \tag{3.17}$$

b. **I(0) component**. *Suppose $\boldsymbol{\varepsilon}_t$ are independent $\mathsf{N}_p(0, \boldsymbol{I}_p)$ for $\underline{t} < t \leq \bar{t} \leq \infty$. Then*
   i. *$\boldsymbol{y}_{\underline{t}}^*$ and $\widetilde{\boldsymbol{y}}_{\underline{t}}^*$ can be given stationary initial distributions.*
   ii. $\min_{\underline{t}+k<t\leq\bar{t}} \min \text{eigen Var}(\widetilde{\boldsymbol{y}}_t^* | \widetilde{\boldsymbol{y}}_s^*, \underline{t} - k < s \leq t - k) > 0.$

As mentioned previously, the identity for the I(1) component uses no distributional assumptions to the VAR errors $\boldsymbol{\varepsilon}_t$. The explicit expressions for the I(0) parts and for the initial observations are consistent with Johansen's implicitly defined expressions.

4 | The Data Generating Process

In this section, we describe the assumptions on the ADL errors, the regressors which are generated by a VAR, and the permitted sequences of data generating processes.

We allow for outliers in both the ADL and the VAR. It will be possible that ADL errors are good while VAR errors are outlying, corresponding to super exogeneity or causal transmission.

4.1 | The ADL Errors

Set of good ADL errors. Let ζ_T be a h set of indices for good observations. Suppose $h/T \to \lambda$ where $1/2 < \lambda \leq 1$.

The good ADL errors are assumed independent standard normal

$$\varepsilon_t \stackrel{\mathsf{D}}{=} \mathsf{IIN}(0, 1) \qquad \text{for } t \in \zeta_T. \tag{4.1}$$

Relaxation of this assumption is discussed in Appendix A.

The outlier ADL errors must be extreme relative to the standard normal ADL good errors. Extreme value theory shows that $\max_{t\in\zeta_T} \varepsilon_t/\sqrt{2\log h} \to 1$ almost surely [40, example 8.13]. We assume

$$|\varepsilon_t| \geq \sqrt{2\log h} \qquad \text{for } t \notin \zeta_T. \tag{4.2}$$

Further, we require independence of

$$\varepsilon_t \qquad \text{and} \qquad \Delta z_{t-s}, \boldsymbol{x}_{t-s-1} \qquad \text{for } t \in \zeta_T \text{ and } s \in \mathbb{N}_0, \tag{4.3}$$

to get a martingale difference structure for the good observations. In Section 4.2, we constrain the magnitude of the outlying VAR errors, which indirectly constrains the ADL errors. There will be no other assumptions to the outlying ADL errors in terms of marginal distribution, dependence structure and relation with the past, current and future observations. The assumptions permit additive and innovative outliers in the sense of [13].

4.2 | The VAR Generating the ADL Regressors

The assumptions to the ADL errors indirectly give a one-dimensional linear constraint to the p-dimensional VAR errors through (3.4). Next, we introduce conditions to the good and outlying VAR errors. The parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{k-1}, \boldsymbol{A}$, and either $\boldsymbol{\tau}_c$ or $\boldsymbol{\mu}_c, \boldsymbol{\tau}_\ell$ do not depend on T.

Set of good VAR errors. Let $\zeta_{VAR,T}$ be a h_{VAR} set of indices of good VAR errors, so that $\zeta_{VAR,T} \subset \zeta_T$ and $h_{VAR} \leq h \leq T$. If $h_{VAR} < h$ super exogeneity or causal transmission may be present. The set $\zeta_{VAR,T}$ is further divided into G episodes of length $h_g = \bar{t}_g - \underline{t}_g$. We require that h_g is non-decreasing in T while G does not depend on T. This condition limits the complexity of the proof, but could potentially be relaxed. Simulations reported in the supplement indicate that this could be the case.

The good and outlying VAR periods are interspersed such that the good episode g starts at $\underline{t}_g + 1$ and ends at $\bar{t}_g$, the next outlier episode runs from $\bar{t}_g + 1$ to $\underline{t}_{g+1}$, and timings satisfy

$$0 = \bar{t}_0 \leq \underline{t}_1 < \bar{t}_1 < \cdots < \underline{t}_g < \bar{t}_g < \cdots < \underline{t}_G < \bar{t}_G \leq \underline{t}_{G+1} = T. \tag{4.4}$$

The good VAR errors are assumed independent normal:

$$\boldsymbol{\varepsilon}_t \stackrel{\mathsf{D}}{=} \mathsf{IIN}_p(0, \boldsymbol{I}_p) \qquad \text{for } t \in \zeta_{VAR,T}.$$

We note that this implies (4.1), through (3.4), as well as

$$\max_{t\in\zeta_{VAR,T}} |\boldsymbol{\varepsilon}_t|^2/(2\log h_{VAR}) \stackrel{a.s.}{=} 1. \tag{4.5}$$

4.3 | Conditions for Boundedness

For the boundedness result, we must check a short list of conditions as shown in Theorem A.1 in the Appendix. The OLS estimator, $\hat{\beta}$ say, in the regression $y_t = \beta x_t + \varepsilon_t$ with scalar y_t, x_t provides some intuition. By the Cauchy-Schwarz inequality, $|\hat{\beta} - \beta|^2 = |\sum_{t=1}^T x_t\varepsilon_t / \sum_{t=1}^T x_t^2|^2 \leq \sum_{t=1}^T \varepsilon_t^2 / \sum_{t=1}^T x_t^2$. When the innovations are i.i.d., we must show that $T^{-1}\sum_{t=1}^T x_t^2$ is not too small. Because of the selection involved in the LTS estimation, we will have a more complicated condition, see Assumption A.1(iv) in the Appendix. Nonetheless, boundedness can be achieved when as many as 1/3 of the observations are outlying and the outliers are rather large.

The outlier errors must be of polynomial order

$$\max_{t \notin \zeta_{VAR,T}} |\boldsymbol{\varepsilon}_t|^2 = \mathrm{O}_{\mathsf{P}}(T^c) \qquad \text{for some } c < \infty. \tag{4.6}$$

Due to the bound (4.5) to the good, normal errors, all errors satisfy

$$\max_{t \le T} |\boldsymbol{\varepsilon}_t|^2 = \mathrm{O}_{\mathsf{P}}(T^c). \tag{4.7}$$

The initial observations $|\boldsymbol{x}_{-\ell}|^2$ for $0 \le \ell < k$ must satisfy the same bound.

Further, the number of good observations is bounded from below as

$$h_{VAR} > 2T/3. \tag{4.8}$$

This assumption will be sufficient to prove boundedness of the LTS estimator $\hat{\beta}$ for the ADL regression that selects h observations, where $h \ge h_{VAR}$. It shows that the LTS estimator remains bounded if up to 1/3 of the observations are outliers.

The present conditions contrast with i.i.d. models with continuous regressors where up to 1/2 of the observations can be outliers. The issue is that LTS looses its robustness if too many regressors are outlying as noted in [41]. In autoregressions, past outlier errors propagate into the regressors, see Remark B.1 in the Appendix for an example.

4.4 | Conditions for Consistent Selection and Expansions

For consistency and asymptotic distribution results we must check the assumptions of Theorem A.2 in the Appendix. Most bindingly, the normalized, infeasible OLS estimator on the good observations should have a standard asymptotic distribution. This requires that the initial observations for each good episode should not be too large as seen in the univariate example in Section 3.2. These initial observations represent the cumulated effect of all outliers in the previous outlier episode both through their number and their magnitude. In practice, outlier innovations appear to be of the same order of magnitude as the largest good innovations. We will adopt this stringent assumption to the magnitude of the outlier innovations in order to maximize the possible number of outliers.

We strengthen the assumptions, so that the regressors are of the same logarithmic order as the good errors. Since the outlier errors propagate autoregressively into the regressors, we now require

$$\max_{t \notin \zeta_{VAR,T}} |\boldsymbol{\varepsilon}_t|^2 = \mathrm{O}_{\mathsf{P}}(2 \log h). \tag{4.9}$$

Thus, in light of (4.5) all errors satisfy

$$\max_{t \le T} |\boldsymbol{\varepsilon}_t|^2 = \mathrm{O}_{\mathsf{P}}(2 \log h). \tag{4.10}$$

The initial observations $|\boldsymbol{x}_{-\ell}|^2$ for $0 \le \ell < k$ are assumed $\mathrm{O}_{\mathsf{P}}(\log T)$.

We will also need to bound the intermediate quantiles of the regressors such that regressors do not concentrate too much. Let the squared (Euclidean) norm of the vector of $\widetilde{\boldsymbol{y}}_t^*$ and $\boldsymbol{\beta}'_{\perp} \boldsymbol{x}_t^* / \sqrt{T}$ have increasing order statistics $q_1 \le \cdots \le q_T$ satisfying

$$\forall 0 < \delta < 1, \exists 0 < r < 1 : \; q_{T - \lfloor T^r \rfloor} / q_T \le \delta \{1 + \mathrm{o}_{\mathsf{P}}(1)\}. \tag{4.11}$$

The maximum of a normal I(1) series is of the same order as the series itself, whereas the maximum of the norm of a stationary, normal VAR diverges, but satisfies the condition (4.11) [42]. Therefore, the condition pertains to the behaviour of the process during outliers episodes.

Finally, we strengthen the lower bound to the number of good observations to

$$h_{VAR} = T - \mathrm{o}(\sqrt{T} / \log T). \tag{4.12}$$

In Section 3.2, we saw that initial values that are $\mathrm{o}_{\mathsf{P}}(T^{1/2})$ do not matter for asymptotic distributions. This is the case here: Given (4.12), (4.10), there are at most $T - h_{VAR} = \mathrm{o}(\sqrt{T}/\log T)$ outliers, each of which is of magnitude $\mathrm{O}_{\mathsf{P}}\{(2 \log h)^{1/2}\}$. The maximal combined effect is the product $\mathrm{o}(\sqrt{T}/\log T)\mathrm{O}_{\mathsf{P}}\{(2 \log h)^{1/2}\} = \mathrm{o}_{\mathsf{P}}(T^{1/2})$.

At first glance, the assumptions (4.9–4.12) may appear restrictive, yet they do cover many situations seen in practice. The bound (4.9) allows outlier errors that are a multiple of the largest good errors. The condition (4.12) permits infinitely many outliers as long as the proportion of outliers shrinks at the indicated rate. The simulation study indicates that this may not be too restrictive in finite samples. The condition (4.12) does however rule out unmodelled level shifts interpreted as a proportion of outliers.

5 | Asymptotic Results for LTS Applied to an ADL

We now consider the asymptotic theory for LTS applied to an ADL. This rests on the general LTS theory from Berenguer-Rico and Nielsen [21], which is summarized in Appendix A. First, we provide a boundedness result. Second, we study consistent selection and asymptotic expansions. Third, we discuss asymptotic distributions for some inferential procedures of interest.

5.1 | Boundedness

We now show boundedness of the LTS estimator $\hat{\beta}$ for the ADL Equation (2.2). As the LTS estimator may not be unique, we let $\mathcal{M}_T$ denote the set of minimizers ζ of $\hat{\sigma}^2_{\zeta}$.

Theorem 2. **Boundedness**. *Consider the setup in Sections 4.1, 4.2, 4.3. Then, the LTS estimator $\hat{\beta}$ for the ADL model (2.2) is* **bounded**: $\max_{\zeta \in \mathcal{M}_T} |\hat{\beta}_{\zeta} - \hat{\beta}_{\zeta_T}| = \mathrm{O}_{\mathsf{P}}(1)$.

For boundedness, nearly a third of the observations can be outliers and the outliers can be rather large as outlined in Section 4.3. The proportion of possible outliers is an asymptotic parallel to the finite sample breakdown point [43] analysed for LTS in [44, section 3.4]. In cross sections the breakdown point is 1/2 with continuous regressors but smaller with discrete regressors

[45]. Here, 1/3 is binding as shown in Remark B.1. For further discussion in the context of M-estimators, see [46].

With Theorem 2 we avoid compactness assumptions for the parameter space.

5.2 | Consistent Selection and Expansions

We give further asymptotic properties for the LTS estimators $\hat{\zeta}, \hat{\beta}, \hat{\sigma}$.

Theorem 3. Consistent selection and expansions. *Consider the setup in Sections 4.1, 4.2, 4.4. Let $\xi_T = \zeta_T$ or $\xi_T = \zeta_{VAR,T}$. Then, the LTS estimators $\hat{\zeta}, \hat{\beta}, \hat{\sigma}$ for the ADL model (2.2) satisfy:*

a. **Consistent selection**: $\forall 0 < \eta < 1$: $\max_{\zeta \in \mathcal{M}_T} \#(\zeta \cap \xi_T^c)/h = \mathrm{O}_\mathsf{P}(h^{\eta-1})$.

b. **Expansion for scale**: $\max_{\zeta \in \mathcal{M}_T} h^{1/2}|\hat{\sigma}^2_\zeta - \hat{\sigma}^2_{\xi_T}| = \mathrm{o}_\mathsf{P}(1)$.

c. **Expansion for regression**:

$$\max_{\zeta \in \mathcal{M}_T} \left| \left(\sum_{i \in \zeta} x_t x_t' \right)^{1/2} (\hat{\beta}_\zeta - \beta) - \left(\sum_{i \in \xi_T} x_t x_t' \right)^{1/2} (\hat{\beta}_{\xi_T} - \beta) \right| = \mathrm{o}_\mathsf{P}(1).$$

The square root matrices are defined through joint diagonalization, see Remark B.2.

For the asymptotic expansion, the number and magnitude of outliers is constrained as outlied in Section 4.4. The number of outliers can still diverge but only as long as the proportion of outliers vanishes. The outliers should be of the same order of magnitude as the large good observations. It is plausible that the result would also apply with fewer, but slightly larger outliers. We expect that such conditions would apply in a range of practical applications.

Theorem 3 gives the oracle property that the LTS estimators $\hat{\beta}, \hat{\sigma}$ have the same asymptotic expansions as the infeasible OLS estimators on the actual set ζ_T of good errors. For an asymptotic distribution theory, we will need to clarify how the propagation of past outlier errors into the regressors matters. This is addressed below.

We note that the polynomial order of the outlier errors required in (4.6) for the boundedness result is here replaced with the logarithmic order in (4.9). This may not be necessary if the number of outliers is restricted further. A case with cointegration and a single outlier of order $\sqrt{T}$ is discussed in [47].

5.3 | Asymptotic Distributions

Theorem 3 shows that the LTS estimators $\hat{\beta}, \hat{\sigma}$ have the same asymptotic distribution as the OLS estimators applied infeasibly to the actual set of good observations. However, as the outliers propagate into the good observations, removing the outliers does not remove their effect fully. Nonetheless, asymptotic inference for the LTS estimators can be applied as if the outliers were completely absent from the data generating process.

The argument for the inferential results is as follows. Lemma B.7 shows that the propagation effect of outliers is asymptotically negligible. In turn, Lemma B.8 shows that limit distributions can be expressed in terms of normal distributions for the I(0) parts and Dickey Fuller type distributions for the I(1) parts matching those from standard models where outliers are absent.

For the ADL model (2.1), the inference results include:

1. The hypothesis of no cointegration, $\alpha = 0$, can be tested using Dickey-Fuller type distributions with a LR-statistic on the good observations selected by LTS [28], possibly with level or trend breaks [48], or t-statistics as in [29, 30].
2. Hypotheses on the cointegration parameter κ can be tested using standard normal inference on the good observations selected by LTS [27].

It should be noted that those results require weak exogeneity. For the autoregressive model (3.6), the unit root hypothesis $\alpha = 0$ can be investigated by

3. The Dickey and Fuller [49, 50] t or F test on the good observations selected by LTS. The model can be augmented with lags and deterministic terms.

Further,

4. A lag-length restriction, such as $\gamma_{k-1} = 0$ in (2.1), can be tested using normal inference on the good observations selected by LTS, see [32] for a derivation.

5.4 | Remarks on Stationary Regressions

The above asymptotic theory extends to regressions with stationary regressors. Suppose, we apply the LTS estimator to the regression

$$y_t = \beta' x_t + \sigma \varepsilon_t, \tag{5.1}$$

where x_t may include a constant and/or a linear trend, while its remaining components are generated by a stationary VAR. The theory developed in Appendix B applies in this situation. The proofs do not require a particular cointegration rank and we can simply ignore parts pertaining to the I(1) components. In more detail, the theory can be applied as follows. *First*, the Granger–Johansen representation in Theorem 1 applies with an empty I(1) component. *Second*, the boundedness result in Theorem 2 applies. For this, it is required in (4.8) that the number of good observations satisfies $h \geq h_{VAR} > 2T/3$. *Third*, the asymptotic expansion in Theorem 3 applies with the additional condition (4.12) that the proportion of outliers vanishes.

6 | Simulations

Assumption (4.12) requires that the number of outliers grows to infinity at a modest rate. We use simulations to investigate how binding this assumption is in finite samples. We find that for some simulation designs more outliers can be tolerated, whereas for other simulation designs it appears to be binding. Here, we focus on the t-tests for the cointegrating coefficient and LR-tests for the hypothesis of no cointegration. Tests on other parameters and variations of the simulation design are reported in Supporting Information. The code was written in Matlab with LTS estimation done using the `libra` code [51].

We consider data generating processes of the form

$$\Delta y_t = \omega \Delta z_t + \alpha(y_{t-1} - \kappa z_{t-1} - \nu) + \sigma_\varepsilon \varepsilon_t, \tag{6.1}$$

$$\Delta z_t = \sigma_\eta \eta_t, \tag{6.2}$$

where $\varepsilon_1, \ldots, \varepsilon_T, \eta_1, \ldots, \eta_T$ are independent. We vary α and κ, but set $\omega = 0.5$, $\sigma_\varepsilon = \sigma_\eta = 1$ and $z_0 = 0$. We let $y_0 = \nu$ when $\alpha \neq 0$ and $y_0 = 0$ when $\alpha = 0$. We study the performance of the indicated tests at a 5% level and computed using OLS and LTS with known h. The number of repetitions is 10^4 giving a Monte Carlo standard deviation of 0.002 for 5% tests.

We do not consider the effect of estimating h. Various procedures have been suggested, but no complete asymptotic theory is available as yet. Berenguer-Rico et al. [20] suggested a method on evaluating normal cumulants for different values of h. In a small simulation study they showed that the proportion of outliers may be consistently estimated. For now, developing an asymptotic theory for LTS using an estimated value of h remains work in progress and we will not persue that here.

6.1 | Inference Under Cointegration

We consider the t-test on the cointegrating parameter κ in the data generating process (6.1), (6.2) as follows. We set $\kappa = 1$ and use either $\alpha = -1$ or $\alpha = -0.2$.

It is convenient to let $\psi = -\alpha\kappa$ and $\mu = -\alpha\nu$ and rewrite model (6.1) as

$$\Delta y_t = \omega \Delta z_t + \alpha y_{t-1} + \psi z_{t-1} + \mu + \sigma_\varepsilon \varepsilon_t. \tag{6.3}$$

We estimate $\theta = (\omega, \alpha, \psi, \mu)$ by regressing Δy_t on $x_t = (\Delta z_t, y_{t-1}, z_{t-1}, 1)$ giving $\hat\theta_s = (\hat\omega_s, \hat\alpha_s, \hat\psi_s, \hat\mu_s)'$ for $s \in \{OLS, LTS\}$. The estimates s_s^2 for σ_ε^2 are degrees of freedom corrected. We test the hypothesis $\mathsf{H}_0 : \kappa = 1$ indirectly using $t_{\kappa,s} = (\hat\kappa_s - 1)/\text{s.e.}(\hat\kappa_s)$ where the standard errors $\text{s.e.}(\hat\kappa_s) = \text{s.e.}(\hat\psi_s/\hat\alpha_s)$ vary with s and are obtained using the δ-method. To that end, let $D = \partial\kappa/\partial\theta$ be the 4-vector of partial derivatives of $\kappa = -\psi/\alpha$ with respect to $\theta = (\omega, \alpha, \psi, \mu)$. Define $\hat D_s$ as the vectors D evaluated at the estimator s. Let $M_{OLS} = \sum_{i=1}^n x_t x_t'$ and $M_{LTS} = \sum_{i \in \hat\zeta} x_t x_t'$. Then, we get $\text{s.e.}(\hat\kappa_s) = (s_s^2 \hat D_s' M_s^{-1} \hat D_s)^{1/2}$.

We vary the sample size and the magnitude of the outliers as follows. Let ζ_T indicate the good observations while ζ_T^c indicates the outliers. For $t \in \zeta_T$, let $\eta_t, \varepsilon_t \sim i.i.d.\mathsf{N}(0,1)$. For $s \notin \zeta_T$, let $\eta_s = \sqrt{2\log h} + \xi_{\eta s}$ while $\varepsilon_s = \sqrt{2\log h} + \xi_{\varepsilon s} + 10$ where $\xi_{\eta s}$ and $\xi_{\varepsilon s}$ are i.i.d. standard uniform.

The number of outliers, $T - h$, varies as $\sqrt{T}/2$, $\sqrt{T}$ and $2\sqrt{T}$. This is more than $o(\sqrt{T}/\log T)$ in (4.12), so that we can explore the boundaries for validity of standard inference. For the same reason, we investigate both small and rather large samples.

Figure 1 shows examples of data generating processes with $T = 100$ observations. Variables y, z are shown in rows 1 and 3, while $y - z$ are shown in rows 2 and 4. The adjustment parameter α is -1 in the upper two rows and -0.2 in the lower two rows. Outliers are generated the same way in each column. Column 1 has $\sqrt{T}/2 = 5$ outliers in 5 episodes, such that observations 20, 40, 60, 80, 100 are outliers. Column 2 has $2\sqrt{T} = 20$ outliers in 5 episodes, such that observations 17–20, 37–40, 57–60, 77–80, 97–100 are outliers. Column 3 has $\sqrt{T}/2 = 5$ outliers in a central episode, such that observations 49-53 are outliers. Column 4 has

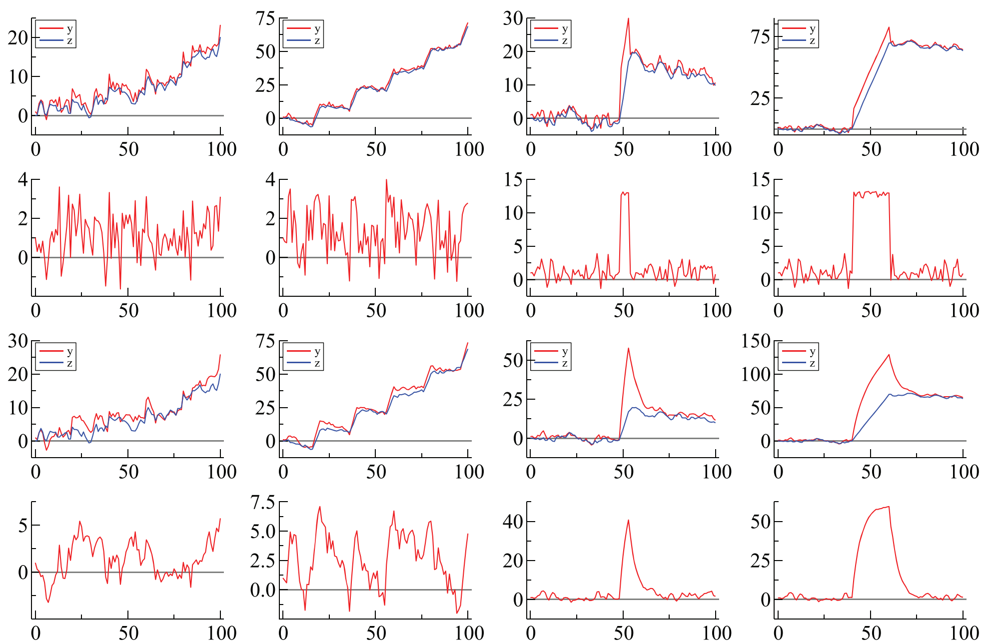

FIGURE 1 | Examples of data generating processes. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 | Outliers distributed in 5 episodes.

Method	$T-h$	$\sqrt{T}/2$	$\sqrt{T}$	$2\sqrt{T}$	$\sqrt{T}/2$	$\sqrt{T}$	$2\sqrt{T}$
	T		$\alpha = -1$			$\alpha = -0.2$	
OLS	25		0.056	0.047		0.488	0.312
	100	0.023	0.011	0.009	0.088	0.050	0.030
	400	0.002	0.000	0.000	0.004	0.000	0.000
	1600	0.005	0.000	0.000	0.001	0.000	0.000
	6400	0.055	0.039	0.010	0.002	0.000	0.000
LTS	25		0.070	0.347		0.108	0.365
	100	0.059	0.058	0.045	0.067	0.063	0.056
	400	0.051	0.052	0.055	0.054	0.053	0.054
	1600	0.053	0.049	0.051	0.055	0.049	0.052
	6400	0.047	0.049	0.049	0.048	0.049	0.049

TABLE 2 | Outliers in one central episode.

Method	$T-h$	$\sqrt{T}/2$	$\sqrt{T}$	$2\sqrt{T}$	$\sqrt{T}/2$	$\sqrt{T}$	$2\sqrt{T}$
	T		$\alpha = -1$			$\alpha = -0.2$	
OLS	25	0.034	0.034	0.055	0.052	0.015	0.005
	100	0.013	0.005	0.006	0.013	0.060	0.313
	400	0.003	0.000	0.000	0.013	0.028	0.113
	1600	0.006	0.000	0.000	0.000	0.003	0.378
	6400	0.027	0.001	0.000	0.000	0.004	0.049
LTS	25	0.071	0.076	0.362	0.089	0.082	0.665
	100	0.057	0.053	0.076	0.063	0.057	0.364
	400	0.056	0.054	0.055	0.058	0.057	0.132
	1600	0.049	0.053	0.049	0.052	0.050	0.029
	6400	0.045	0.052	0.051	0.047	0.000	0.047

$2\sqrt{T} = 20$ outliers in a central episode, such that observations 41–60 are outliers.

In Table 1, the $T-h$ outliers occur in $G = 5$ episodes and in both ε_t and η_t. In each episode there are $\lfloor (T-h)/G \rfloor$ outliers. Outlier episodes are equally spaced by $\lfloor h/G \rfloor$ good observations. Specifically, the system starts with $\lfloor h/G \rfloor$ good observations, after which there is an episode with $\lfloor (T-h)/G \rfloor$ outliers. This is followed by another $\lfloor h/G \rfloor$ good observations, after which another episode with $\lfloor (T-h)/G \rfloor$ outliers follows. This repeats for $G = 5$ episodes in the sample. We find that OLS inference is misleading. LTS performs quite well except for $T = 25$ with $2\sqrt{T} = 10 > T/3$ outliers, so that the boundedness condition (4.8) fails.

In Table 2, the outliers are located in the middle of the sample so that

$$\zeta_h^c = \{\lceil h/2 \rceil + 1, \lceil h/2 \rceil + 2, \dots, \lceil h/2 \rceil + (T-h)\},$$

where $\lceil \cdot \rceil$ denotes the ceiling function. For z_t, the cumulated effect of these outliers is a level shift of magnitude $(T-h)\sqrt{2 \log h}$.

Again, OLS performs poorly. LTS is not quite as good as before. The performance is good with less persistence, $\alpha = -1$, apart from when $T-h = 2\sqrt{T}$ with $T = 25$. With more persistence, $\alpha = -0.2$, LTS works well for $T-h = \sqrt{T}/2$ and for small values of $\sqrt{T}$ but breaks down otherwise.

Overall, the simulations support the validity of the asymptotic theory when the number of outliers is $o(\sqrt{T}/\log T)$ as required in (4.12). The conclusions from the asymptotic theory also appear to be valid when the number of outliers is $\sqrt{T}/2$, but not necessarily with a larger number of outliers and in particular not if the outliers are very concentrated.

6.2 | Power of Test on Cointegrating Parameter

We now study the power of the test on the cointegrating parameter κ in (6.1), (6.2) by testing for $\kappa = 1$ as before, but setting $\kappa = 1 + \delta$ in the data generating processes. Otherwise, the setup is the same as in the size simulations reported in Tables 1, 2. We only consider the case with feedback coefficient $\alpha = -0.2$ and $\sqrt{T}/2$ outliers as in the fourth columns of those tables. Thus, Table 3 has the outliers in five episodes, whereas Table 4 has the outliers

TABLE 3 | Outliers in 5 episodes. Power. $\kappa = 1 + \delta$.

Method	T	$\delta = 1/64$	1/32	1/16	1/8	1/4	1/2
OLS	100	0.106	0.129	0.183	0.324	0.597	0.855
	400	0.014	0.053	0.276	0.710	0.935	0.997
	1600	0.042	0.437	0.805	0.953	0.998	1.000
	6400	0.697	0.910	0.983	1.000	1.000	1.000
LTS	100	0.073	0.087	0.150	0.349	0.704	0.942
	400	0.146	0.375	0.766	0.961	0.999	1.000
	1600	0.798	0.974	1.000	1.000	1.000	1.000
	6400	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 4 | Outliers in one central episode. Power. $\kappa = 1 + \delta$.

Method	T	$\delta = 1/64$	1/32	1/16	1/8	1/4	1/2
OLS	100	0.011	0.008	0.005	0.002	0.002	0.030
	400	0.007	0.004	0.001	0.000	0.002	0.066
	1600	0.000	0.000	0.000	0.000	0.030	0.656
	6400	0.000	0.001	0.066	0.250	0.311	0.886
LTS	100	0.077	0.112	0.232	0.572	0.919	0.996
	400	0.248	0.667	0.969	1.000	1.000	1.000
	1600	0.986	1.000	1.000	1.000	1.000	1.000
	6400	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 5 | Testing for no cointegration with break inducing outliers.

T	Central episode			5 distinct episodes		
	OLS	*LTS*	*IOLS*	*OLS*	*LTS*	*IOLS*
25	0.058	0.018	0.018	0.100	0.028	0.024
100	0.258	0.013	0.013	0.042	0.015	0.015
400	0.453	0.010	0.010	0.170	0.013	0.013
1600	0.794	0.011	0.011	0.659	0.012	0.012
6400	0.974	0.010	0.010	0.967	0.013	0.013

in the middle of the sample. As before, we consider tests based on full sample OLS and on LTS.

The reported rejection frequencies use the asymptotic critical values as it is unclear how to size correct the OLS test. On this measure the LTS-test has better power. This is perhaps more meaningful for the LTS-test, which typically has good size control as seen in Tables 1, 2.

The OLS test has better power properties with 5 episodes in Table 3 than with one central episode in Table 4. In Table 3, OLS is slower to reach high power than the LTS test. This is compensated by the low sizes reported in Table 1 even if size is poorly controlled. In Table 4, OLS struggles to reach the very low sizes reported in Table 2.

6.3 | Testing for No Cointegration

We now consider the test for no cointegration by testing for $\alpha = 0$ in (6.1), (6.2). This restriction implies $\psi = -\alpha\kappa = 0$ and $\mu = -\alpha\nu = 0$ in (6.3), so that the regressors y_{t-1}, z_{t-1} and the intercept are excluded in the restricted model. We use the Likelihood Ratio (LR) test of [28].

It is well known that intercepts play a very different role in autoregressions with a unit root and with a stationary role. This lead [50] to restrict the coefficient to the intercept when imposing the unit root. The present cointegration test takes this effect into account. Now, this interplay between autoregressive roots and deterministic terms carries over to outliers. The effect of a large positive outlier or of many positive outliers vanishes in the stationary case, but gives a level shift in the unit root case. Also in the case of unit root testing, the location of outliers may matter [10]. We will explore some of these effects.

In Table 5, the outliers cumulate to a break under the hypothesis of no cointegration. We consider the two scenarios where outliers occur in either (i) one central episode or (ii) five distinct episodes. The outliers are generated as in Section 6.1. The

TABLE 6 | Testing for no cointegration when outliers do not induce break.

T	Central episode			5 distinct episodes		
	OLS	LTS	$IOLS$	OLS	LTS	$IOLS$
25	0.333	0.201	0.066	0.197	0.927	0.080
100	0.016	0.117	0.054	0.756	0.050	0.050
400	0.050	0.050	0.050	0.716	0.054	0.054
1600	0.038	0.053	0.053	0.596	0.054	0.054
6400	0.035	0.052	0.052	0.401	0.049	0.049

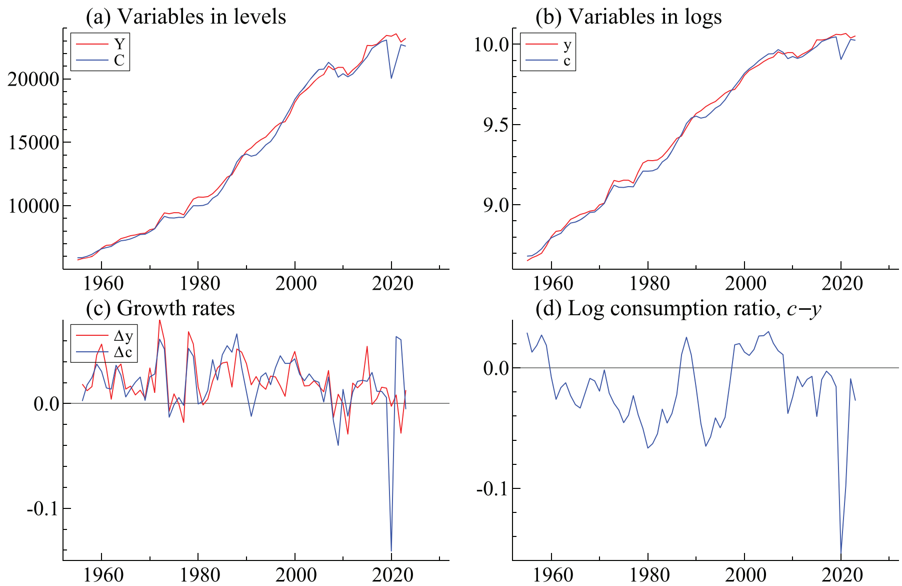

FIGURE 2 | UK individual income and consumption expenditure. [Colour figure can be viewed at wileyonlinelibrary.com]

number of outliers is $2 \times round(\sqrt{T}/4)$ in scenario (i) and $5 \times round(\sqrt{T}/10)$ in scenario (ii). Since there is no cointegration, these outliers induce a level shift in the time series y_t.

The induced level shift is non-negligible as the number of outliers is higher than permitted by the theory. Thus, the standard [28] limit distribution is not applicable. One should develop theory along the lines of [48], but the particular limit distribution is not of general interest as users may prefer modelling the level shifts in this situation. Rather, we can test the oracle property in the simulation by including tests based on OLS applied infeasibly to the actual good observations (IOLS). We see that the OLS test has poor size control, whereas the LTS test matches the IOLS test very well. Thus, the oracle property for LTS in Theorem 3 appears to apply more widely than its assumptions indicate.

In Table 6 the outliers are constructed so as not to cumulate to a break. This is done as follows. In scenario (i), the first half of the outliers have $\eta_s = \sqrt{2\log h} + \xi_{\eta s}$ and $\varepsilon_s = \sqrt{2\log h} + \xi_{\varepsilon s} + 10$, whereas the second half of the outliers have $\eta_s = -(\sqrt{2\log h} + \xi_{\eta s})$ and $\varepsilon_s = -(\sqrt{2\log h} + \xi_{\varepsilon s} + 10)$. In scenario (ii), outliers come in pairs with a positive outlier followed by a negative outlier.

The IOLS columns indicate that the [28] limit distribution is now correct. The LTS based tests performs well except for small sample sizes. The OLS test performs reasonably well for scenario (i) although the size control is wobbly. OLS does not perform well in scenario (ii).

7 | Empirical Illustration

We illustrate the theory through a consumption function analysis. For simplicity we only consider consumption and income, although it has been argued that changing housing collateral and credit constraints should be taken into account [52]. We use the R [53] package robustbase for LTS estimation and PcGive [54] for other calculations.

Figure 2 shows annual series of individual consumption (C_t) and income (Y_t) for the United Kingdom.[1] Panel (a) shows Y_t and C_t in levels. Panel (b) shows the series in logs (y_t and c_t). The trending, non-stationary behavour of the series is evident with large drops in consumption in the 2009 financial crisis and in the 2020 pandemic. Panel (c) shows the growth rates, which could be I(0). Panel (d) shows the log consumption ratio $\log(C_t/Y_t) = c_t - y_t$. This is a candidate cointegrating relation.

Full sample OLS estimation. We start by fitting a full sample ADL model with two lags and linear trend using OLS. This gives

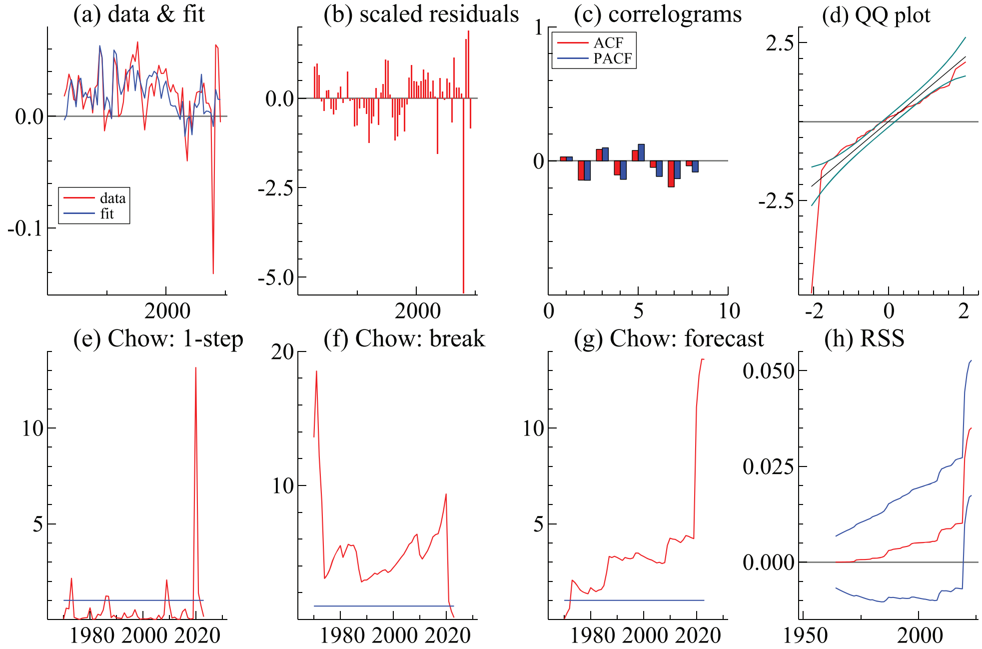

FIGURE 3 | Mis-specification graphics for model estimated by OLS. [Colour figure can be viewed at wileyonlinelibrary.com]

$$\underset{(s.e.)}{\Delta\hat{c}_t} = -\underset{(0.113)}{0.402}c_{t-1} + \underset{(0.142)}{0.460}y_{t-1} - \underset{(0.00130)}{0.0014}\,t$$
$$-\underset{(0.492)}{0.510} + \underset{(0.158)}{0.796}\Delta y_t + \underset{(0.132)}{0.223}\Delta c_{t-1} - \underset{(0.182)}{0.179}\Delta y_{t-1},$$
$$\hat{\sigma} = 0.0242, \quad \hat{\ell} = 158.021, \quad T = 67,$$
$$\mathsf{F}_{ar1-2}(2,58) = 2.46 \ [p = 0.09], \chi^2_{normal}(2) = 39.7 \ [p = 0.00],$$
$$\mathsf{F}_{arch1}(1,65) = 0.22 \ [p = 0.64], \mathsf{F}_{hetero}(12,54) = 1.15 \ [p = 0.34]. \tag{7.1}$$

The OLS estimates are somewhat surprising. The income to consumption ratio is $0.460/0.402 = 1.14$. This is quite a bit larger than unity, perhaps driven by the data during the pandemic. Compensating for this, the slope coefficient is negative. As usual, it is a good idea to check the validity of the model before drawing any inferences.

We check for mis-specification tests using standard output from PcGive. The test statistics are F_{ar1-2} for residual second order autocorrelation [55], F_{arch1} for first order autocorrelated conditional heteroskedasticity [56], χ^2_{normal} for non-normality using the [57] version of the cumulant based test developed in 1880 by Thiele, and F_{hetero} for heteroskedasticity [58]. These papers do not cover the cointegration setting. Cointegration is considered by [32] for F_{ar1-2} and [59] for χ^2_{normal}. The normality test statistic is very extreme, but the other statistics do not reject the model.

Figure 3 gives mis-specification graphics from PcGive. Data and fit are shown in panel (a), scaled residuals in (b), correlograms for residuals in (c), QQ-plot of the quantiles of the residuals versus the quantiles of the fitted normal distribution in (d), three versions of the [60] test in (e,f,g) and recursive residual sum of squares (RSS) in (h). The error bands are pointwise with level of 5% for (d) and 1% for (e,f,g,h). Theory for the cointegration setting is available for the QQ plot [61], the 1-step Chow test [62] and for the RSS plot [63]. We see evidence of big outliers around the pandemic and all recursive tests reject the model strongly.

LTS estimation. We now fit an ADL model using LTS. We set the number of outliers to be $T - h = 5$ and provide evidence in favour of this choice below. LTS finds outliers for 2009, 2020, 2021, 2022, 2023 matching the financial crisis, the pandemic and the Ukranian war. The sample covers many years with many crises. These include the two oilcrises in 1973-74 and 1979 and the 1991 recession. These crises appear to be small relative to the 2009 financial crisis and the pandemic and are not selected by LTS.

The model is now estimated by full sample OLS using 5 impulse indicators, that is,

$$\underset{(s.e.)}{\Delta\hat{c}_t} = -\underset{(0.065)}{0.147}c_{t-1} + \underset{(0.076)}{0.152}y_{t-1} - \underset{(0.0008)}{0.0001}\,t - \underset{(0.295)}{0.035}$$
$$+\underset{(0.086)}{0.630}\Delta y_t + \underset{(0.117)}{0.362}\Delta c_{t-1} - \underset{(0.110)}{0.261}\Delta y_{t-1} - \underset{(0.013)}{0.046}I_{2009}$$
$$-\underset{(0.013)}{0.144}I_{2020} + \underset{(0.022)}{0.082}I_{2021} + \underset{(0.017)}{0.039}I_{2022} - \underset{(0.017)}{0.048}I_{2023},$$
$$\hat{\sigma} = 0.0123, \quad \hat{\ell} = 206.405, \quad T = 67,$$
$$\mathsf{F}_{ar1-2}(2,53) = 1.23 \ [p = 0.23], \chi^2_{normal}(2) = 0.01 \ [p = 0.99],$$
$$\mathsf{F}_{arch1}(1,65) = 1.93 \ [p = 0.17], \mathsf{F}_{hetero}(12,49) = 2.28 \ [p = 0.02]. \tag{7.2}$$

The income to consumption ratio $0.152/0.147 = 1.030$ is now closer to unity.

We apply the same mis-specification tests as before. The previously mentioned papers do not cover the outlier detection. The lag length test is discussed in Section 5.3. Outlier selection is considered by [64] for χ^2_{normal} and [65] for F_{hetero}. The theory suggests that it is plausible that all tests are valid with LTS estimation using the assumptions in Section 4. Figure 4 gives mis-specification graphics. None of the mis-specification tests reject the model.

The theory presented in Section 5.3 shows that the t-statistics formed by dividing coefficients in (7.2) by their standard errors

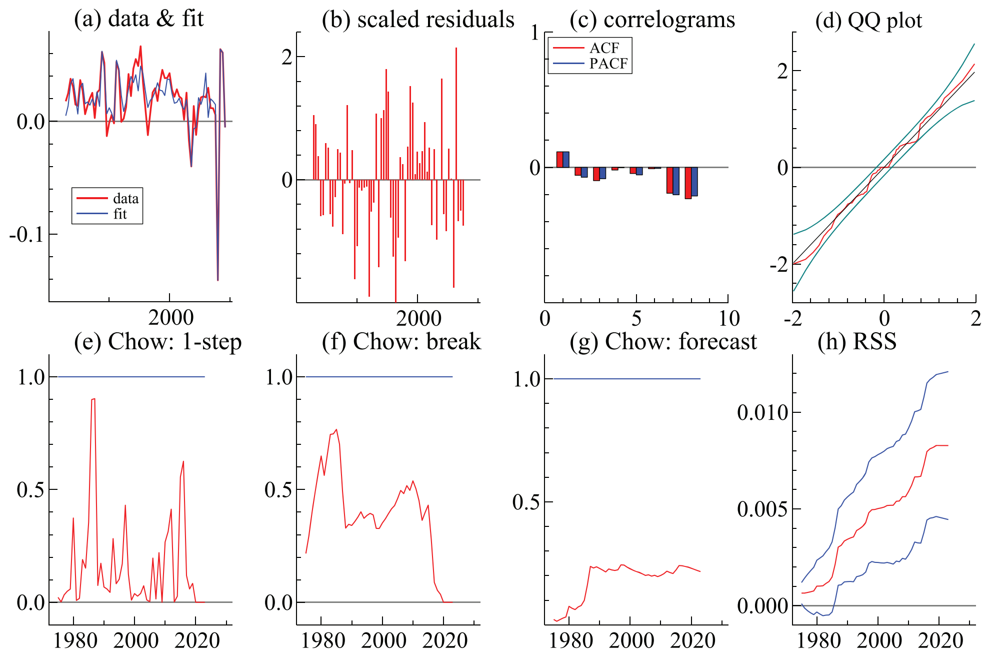

FIGURE 4 | Mis-specification graphics for model estimated by LTS. [Colour figure can be viewed at wileyonlinelibrary.com]

have the same asymptotic distributions as in models without outliers and outlier selection.

The hypothesis of no cointegration is tested using a likelihood ratio statistic for zero coefficient on c_{t-1}, y_{t-1}, t giving LR-statistic $2(206.41 - 203.39) = 6.0$. The 80% critical value is 11.0 [28, table 2], thus indicating absence of cointegration.

Imposing cointegration nonetheless, the equilibrium correction form of the model is

$$\underset{(s.e.)}{\Delta\hat{c}_t} = -\underset{(0.065)}{0.147}\left\{ c_{t-1} - \underset{(0.231)}{1.030} y_{t-1} + \underset{(0.0054)}{0.0005}\, t \right\} + \cdots. \tag{7.3}$$

The t-statistic for homogeneity in the consumption function is $(1 - 1.030)/0.231 = -0.13$, which is not significant when comparing with a standard normal distribution.

Testing weak exogeneity. The above inferences assume weak exogeneity of the income variable. Thus, we consider the following model for income:

$$\begin{aligned}\underset{(s.e.)}{\Delta\hat{y}_t} &= \underset{(0.099)}{0.181}(c_{t-1} - y_{t-1}) - \underset{(0.0001)}{0.0002}\, t + \underset{(0.007)}{0.026} - \underset{(0.167)}{0.024}\Delta y_{t-1} \\ &+ \underset{(0.180)}{0.322}\Delta c_{t-1} - \underset{(0.021)}{0.002} I_{2009} - \underset{(0.020)}{0.013} I_{2020} \\ &+ \underset{(0.033)}{0.070} I_{2021} - \underset{(0.025)}{0.042} I_{2022} - \underset{(0.024)}{0.016} I_{2023}, \\ \hat{\sigma} &= 0.0193, \quad \hat{\ell} = 174.728, \quad T = 67,\end{aligned}$$

$$F_{ar1-2}(2, 55) = 5.19 \;\; [0.01], \; \chi^2_{normal}(2) = 5.88 \;\; [0.05],$$

$$F_{arch1}(1, 65) = 1.71 \;\; [0.20], F_{hetero}(8, 53) = 1.32 \;\; [0.25].$$

Here, estimation is by OLS as the indicator dummies are taken as given and are not necessarily significant. The t-statistic for the cointegrating relation $c_{t-1} - y_{t-1}$ is $0.181/0.099 = 1.83$ which is smaller than, but close to the two-sided 5% critical value from a normal distribution. This is marginal evidence in favour of weak exogeneity.

Estimating the number of good observations. Next, we consider two formal methods for estimating h. It should be pointed out, though, that both methods have incomplete theory and it has not been established yet whether the LTS estimator based on the estimated h has oracle properties. The first method, following [20], estimates the model by LTS for different h, computes normality test statistics and then minimize over h.

Table 7 reports normality test statistics for large values of h. The first, top panel shows h lumping values $38 - 60$ together. The corresponding number of outliers $n - h$ is also shown. The second panel shows normality test statistics T_h, which are minimized for $h = 62$. The value $h = T = 67$ is clearly not attractive. The third panel shows the estimated coefficients for c_{t-1}, y_{t-1} and t, which are relevant for the cointegrating vector. We see a large difference for $h < T$ and for $h = T = 67$ with smaller variation in the first group. The fourth panel shows the identified outliers. We see that the sets of outliers are expanding in a nested way (from the left) when increasing the number of outliers up to 4. Then there is some instability with 2022 declared an outlier for 5 outliers, but not for 6 outliers.

The second method uses Impulse Indicator Saturation (IIS) in PcGive [8]. The non-indicator variables in (7.2) were not selected over. Choosing user defined gauges of 0.5%–0.9% gave four outliers 2009, 2020, 2021, 2023 matching the LTS estimation with 4 outliers. To appreciate the sentitivity on the choice of gauge, we also used some of the default values: a tiny gauge (0.1%) gave three outliers in 2020, 2021, 2023 while a small gauge (1%) gave six outlier in 2009, 2010, 2016, 2020, 2021, 2023, where 2016 is the year of

TABLE 7 | Determining h.

h	38-60	61	62	63	64	65	66	67
$T-h$	29-7	6	5	4	3	2	1	0
T_h	>1.2	2.057	*0.699*	1.427	2.185	4.215	0.901	647.072
α		−0.190	−0.147	−0.198	−0.224	−0.238	−0.299	−0.402
$\alpha\kappa$		−0.195	−0.152	−0.178	−0.197	−0.246	−0.273	−0.460
κ		1.028	1.030	0.897	0.877	1.032	0.913	1.146
t		−0.000	−0.000	0.001	0.001	−0.000	0.001	−0.001
Outliers		2009	2009	2009				
		2010						
		2016						
		2020	2020	2020	2020	2020	2020	
		2021	2021	2021	2021	2021		
			2022					
		2023	2023	2023	2023			

the Brexit referendum. IIS could not be brought to select 5 outliers by changing the gauge.

We decided to go with 5 outliers as suggested by Table 7. The t-statistic for the 2022 dummy is a modest 2.29 in (7.2). Due to the instability in the identified outliers, we also tried a model with 4 outliers giving results of similar quality.

8 | Discussion

We have derived conditions for oracle properties of LTS inference in a cointegrated autoregressive distributed lag (ADL) model. The key assumptions are that outlier errors are more extreme than good errors and that the proportion of outliers, $(T-h)/h$, is asymptotically vanishing. With these assumptions the LTS estimator has the same asymptotic properties as an OLS estimator applied to a model generated from the good errors with absence of any outliers.

The analysis assumes that the number of good observations h is known. In practice, one would want to estimate this number. A number of algorithms are available for this purpose: the index plot method [44], a method based on the normal cumulants [20] and a bootstrap method [66]. Related algorithms include the Forward Search [5] and the Impulse Indicator Saturation implemented as Autometrics in PcGive [8], as Gets in R [9] and in the Eviews software. Some asymptotic theory is available for data generating processes with normal errors and no outliers [7, 10–12]. It would be desirable to develop a theory for selection of h in the presence of outliers.

ADL inference rests on weak exogeneity. This can be a questionable assumption in practice as seen in the empirical illustration. The standard advice is then to use the VAR methods developed by Johansen [24, 25]. We would then need a systems version of Least Trimmed Squares. One approach is to use the Minimum Covariance Determinant approach [67]. Extensions are available to a VAR [16] and to a VAR with different outliers in different equations [68]. Theory for these methods would be desirable.

Funding

Support from Aarhus Center for Econometrics (ACE) funded by the Danish National Foundation grant DNRF186 is gratefully acknowledged. Comments from Jurgen Doornik, Otso Hao and the referees are gratefully acknowledged. This paper is for inclusion in the Oxford Bulletin of Economics & Statistics Special Issue entitled Celebrating 40 Years of Unit Roots and Cointegration.

Endnotes

[1] Household data is from https://www.ons.gov.uk in quarterly accounts for Q3 2024, 2nd release in December 2024. ONS codes CRXX: Real disposable Income per head, current prices. CRYJ: Final consumption expenditure per head, current prices. Data is included in Supporting Information.

References

1. R. J. Ball and T. Burns, "An Econometric Approach to Short-Run Analysis of the U.K. Economy, 1955-66," *Operational Research Quarterly* 19 (1968): 225–256.

2. R. J. Ball, B. D. Boatwright, T. Burns, P. W. M. Lobban, and G. W. Miller, "The London Business School Quarterly Econometric Model of the UK Economy," in *Modelling the Economy*, ed. G. A. Renton (Heinemann Educational Books, 1975), 3–37.

3. J. E. H. Davidson, D. F. Hendry, F. Srba, and S. Yeo, "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal* 88 (1978): 661–692.

4. A. H. Welsh and E. Ronchetti, "A Journey in Single Steps: Robust One-Step M-Estimation in Linear Regression," *Journal of Statistical Planning and Inference* 103 (2002): 287–310.

5. A. Atkinson and M. Riani, *Robust Diagnostic Regression Analysis* (Springer, 2000).

6. M. Riani, "Extensions of the Forward Search to Time Series," *Studies in Nonlinear Dynamics and Econometrics* 8 (2004): 2.

7. D. F. Hendry, S. Johansen, and C. Santos, “Automatic Selection of Indicators in Fully Saturated Regression,” *Computational Statistics* 23 (2008): 317–335.

8. J. A. Doornik, “Autometrics,” in Castle and Shephard [69], (2009), 88–121.

9. F. Pretis, J. Reade, and G. Sucarrat, “Automated General-To-Specific (Gets) Regression Modeling and Indicator Saturation for Outliers and Structural Breaks,” *Journal of Statistical Software* 86 (2018): 1–44.

10. S. Johansen and B. Nielsen, “Saturation by indicators in regression models,” in Castle and Shephard [69], (2009), 1–36.

11. S. Johansen and B. Nielsen, “Analysis of the Forward Search Using Some New Results for Martingales and Empirical Processes,” *Bernoulli* 22 (2016): 1131–1183 Corrigendum (2019) 25, 3201.

12. S. Johansen and B. Nielsen, “Asymptotic Theory of Outlier Detection Algorithms for Linear Time Series Regression Models (With Discussion),” *Scandinavian Journal of Statistics* 43 (2016): 321–381.

13. A. J. Fox, “Outliers in Time Series,” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 34 (1972): 350–363.

14. H. B. Nielsen, “Cointegration Analysis in the Presence of Outliers,” *Econometrics Journal* 7 (2004): 249–271.

15. P. Burridge and A. R. M. Taylor, “Additive Outlier Detection via Extreme-Value Theory,” *Journal of Time Series Analysis* 27 (2006): 685–701.

16. C. Croux and K. Joossens, “Robust Estimation of the Vector Autoregressive Model by a Least Trimmed Squares Procedure,” in *COMPSTAT 2008*, Proceedings in Computational Statistics, ed. P. Brito (Physica Verlag, 2008), 489–501.

17. P. J. Rousseeuw, “Least Median of Squares Regressions,” *Journal of the American Statistical Association* 79 (1984): 871–880.

18. P. J. Huber, “Robust Estimation of a Location Parameter,” *Annals of Mathematical Statistics* 35 (1964): 73–101.

19. J. Á. Víšek, “The Least Trimmed Squares; Part III: Asymptotic Normality,” *Kybernetika* 42 (2006): 203–224.

20. V. Berenguer-Rico, S. Johansen, and B. Nielsen, “A Model Where the Least Trimmed Squares Estimator Is Maximum Likelihood,” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 85 (2023): 886–912.

21. V. Berenguer-Rico and B. Nielsen, “Least Trimmed Squares: Nuisance Parameter Free Asymptotics,” *Econometric Theory* (2024) To appear.

22. R. F. Engle and C. W. J. Granger, “Co-Integration and Error Correction: Representation, Estimation, and Testing,” *Econometrica* 55 (1987): 251–276.

23. C. W. J. Granger, “Developments in the Study of Cointegrated Economic Variables,” *Oxford Bulletin of Economics and Statistics* 48 (1986): 213–228.

24. S. Johansen, “Statistical Analysis of Cointegration Vectors,” *Journal of Economic Dynamics and Control* 12 (1988): 231–254.

25. S. Johansen, *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model* (Oxford University Press, 1995).

26. R. F. Engle, D. F. Hendry, and J. F. Richard, “Exogeneity,” *Econometrica* 51 (1983): 277–304.

27. S. Johansen, “Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis,” *Journal of Econometrics* 52 (1992): 389–402.

28. I. Harbo, S. Johansen, B. Nielsen, and A. Rahbek, “Asymptotic Inference on Cointegrating Rank in Partial Systems,” *Journal of Business & Economic Statistics* 16 (1998): 388–399.

29. A. Banerjee, J. J. Dolado, and R. Mestre, “Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework,” *Journal of Time Series Analysis* 19 (1998): 267–283.

30. N. R. Ericsson and J. G. MacKinnon, “Distributions of Error Correction Tests for Cointegration,” *Econometrics Journal* 5 (2002): 285–318.

31. P. C. B. Phillips, “Reflections on Econometric Methodology,” *Economic Record* 64 (1988): 344–359.

32. B. Nielsen, “Order Determination in General Vector Autoregressions,” in *Time Series and Related Topics: In Memory of Ching-Zong Wei, Volume 52*, Lecture Notes–Monograph Series, ed. H. C. Ho, C. K. Ing, and T. L. Lai (Institute of Mathematical Statistics, 2006), 93–112.

33. P. J. Rousseeuw and K. van Driessen, “An Algorithm for Positive-Breakdown Regression Based on Concentration Steps,” in *Data Analysis: Scientific Modeling and Practical Application*, ed. W. Gaul, O. Opitz, and M. Schader (Springer Verlag, 2000), 335–346.

34. P. J. Rousseeuw and K. van Driessen, “Computing LTS Regression for Large Data Sets,” *Data Mining and Knowledge Discovery* 12 (2006): 29–45.

35. D. F. Hendry and C. Santos, “An Automatic Test of Super Exogeneity,” in *Volatility and Time Series Econometrics*, ed. T. Bollerslev, J. R. Russell, and M. W. Watson (Oxford University Press, 2010).

36. V. Bazinas and B. Nielsen, “Causal Transmission in Reduced-Form Models,” *Econometrics* 10 (2022): 14.

37. P. Billingsley, *Convergence of Probability Measures* (John Wiley & Sons, 1968).

38. B. Nielsen, “The Asymptotic Distribution of Unit Root Tests of Unstable Autoregressive Processes,” *Econometrica* 69 (2001): 211–219.

39. B. Nielsen, “Test for Cointegration Rank in General Vector Autoregressions,” Discussion paper 2009-w10, Nuffield College, 2009.

40. A. DasGupta, *Asymptotic Theory of Statistics and Probability* (Springer, 2008).

41. P. J. Rousseeuw, “Unconventional Features of Positive-Breakdown Estimators,” *Statistics & Probability Letters* 19 (1994): 417–431.

42. V. Watts, H. Rootzén, and M. R. Leadbetter, “On Limiting Distributions of Intermediate Order Statistics From Stationary Sequences,” *Annals of Probability* 10 (1982): 653–662.

43. D. L. Donoho and P. J. Huber, *The Notion of Breakdown Point* (Wadsworth, 1983), 157–184.

44. P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection* (John Wiley & Sons, 1987).

45. O. Hao and B. Nielsen, *Outlier Robust Estimation with Discrete Covariates* (University of Oxford, 2025).

46. J. Klooster and B. Nielsen, “On (in)Consistency of m-Estimators Under Contamination,” Technical report, arXiv:2502.09145v1, 2025.

47. J. A. Doornik, D. F. Hendry, and B. Nielsen, “Inference in Cointegrating Models: UK M1 Revisited,” *Journal of Economic Surveys* 12 (1998): 533–572.

48. T. Kurita and B. Nielsen, “Partial Cointegrated Vector Autoregressive Models With Structural Breaks in Deterministic Terms,” *Econometrics* 7 (2019): 42.

49. D. A. Dickey and W. A. Fuller, “Distribution of the Estimators for Autoregressive Time Series With a Unit Root,” *Journal of the American Statistical Association* 74 (1979): 427–431.

50. D. A. Dickey and W. A. Fuller, “Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root,” *Econometrica* 49 (1981): 1057–1072.

51. S. Verboven and M. Hubert, “Matlab Library Libra,” *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (2010): 509–515.

52. J. Aron, J. V. Duca, J. Muellbauer, K. Murata, and A. Murphy, "Credit, Housing Collateral, and Consumption: Evidence From Japan, the U.K. and the U.S," *Review of Income and Wealth* 58 (2012): 397–423.

53. R Core Team, "R: A Language and Environment for Statistical Computing," (2024).

54. J. A. Doornik and D. F. Hendry, *Empirical Econometric Modelling – PcGive 16, Volume 1* (Timberlake, 2022).

55. L. G. Godfrey, "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica* 46 (1978): 1293–1301.

56. R. F. Engle, "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50 (1982): 987–1108.

57. J. A. Doornik and H. Hansen, "An Omnibus Test for Univariate and Multivariate Normality," *Oxford Bulletin of Economics and Statistics* 70 (2008): 927–939.

58. H. White, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48 (1980): 817–838.

59. L. Kilian and U. Demiroglu, "Residual-Based Tests for Normality in Autoregressions: Asymptotic Theory and Simulation Evidence," *Journal of Business & Economic Statistics* 18 (2000): 40–50.

60. G. C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica* 28 (1960): 591–605.

61. E. Engler and B. Nielsen, "The Empirical Process of Autoregressive Residuals," *Econometrics Journal* 12 (2009): 367–381.

62. B. Nielsen and A. Whitby, "A Joint Chow Test for Structural Instability," *Econometrics* 3 (2015): 156–186.

63. B. Nielsen and J. Sohkanen, "Asymptotic Behaviour of the CUSUM of Squares Test Under Stochastic and Deterministic Time Trends," *Econometric Theory* 27 (2011): 913–927.

64. V. Berenguer-Rico and B. Nielsen, "Normality Testing After Outlier Removal," *Econometrics and Statistics* 38 (2026): 74–97.

65. V. Berenguer-Rico and I. Wilms, "Heteroscedasticity Testing After Outlier Removal," *Econometric Reviews* 40 (2021): 51–85.

66. Q. Heng and K. Lange, "Bootstrap Estimation of the Proportion of Outliers in Robust Regression," *Statistics and Computing* 35 (2025): 3.

67. P. Rousseeuw, "Multivariate Estimation With High Breakdown Point," in *Mathematical Statistics and Applications*, ed. W. Grossmann, G. Pflug, I. Vincze, and W. Wertz (Reidel, 1985), 283–297.

68. J. Raymaekers and P. J. Rousseeuw, "The Cellwise Minimum Covariance Determinant Estimator," *Journal of the American Statistical Association* 119 (2024): 2610–2621.

69. J. L. Castle and N. Shephard, *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry* (Oxford University Press, 2009).

70. S. Johansen and B. Nielsen, "Boundedness of M-Estimators for Multiple Linear Regression in Time Series," *Econometric Theory* 35 (2019): 653–683.

71. N. Chan and C. Wei, "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics* 16 (1988): 367–401.

72. N. Chan, "Asymptotic Inference for Unstable Autoregressive Time Series With Drifts," *Journal of Statistical Planning and Inference* 23 (1989): 301–312.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.

Appendix A

General Asymptotic Theory for LTS

Consider scalar observations y_t and normalized regressors vectors x_{tT} satisfying

$$y_t = x'_{tT}\beta + \sigma\varepsilon_t \qquad \text{for } t = 1, \ldots, T. \tag{A1}$$

The β appearing here and the resulting LTS estimator $\hat{\beta}$ are normalized versions of those appearing in model Equation (2.2) and in Section 2.2. We use the same notation as the distinction only matters in the end of the proof of Theorem 3 in the of this Appendix.

The sequence of data generating processes has common β, σ. For each T there are $h = \lfloor \lambda T \rfloor$ good observations for a common $1/2 < \lambda \leq 1$ and a h index set ζ_T of good observations. The regularity conditions are as follows.

Assumption A.1. Suppose

i. **Frequency of 'good' observations**: $h/T \to \lambda$ where $\lambda > 1/2$.
ii. **'Good' errors** ε_t are independent $\mathsf{N}(0,1)$ for $t \in \zeta_T$.
iii. **'Outlier' errors**: $\min_{t \notin \zeta_T} \varepsilon_t^2 \geq (2\log h)\{1 + \mathrm{o}_{\mathsf{P}}(1)\}$.
iv. **Frequency of regressors near hyperplanes**: Define

$$F_{Th}(a) = \max_{\zeta:\#\zeta=h} \sup_{\delta:|\delta|=1} h^{-1} \sum_{t\in\zeta} 1_{(|x'_{tT}\delta| \leq a)}. \tag{A2}$$

Let ξ satisfy $0 < \xi < 2 - \lambda^{-1}$ and suppose

$$\lim_{(a,T)\to(0,\infty)} \mathsf{P}\{F_{Th}(a) > \xi\} = 0, \tag{A3}$$

that is $\forall \epsilon > 0,\ \exists (a_0, n_0) > 0\colon\ \forall a \leq a_0, T \geq T_0\colon\ \mathsf{P}\{F_{Th}(a) > \xi\} < \epsilon$.

v. **Regressors**: $\|\sum_{i=1}^n x_{tT}x'_{tT}\| = \mathrm{O}_{\mathsf{P}}(T)$.
vi. **Regressors**: Let $|x_{tT}|$ have order statistics $x_{(1)} \leq \cdots \leq x_{(T)}$ satisfying either
   a. $x_{(T)} = \mathrm{O}_{\mathsf{P}}(1)$; or
   b. $x_{(T)}^2 = \mathrm{O}_{\mathsf{P}}(\log T)$ and $\forall 0 < \delta < 1,\ \exists 0 < r < 1\colon\ x^2_{(T-\lfloor T^r \rfloor)}/x^2_{(T)} \leq \delta\{1 + \mathrm{o}_{\mathsf{P}}(1)\}$.
vii. **Infeasible OLS estimator**: $(\hat{\beta}_{\zeta_T} - \beta)'(\sum_{t\in\zeta_T} x_{tT}x'_{tT})(\hat{\beta}_{\zeta_T} - \beta) = \mathrm{O}_{\mathsf{P}}(1)$.

We comment on the assumptions. In (ii), the good errors are normal. This can be relaxed for known h [21], but a distributional assumption seems necessary for estimating h [20]. In (iii) the outlier errors are more extreme than the good errors noting that under normality $\max_{t\in\zeta_T} \varepsilon_t^2/(2\log h) \to 1$ *a.s.* In (iv) the concentration of the regressors is bounded. This implies that $\sum_{t\in\zeta} x_{tT}x'_{tT}$ is invertible for any h-set ζ [70]. Condition (v) has a trade-off with (v), (vi) which limit the magnitude of the regressors.

We quote the general LTS asymptotic theory [21]. As before, let $\mathcal{M}_T$ denote the set of minimizers ζ of $\hat{\sigma}^2_\zeta$.

Theorem A.1. **Boundedness**. *Suppose Assumption A.1*(i, ii, iv). *Then the LTS estimator $\hat{\beta}$ for (A1) is* **bounded**: $\max_{\zeta\in\mathcal{M}_T} |\hat{\beta}_\zeta - \hat{\beta}_{\zeta_n}| = \mathrm{O}_{\mathsf{P}}(1)$.

Theorem A.2. **Consistent selection and expansions**. *Suppose Assumption A.1. Then the LTS estimators $\hat{\zeta}, \hat{\beta}, \hat{\sigma}$ for (A1) satisfy*:

a. **Consistent selection by** $\hat{\zeta}$. $\forall 0 < \beta < 1\colon\ \max_{\zeta\in\mathcal{M}_T} \#(\zeta \cap \zeta_T^c)/h = \mathrm{O}_{\mathsf{P}}(h^{\beta-1})$.

b. **Expansion for** $\hat{\sigma}^2$. $\max_{\zeta\in\mathcal{M}_T} h^{1/2}|\hat{\sigma}^2_\zeta - \hat{\sigma}^2_{\zeta_n}| = \mathrm{o}_{\mathsf{P}}(1)$.

c. **Expansion for $\hat{\beta}$.**

$$\max_{\zeta \in \mathcal{M}_T} \left| \left(\sum_{i \in \zeta} x_{iT} x_{iT}' \right)^{1/2} (\hat{\beta}_\zeta - \beta) - \left(\sum_{i \in \zeta_T} x_{iT} x_{iT}' \right)^{1/2} (\hat{\beta}_{\zeta_T} - \beta) \right| = o_P(1).$$

The square root matrices are defined through joint diagonalization, see Remark B.2.

The conditions (v), (vi) are sufficient for consistent selection and expansions, but not necessary. Berenguer-Rico and Nielsen [21] provide a set of alternative conditions.

Appendix Proofs B

Proof of Representation

Proof of Theorem 1. *Part* (a). Start from the homogeneous model Equation (3.1). Subtract $\sum_{j=1}^{k-1} \boldsymbol{\Gamma}_j \Delta \boldsymbol{x}_t^*$ on both sides and use that $\boldsymbol{\Psi} = I_p - \sum_{j=1}^{k-1} \boldsymbol{\Gamma}_j$ to get

$$\boldsymbol{\Psi} \Delta \boldsymbol{x}_t^* = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-1}^* + \sum_{j=1}^{k-1} \boldsymbol{\Gamma}_j \left(\Delta \boldsymbol{x}_{t-j}^* - \Delta \boldsymbol{x}_t^* \right) + \boldsymbol{A}\boldsymbol{\varepsilon}_t.$$

Insert $\Delta \boldsymbol{x}_{t-j}^* - \Delta \boldsymbol{x}_t^* = -\sum_{s=0}^{j-1} \Delta^2 \boldsymbol{x}_{t-s}^*$ and interchange the two sums to get

$$\boldsymbol{\Psi} \Delta \boldsymbol{x}_t^* = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-1}^* - \sum_{s=0}^{k-2} \left(\sum_{j=s+1}^{k-1} \boldsymbol{\Gamma}_j \right) \Delta^2 \boldsymbol{x}_{t-j}^* + \boldsymbol{A}\boldsymbol{\varepsilon}_t.$$

On the left, pre-multiply $\boldsymbol{x}_t^*$ by the identity $I_p = \boldsymbol{\beta}_\perp \overline{\boldsymbol{\beta}}_\perp' + \overline{\boldsymbol{\beta}}\boldsymbol{\beta}'$ and move the $\boldsymbol{\beta}' \boldsymbol{x}_t^*$ term to the right. Also, pre-multiply both sides by $\boldsymbol{\alpha}_\perp'$. This gives

$$\boldsymbol{\alpha}_\perp' \boldsymbol{\Psi} \boldsymbol{\beta}_\perp \overline{\boldsymbol{\beta}}_\perp' \Delta \boldsymbol{x}_t^* = -\boldsymbol{\alpha}_\perp' \boldsymbol{\Psi} \overline{\boldsymbol{\beta}} \boldsymbol{\beta}' \Delta \boldsymbol{x}_t^* - \boldsymbol{\alpha}_\perp' \sum_{s=0}^{k-2} \left(\sum_{j=s+1}^{k-1} \boldsymbol{\Gamma}_j \right) \Delta^2 \boldsymbol{x}_{t-j}^* + \boldsymbol{\alpha}_\perp' \boldsymbol{A}\boldsymbol{\varepsilon}_t.$$

Pre-multiply by the inverse of $\boldsymbol{\alpha}_\perp' \boldsymbol{\Psi} \boldsymbol{\beta}_\perp$, which exists by Assumption 1. Then pre-multiply by $\boldsymbol{\beta}_\perp' \boldsymbol{\beta}_\perp$. Use the definition $\boldsymbol{C} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}_\perp' \boldsymbol{\Psi} \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}_\perp'$. This gives

$$\boldsymbol{\beta}_\perp' \Delta \boldsymbol{x}_t^* = -\boldsymbol{\beta}_\perp' \boldsymbol{C} \boldsymbol{\Psi} \overline{\boldsymbol{\beta}} \boldsymbol{\beta}' \Delta \boldsymbol{x}_t^* - \boldsymbol{\beta}_\perp' \boldsymbol{C} \sum_{s=0}^{k-2} \left(\sum_{j=s+1}^{k-1} \boldsymbol{\Gamma}_j \right) \Delta^2 \boldsymbol{x}_{t-j}^* + \boldsymbol{\beta}_\perp' \boldsymbol{C} \boldsymbol{A}\boldsymbol{\varepsilon}_t.$$

Using the definition of $\boldsymbol{v}$, we can write this in compact form as

$$\boldsymbol{\beta}_\perp' \Delta \boldsymbol{x}_t^* = \boldsymbol{v} \Delta \boldsymbol{y}_t^* + \boldsymbol{\beta}_\perp' \boldsymbol{C} \boldsymbol{A}\boldsymbol{\varepsilon}_t.$$

Sum over t to get the desired expression.

Part (b, i). The normality assumption and assumption 1 ensure that $\boldsymbol{y}_t^*$ can be given a stationary initial distribution. Now, $\widetilde{\boldsymbol{y}}_t^*$ can also be given a stationary initial distribution as it is a linear function of $\boldsymbol{y}_{t+1}^*, \boldsymbol{y}_t^*$.

Part (b, ii). Rearrange the homogenous model Equation (3.1) as by subtracting the intermediate differences $\boldsymbol{\beta}' \Delta \boldsymbol{x}_{t-j}^*$ from $\boldsymbol{\beta}' \boldsymbol{x}_{t-k}^*$ and defining $\boldsymbol{\Gamma}_j^\dagger = \boldsymbol{\Gamma}_j + \boldsymbol{\alpha}\boldsymbol{\beta}'$ to get

$$\Delta \boldsymbol{x}_t^* = \sum_{j=1}^{k-1} \boldsymbol{\Gamma}_j^\dagger \Delta \boldsymbol{x}_{t-j}^* + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-k}^* + \boldsymbol{A}\boldsymbol{\varepsilon}_t. \tag{B1}$$

This matches the formulation in [24]. In the same vein, let

$$\boldsymbol{y}_t^\dagger = \begin{pmatrix} \Delta \boldsymbol{x}_t^* \\ \vdots \\ \Delta \boldsymbol{x}_{t-k+2}^* \\ \boldsymbol{\beta}' \boldsymbol{x}_{t-k+1}^* \end{pmatrix}, \quad \boldsymbol{Y}^\dagger = \begin{pmatrix} \boldsymbol{\Gamma}_1^\dagger & \cdots & \cdots & \cdots & \boldsymbol{\Gamma}_{k-1}^\dagger & \boldsymbol{\alpha} \\ I_p & 0 & \cdots & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & & \vdots & \vdots \\ & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & I_p & 0 & 0 \\ 0 & \cdots & 0 & 0 & \boldsymbol{\beta}' & I_p \end{pmatrix}, \quad \boldsymbol{e}_p = \begin{pmatrix} I_p \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

such that there is linear, bijective mapping between $\boldsymbol{y}_t^*$ and $\boldsymbol{y}_t^\dagger$ and

$$\boldsymbol{y}_t^\dagger = \boldsymbol{Y}^\dagger \boldsymbol{y}_{t-1}^\dagger + \boldsymbol{e}_p \boldsymbol{A}\boldsymbol{\varepsilon}_t. \tag{B2}$$

Apply the autoregressive equation k times to get

$$\boldsymbol{y}_t^\dagger = \begin{pmatrix} I_p & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & I_p & * \\ 0 & \cdots & 0 & \boldsymbol{\beta}' \end{pmatrix} \begin{pmatrix} \boldsymbol{A}\boldsymbol{\varepsilon}_t \\ \vdots \\ \vdots \\ \boldsymbol{A}\boldsymbol{\varepsilon}_{t-k+1} \end{pmatrix} + \boldsymbol{Y}^{\dagger k} \boldsymbol{y}_{t-k}^\dagger, \tag{B3}$$

where $*$ represents quantities that are not of importance. As $\boldsymbol{\beta}'$ has full row rank by Assumption 1, so does the first matrix in (B3). The vector of errors in (B3) has an invertible covariance matrix whenever $t - k + 1 > \underline{t}$. It is also independent of the σ-algebra $\mathcal{G}_{t-k}$ generated by $\boldsymbol{y}_{\underline{t}-k+1}^\dagger, \ldots, \boldsymbol{y}_{t-k}^\dagger$, while $\boldsymbol{y}_{t-k}^\dagger$ is $\mathcal{G}_{t-k}$ measurable. Therefore, $\mathsf{Var}(\boldsymbol{y}_t^\dagger | \mathcal{G}_{t-k})$ is constant and invertible for $t \geq \underline{t} + k$.

We now concatenate $\boldsymbol{y}_t^\dagger$ with Δx_{t+1}^*. By the model Equation (B1), we have

$$\Delta \boldsymbol{x}_{t+1}^* = \boldsymbol{A}\boldsymbol{\varepsilon}_{t+1} + \sum_{j=1}^{k-1} \boldsymbol{\Gamma}_j^\dagger \Delta \boldsymbol{x}_{t-j+1}^* + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-k+1}^* = \boldsymbol{A}\boldsymbol{\varepsilon}_{t+1} + \boldsymbol{v}_\dagger' \boldsymbol{y}_t^\dagger,$$

for a suitably defined $\boldsymbol{v}_\dagger$. Thus, we have

$$\widetilde{\boldsymbol{y}}_t^\dagger = \begin{pmatrix} \Delta \boldsymbol{x}_{t+1}^* \\ \boldsymbol{y}_t^\dagger \end{pmatrix} = \begin{pmatrix} I_p & \boldsymbol{v}_\dagger' \\ 0 & I_{\dim y^*} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}\boldsymbol{\varepsilon}_{t+1} \\ \boldsymbol{y}_t^\dagger \end{pmatrix}.$$

As $\boldsymbol{A}\boldsymbol{\varepsilon}_{t+1}$ and $\boldsymbol{y}_t^\dagger$ are independent and each has invertible, constant covariance, we get that $\mathsf{Var}(\widetilde{\boldsymbol{y}}_t^\dagger | \mathcal{G}_{t-k})$ is constant and invertible for $t \geq \underline{t} + k$.

Finally, the σ-algebra $\mathcal{G}_{t-k}$ generated by $\boldsymbol{y}_s^\dagger$ for $\underline{t} - k < s \leq t - k$ can equivalently be generated by $\widetilde{\boldsymbol{y}}_s^\dagger$ for $\underline{t} - k < s \leq t - k - 1$ due to the concatenation with Δx_{t+1}^*, or by $\widetilde{\boldsymbol{y}}_s^*$ for $\underline{t} - k < s < t - k$ by a linear, bijective transformation. □

The Normalized Regressor Vector

Assumption A.1 uses a normalized regressor vector x_{tT}, which we define here. The ADL Equation (2.1) with a constant has regressors

$$\Delta z_t, \boldsymbol{x}_{t-1}, \Delta \boldsymbol{x}_{t-1}, \ldots, \Delta \boldsymbol{x}_{t-k+1}, 1, \tag{B4}$$

where $\boldsymbol{x}_t = (y_t, z_t)'$. The unobserved components formulation (3.1) has $\boldsymbol{x}_t = \boldsymbol{x}_t^* + \boldsymbol{\tau}_c$ so that $\Delta \boldsymbol{x}_t = \Delta \boldsymbol{x}_t^*$ and where $\boldsymbol{x}_t^*$ satisfies the VAR in (3.2). Thus, the regressors in (B4) form a bijective, linear function of

$$\Delta z_t, \boldsymbol{x}_{t-1}^*, \Delta \boldsymbol{x}_{t-1}^*, \ldots \Delta \boldsymbol{x}_{t-k+1}^*, 1.$$

Now, $\boldsymbol{x}_{t-1}^*$ is a linear combination of $\boldsymbol{\beta}' \boldsymbol{x}_{t-1}^*$ and $\boldsymbol{\beta}_\perp' \boldsymbol{x}_{t-1}^*$. Concatenate $\boldsymbol{\beta}' \boldsymbol{x}_{t-1}^*, \Delta \boldsymbol{x}_{t-1}^*, \ldots, \Delta \boldsymbol{x}_{t-k+1}^*$ as $\boldsymbol{y}_{t-1}^*$, see (3.12). The Granger–Johansen representation Theorem 1 writes $\boldsymbol{\beta}_\perp' \boldsymbol{x}_{t-1}^*$ as a linear combination of a random walk, $\boldsymbol{y}_{t-1}^*$ and initial values. Normalize the random walk by $\sqrt{T}$. Then the regressors in (B4) are a linear function of

$$x_{tT} = \left(\Delta z_t \ , \ \boldsymbol{y}_{t-1}^{*\prime} \ , \ T^{-1/2} \sum_{s=1}^{t-1} \varepsilon_s' \ , \ 1 \right)', \tag{B5}$$

which has dimension $p - 1 + p + r + (k - 1)p + 1 = (k + 1)p + r$.

When checking parts (iv, v) of Assumption A.1, we extend the vector x_{tT} with $\Delta y_t, t/T$. Since $\Delta y_t, \Delta z_t, \Delta \mathbf{y}^*_{t-1}$ concatenate as $\widetilde{\mathbf{y}}^*_{t-1}$ we get that x_{tT} is a subvector of

$$\widetilde{x}_{tT} = \left(\widetilde{\mathbf{y}}^{*\prime}_{t-1} \ , \ T^{-1/2}\sum_{s=1}^{t-1}\epsilon'_s \ , \ t/T \ , \ 1 \right)', \tag{B6}$$

which has dimension $(k+1)p + r + 2$. We remove Δy_t in martingale arguments:

$$\overline{x}_{tT} = \{0, I_{(k+1)p+r+1}\}\widetilde{x}_{tT}. \tag{B7}$$

Conditions for Boundedness

We check the boundedness Assumption A.1(iv) for x_{tT} defined in (B5). This is a subvector of $\widetilde{x}_{tT}$ in (B6). We link the $F_{Th}(a)$ functions for x_{tT} and $\widetilde{x}_{tT}$.

Lemma B.1. *If x_{tT} is a subvector of $\widetilde{x}_{tT}$ then*

$$F^x_{Th}(a) = \max_{\zeta:\#\zeta=h}\sup_{\delta:|\delta|=1}\frac{1}{h}\sum_{i\in\zeta}1_{(|\delta' x_{iT}|\le a)}$$
$$\le \max_{\zeta:\#\zeta=h}\sup_{\widetilde{\delta}:|\widetilde{\delta}|=1}\frac{1}{h}\sum_{i\in\zeta}1_{(|\widetilde{\delta}'\widetilde{x}_{iT}|\le a)} = F^{\widetilde{x}}_{Th}(a).$$

Proof of Lemma B.1. Write $x_{tT} = s'\widetilde{x}_{tT}$ for a selection matrix s of dimension $\dim\widetilde{x}\times\dim x$ with zero coeffients apart from one unit coefficient in each column. If δ is a unit vector of length $\dim x$, then $\widetilde{\delta} = S\delta$ is a unit vector of length $\dim\widetilde{x}$. Therefore,

$$\sup_{\delta:|\delta|=1}\sum_{i\in\zeta}1_{(|\delta' x_{iT}|\le a)} = \sup_{\widetilde{\delta}=S\delta:|\delta|=1}\sum_{i\in\zeta}1_{(|\widetilde{\delta}'\widetilde{x}_{iT}|\le a)} \le \sup_{\widetilde{\delta}:|\widetilde{\delta}|=1}\sum_{i\in\zeta}1_{(|\widetilde{\delta}'\widetilde{x}_{iT}|\le a)}.$$

Divide by h and take maximum over h-sets ζ to get $F^x_{Th}(a) \le F^{\widetilde{x}}_{Th}(a)$. □

Lemma B.2. *If $x, y \in \mathbb{R}$, $a > 0$ then $|1_{(|x|\le a)} - 1_{(|y|\le a)}| \le 1_{(|x-a|\le|y-x|)} + 1_{(|x+a|\le|y-x|)}$.*

Proof of Lemma B.2. Let $d = |x - y|$. Rewrite the difference of indicators as

$$1_{(|x|\le a)} - 1_{(|y|\le a)} = 1_{(-a\le x\le a)} - 1_{(-a+x-y\le x\le a+x-y)}.$$

This difference is zero outside the sets $-a-d \le x \le -a+d$ and $a-d \le x \le a+d$. On those sets, their difference may be -1, 0 or 1. Hence, the bound applies. □

We give conditions ensuring that the $F^{\widetilde{x}}_{tT}$ function for $\widetilde{x}_{tT}$ vanishes within a good episode indexed by $t = 1, \ldots, T$. We combine the separate arguments for stationary processes, random walks and linear trends in [70].

Lemma B.3. *Let $z_{tT} = (u'_t, v'_t/\sqrt{T}, t/T, 1)'$ where $u_t \in \mathbb{R}^{\dim u}$, $v_t \in \mathbb{R}^{\dim v}$. Let $F^z_{T\delta}(a) = T^{-1}\sum_{t=1}^T 1_{(|z'_{tT}\delta|\le a)}$. Suppose that, for some $q_T \ge 1$ such that $q_T/T \to 0$,*

- *i.* **I(0) component**.
  - *a.* $\forall\epsilon > 0, \exists C > 0$: $\max_{q_T<t\le T}\mathsf{P}(|u_t| > C) \le \epsilon$.
  - *b.* $\exists C > 0$: $\max_{q_T<t\le T}\sup_{\delta_u:|\delta_u|=1}\sup_{v\in\mathbb{R}}\mathsf{P}(|u'_t\delta_u + v| < \alpha) \le \alpha C$,
- *ii.* **I(1) component**.
  - *a.* $\forall\epsilon > 0, \exists C > 0$: $\mathsf{P}(\max_{q_T<t\le T}|v_t/\sqrt{T}| > C) \le \epsilon$.
  - *b.* $\exists C > 0$: $\max_{q_T<t\le T}\sup_{\delta_v:|\delta_v|=1}\sup_{v\in\mathbb{R}}\mathsf{P}(|v'_t\delta_v/\sqrt{t} + v| < \alpha) \le \alpha C$.

Then $\sup_{\delta:|\delta|=1}F^z_{T\delta}(a) = \mathrm{o}_{\mathsf{P}}(1)$ as $(a, T) \to (0, \infty)$.

Proof of Lemma B.3. Truncation. Write $F^z_{T\delta}(a) = N_{T\delta}(a) + Q_{T\delta}(a) + R_{T\delta}(a, 0)$ where $N_{T\delta}(a) = T^{-1}\sum_{t\le q_T}1_{(|z'_{tT}\delta|\le a)} \le q_T/T \to 0$ by assumption, while

$$Q_{T\delta}(a) = \frac{1}{T}\sum_{t>q_T}1_{(|z'_{tT}\delta|\le a, |z_{tT}|>A)},$$

$$R_{T\delta}(\alpha, \mu) = \frac{1}{T}\sum_{t>q_T}1_{(|z'_{tT}\delta-\mu|\le\alpha, |z_{tT}|\le A)},$$

for an $A > 0$ to be chosen. We show that $Q_{T\delta}$ and $R_{T\delta}$ vanish uniformly in δ.

The term Q vanishes. The process $z_{tT} = (u'_t, v'_t/\sqrt{T}, t/T, 1)'$ satisfies

$$|z_{tT}| \le |u_t| + |v_t|/\sqrt{T} + |t/T| + 1 \le |u_t| + \max_{q_T<t\le T}|v_t|/\sqrt{T} + 2$$

by the triangle inequality. Thus, we get the set inclusions, for $A \ge 4$,

$$\left(|z'_{tT}\delta| \le a, |z_{tT}| > A\right) \subset \left(|z_{tT}| > A\right)$$
$$\subset \left(|u_t| > A/4\right) \cup \left(\max_{q_T<t\le T}\left|v_t/\sqrt{T}\right| > A/4\right),$$

uniformly in δ, a. Thus, we can bound, uniformly in δ, a and for $A \ge 4$,

$$Q_{T\delta}(a) \le \widetilde{Q}_T = \frac{1}{T\text{-}q_T}\sum_{t>q_T}1_{(|u_t|>A/4)} + 1_{(\max_{q_T<t\le T}|v_t/\sqrt{T}|>A/4)}.$$

Take supremum and then expectation to bound

$$\mathsf{E}\sup_{\delta}Q_{T\delta}(a) \le \mathsf{E}\widetilde{Q}_T = \max_{q_T<t\le T}\mathsf{P}\left(|u_t| > A/4\right)$$
$$+ \mathsf{P}\left(\max_{q_T<t\le T}\left|v_t/\sqrt{T}\right| > A/4\right).$$

This is small for large A since $u_t = \mathrm{O}_{\mathsf{P}}(1)$ uniformly in t while $\max_{t\le T}\left|v_t/\sqrt{T}\right| = \mathrm{O}_{\mathsf{P}}(1)$ by conditions $(i, a; ii, a)$. Thus, $\forall\epsilon > 0$, $\exists A > 0$ such that $\mathsf{E}\sup_{\delta:|\delta|=1}Q_{T\delta}(a) < \epsilon$.

The term R. We parametrize the unit vector δ as

$$\delta = \left(\delta'_u\cos\psi\cos\phi\cos\theta \ , \ \delta'_v\sin\psi\cos\phi\cos\theta \ , \ \sin\phi\cos\theta \ , \ -\sin\theta\right)', \tag{B8}$$

where $\delta_u \in \mathbb{R}^{\dim u}$, $\delta_v \in \mathbb{R}^{\dim v}$ such that $|\delta_u| = |\delta_v| = 1$ while $0 \le \psi \le \pi/2$ and $|\phi|, |\theta| \le \pi/2$. Initially, we distinquish between $\cos\theta = 0$ and $\cos\theta > 0$.

The case $\cos\theta = 0$. If $\cos\theta = 0$ then $|z'_{tT}\delta| = 1$ so that $R_{T\delta}(a) = 0$ for all $a < 1$.

The case $\cos\theta > 0$. We bound $1/\cos\theta$; we chain over δ and analyse the oscillation term; we remove the truncation; and, finally, we consider three subcases.

Bounding $1/\cos\theta$. Write $z'_{tT}\delta = z'_{tT}\delta_{\theta=0}\cos\theta - \sin\theta$ where $\delta_{\theta=0}$ has the form (B8) with $\theta = 0$, so that $\cos\theta = 1$ and $\sin\theta = 0$. As $|\delta_{\theta=0}| = 1$ then $|z'_{tT}\delta_{\theta=0}| \le |z_{tT}| \le A$. By [70, lemma 3.1], we find for $a \le 1/2$ and $|\theta| \le \pi/2$ that $|-\sin\theta + z'_{tT}\delta_{\theta=0}\cos\theta| \le a$ implies $1/\cos\theta \le 2(1 + |z'_{tT}\delta_{\theta=0}|) \le 2(1 + A)$.

Chaining. The set $|\delta| = 1$ is compact. For $\epsilon > 0$ we make a finite cover with L balls with centers δ_ℓ and radius ϵ. Linear chaining gives

$$\sup_{\delta}R_{T\delta}(a, 0) \le \max_{\ell\le L}\{R_{T\delta_\ell}(a, 0) + \sup_{\delta:|\delta-\delta_\ell|\le\epsilon}$$
$$\left|R_{T\delta}(a, 0) - R_{T\delta_\ell}(a, 0)\right|\}. \tag{B9}$$

Oscillation term in (B9). By the triangle inequality,

$$\left|R_{T\delta}(a, 0) - R_{T\delta_\ell}(a, 0)\right|$$
$$\le \frac{1}{T}\sum_{t>q_T}\left|1_{(|z'_{tT}\delta|\le a, |z_{tT}|\le A)} - 1_{(|z'_{tT}\delta_\ell|\le a, |z_{tT}|\le A)}\right|.$$

Apply the inequality $|1_{(|x|\le a)} - 1_{(|y|\le a)}| \le 1_{(|x-a|\le|y-x|)} + 1_{(|x+a|\le|y-x|)}$ from Lemma B.2 with $x = z'_{tT}\delta_\ell$ and $y = z'_{tT}\delta$ so that $|y - x| \le |z_{tT}||\delta_\ell - \delta| \le A\epsilon$. Thus, uniformly in δ,

$$\left|R_{T\delta}(a,0)-R_{T\delta_\ell}(a,0)\right|$$
$$\leq \frac{1}{T}\sum_{t>q_T}\left\{1_{(|z_{tT}'\delta_\ell-a|\leq\epsilon A,|z_{tT}|\leq A)}+1_{(|z_{tT}'\delta_\ell+a|\leq\epsilon A,|z_{tT}|\leq A)}\right\}$$
$$= R_{T\delta_\ell}(\epsilon A,a)+R_{T\delta_\ell}(\epsilon A,-a). \tag{B10}$$

Remove truncation. We can now remove the truncation, so that

$$R_{T\delta}(\alpha,\mu)\leq \frac{1}{T}\sum_{t>q_T}1_{(|z_{tT}'\delta_\ell-\mu|\leq\alpha)}=S_{T\delta}(\alpha,\mu). \tag{B11}$$

Return to the chaining inequality (B9), apply the bounds (B10), (B11) to bound

$$\sup_\delta R_{T\delta}(a,0)\leq \max_{\ell\leq L}S_{T\delta_\ell}(a,0)+\max_{\ell\leq L}S_{T\delta_\ell}(\epsilon A,a)$$
$$+\max_{\ell\leq L}S_{T\delta_\ell}(\epsilon A,-a).$$

We show that these terms vanish. For variables $S_\ell\geq 0$, the Boole and Markov inequalities give $\mathsf{P}(\max_\ell S_\ell>\eta)=\mathsf{P}\cup_\ell(S_\ell>\eta)\leq\sum_\ell \mathsf{P}(S_\ell>\eta)\leq(L/\eta)\max_\ell \mathsf{E}S_\ell$. Apply this to $S_{T\delta_\ell}$ as defined in (B11) noting $T-q_T\leq T$ to get, for any $\alpha\geq 0$, $\mu\in\mathbb{R}$,

$$\mathsf{P}\left\{\max_{\ell\leq L}S_{T\delta_\ell}(\alpha,\mu)>\eta\right\}\leq\frac{L}{T\eta}\max_{\ell\leq L}\sum_{t>q_T}\mathsf{P}\left(|z_{tT}'\delta_\ell-\mu|\leq\alpha\right)$$
$$\leq\frac{L}{\eta}\max_{\ell\leq L}\max_{t>q_T}\mathsf{P}\left(|z_{tT}'\delta_\ell-\mu|\leq\alpha\right). \tag{B12}$$

We must bound $\mathsf{P}(|z_{tT}'\delta-\mu|\leq\alpha)$. We distinguish between the cases $\sin^2\phi\geq 1/2$ and $\cos^2\phi\geq 1/2$, $\cos^2\psi\geq 1/2$ and $\cos^2\phi\geq 1/2$, $\sin^2\psi\geq 1/2$.

The case $\sin^2\phi\geq 1/2$ *and the linear trend term*. Since $\sin\phi\neq 0,\cos\theta>0$, we find

$$\frac{z_{tT}'\delta-\mu}{\sin\phi\cos\theta}=\frac{t}{T}+v_1\quad\text{with}$$
$$v_1=\left(u_t'\delta_u\cos\psi+\frac{1}{\sqrt{T}}v_t'\delta_v\sin\psi\right)\tan\phi-\frac{\sin\theta+\mu}{\sin\psi\cos\theta}.$$

Noting that $1/\cos\theta\leq 2(1+A)$ as found above while $\sin^2\phi\geq 1/2$ is assumed, we can bound $\alpha/(|\sin\phi\cos\theta|)\leq\alpha 2(1+A)\sqrt{2}=\widetilde{\alpha}_1$. Taken together, we get

$$\left(|z_{tT}'\delta-\mu|\leq\alpha\right)\subset\left(\left|\frac{z_{tT}'\delta-\mu}{\sin\phi\cos\theta}\right|\leq\widetilde{\alpha}_1\right)=\left(\left|\frac{t}{T}+v_1\right|\leq\widetilde{\alpha}_1\right).$$

This describes an interval for t of length $2T\widetilde{\alpha}_1$. Thus, the indicator for $(|z_{tT}'\delta-\mu|\leq\alpha)$ is unity for at most $2T\widetilde{\alpha}_1+1$ values of t. It follows that, as $(\alpha,T)\to(0,\infty)$,

$$S_{T\delta}(a)\leq T^{-1}\left(2T\widetilde{\alpha}_1+1\right)\leq\alpha 4(1+A)\sqrt{2}+T^{-1}\to 0.$$

The case $\cos^2\psi,\cos^2\phi\geq 1/2$ *and the I(0) term*. As $\cos\psi,\cos\phi,\cos\theta>0$, we find

$$\frac{z_{tT}'\delta-\mu}{\cos\psi\cos\phi\cos\theta}=u_t'\delta_u+v_2\quad\text{with}$$
$$v_2=\frac{1}{\sqrt{T}}v_t'\delta_v\tan\psi+\frac{t\tan\phi}{T\cos\psi}-\frac{\sin\theta+\mu}{\cos\psi\cos\phi\cos\theta}.$$

Noting that $1/\cos\theta\leq 2(1+A)$ as found above while $\cos^2\psi,\cos^2\phi\geq 1/2$ are assumed, we can bound $\frac{\alpha}{/}(|\cos\psi\cos\phi\cos\theta|)\leq\alpha 4(1+A)=\widetilde{\alpha}_2$. Taken together, we get

$$\left(|z_{tT}'\delta-\mu|\leq\alpha\right)\subset\left(\left|\frac{z_{tT}'\delta-\mu}{\cos\psi\cos\phi\cos\theta}\right|\leq\widetilde{\alpha}_2\right)=\left(|u_t'\delta_u+v_2|\leq\widetilde{\alpha}_2\right).$$

Taking probability and applying condition (i,b), we find a $C>0$ exists such that

$$\mathsf{P}\left(|z_{tT}'\delta-\mu|\leq\alpha\right)$$
$$\leq\max_{q_T<t\leq T}\sup_{\delta_u:|\delta_u|=1}\sup_{v_2\in\mathbb{R}}\mathsf{P}\left(|u_t'\delta_u+v_2|\leq\widetilde{\alpha}_2\right)\leq\widetilde{\alpha}_2 C. \tag{B13}$$

Combine the inequalities (B12), (B13) with the definition of $\widetilde{\alpha}_2$ to get

$$\mathsf{P}\left\{\max_{\ell\leq L}S_{T\delta_\ell}(\alpha,\mu)>\eta\right\}\leq\frac{L}{\eta}\widetilde{\alpha}_2 C=\frac{L}{\eta}\alpha 4(1+A)C\to 0, \tag{B14}$$

as $(\alpha,T)\to(0,\infty)$ since η,L,A,C are fixed.

The case $\sin^2\psi,\cos^2\phi\geq 1/2$ *and the I(1) term*. As $\sin\psi,\cos\phi,\cos\theta>0$, we find

$$\frac{z_{tT}'\delta-\mu}{\sin\psi\cos\phi\cos\theta}=\frac{1}{\sqrt{T}}v_t'\delta_v+v_3\quad\text{with}$$
$$v_3=u_t'\delta_u\cot\psi+\frac{t\tan\phi}{T\sin\psi}-\frac{\sin\theta+\mu}{\sin\psi\cos\phi\cos\theta},$$

Noting that $1/\cos\theta\leq 2(1+A)$ as found above while $\sin^2\psi,\cos^2\phi\geq 1/2$ is assumed, we get $\alpha/(|\sin\psi\cos\phi\cos\theta|)\leq\alpha 4(1+A)=\widetilde{\alpha}_3$. Taken together, we get

$$\left(|z_{tT}'\delta-\mu|\leq\alpha\right)\subset\left(\left|\frac{z_{tT}'\delta-\mu}{\sin\psi\cos\phi\cos\theta}\right|\leq\widetilde{\alpha}_3\right)$$
$$=\left(\left|v_t'\delta_v/\sqrt{T}+v_3\right|\leq\widetilde{\alpha}_3\right).$$

Taking probability, normalizing by $\sqrt{T/t}$ and applying condition (ii,b), we find a $C>0$ exists such that, uniformly in $q_T<t\leq T$, $\delta_v:|\delta_v|=1$, $v_3\in\mathbb{R}$

$$\mathsf{P}\left(|z_{tT}'\delta-\mu|\leq\alpha\right)\leq\mathsf{P}\left(\left|\frac{v_t'\delta_v}{\sqrt{T}}+v_3\right|\leq\widetilde{\alpha}_3\right)$$
$$=\mathsf{P}\left(\left|\frac{v_t'\delta_v}{\sqrt{t}}+v_3\right|\leq\widetilde{\alpha}_3\frac{\sqrt{T}}{\sqrt{t}}\right)\leq\widetilde{\alpha}_3 C\frac{\sqrt{T}}{\sqrt{t}}. \tag{B15}$$

Combine the inequalities (B12), (B15) and $\sum_{t=2}^T t^{-1/2}\leq\int_1^T t^{-1/2}dt<T^{1/2}/2$ with the definition of $\widetilde{\alpha}_3$ to get

$$\mathsf{P}\left\{\max_{\ell\leq L}S_{T\delta_\ell}(\alpha,\mu)>\eta\right\}\leq\frac{L}{T\eta}\max_{\ell\leq L}\sum_{t>q_T}\widetilde{\alpha}_3 C\frac{\sqrt{T}}{\sqrt{t}}$$
$$\leq\frac{L}{\eta}\alpha 2(1+A)C\to 0,$$

as $(\alpha,T)\to(0,\infty)$ since η,L,A,C are fixed. □

We bound the F_{Th} function for the ADL model with finitely many good episodes.

Lemma B.4. *Consider the setup in Section 4. Then*

$$F_{Th}(a)=\max_{\zeta:\#\zeta=h}\sup_{\delta:|\delta|=1}\frac{1}{h}\sum_{t\in\zeta}1_{(|x_{tT}'\delta|\leq a)}\leq\frac{T-h_{VAR}}{h}+\mathrm{o}_{\mathsf{P}}(1).$$

Proof of Lemma B.4. Objective. First, we bound $F_{Th}(a)\leq F_{hh}(a)+(T-h)/h$ where $F_{hh}(a)=\sup_{\delta:|\delta|=1}h^{-1}\sum_{t\in\zeta_T}1_{(|\delta' x_{tT}|\leq a)}$ by [21, Equation 4.2]. As F_{Th} sums over ζ, bound $\zeta\subset\zeta_T\cup\zeta_T^c$ and then bound the indicators on ζ_T^c by unity.

Second, in a similar fashion, bound $F_{hh}(a)\leq(h_{VAR}/h)F_{VAR}(a)+(h-h_{VAR})/h$, where $F_{VAR}(a)=\sup_{\delta:|\delta|=1}h_{VAR}^{-1}\sum_{t\in\zeta_{VAR,T}}1_{(|\delta' x_{tT}|\leq a)}$. Note that $h_{VAR}/h\leq 1$.

Third, the number G of periods with good ADL and VAR errors is finite by assumption. The indices for the good periods are, $\underline{t}_g < t \leq \bar{t}_g$ for some $g \leq G$ where G is finite, see (4.4). We bound $F_{VAR}(a) \leq \sum_{g=1}^{G}(h_g/h_{VAR})F_g(a)$ where $h_g = \bar{t}_g - \underline{t}_g$ noting $h_g \leq h_{VAR}$ and where $F_g(a) = \sup_{\delta:|\delta|=1} h_g^{-1}\sum_{t=\underline{t}_g+1}^{\bar{t}} 1_{(|\delta' x_{tT}|\leq a)}$.

Fourth, we have that h_g is non-decreasing. Suppose $h_g/h \to 0$ as $h \to \infty$ and note that $F_g(a) \leq 1$ by construction. Then $(h_g/h)F_g(a) \to 0$. For groups g where $h_g/\sqrt{T}$ diverges, we will bound $h_g/h \leq 1$. Combine all the bounds as

$$F_{Th}(a) \leq \frac{h - h_{VAR}}{h} + \sum_{g=1}^{G} 1_{(h_g/\sqrt{T} \text{ diverges})} F_g(a) + \mathrm{o}(1). \tag{B16}$$

Each good episode. Consider $F_g(a)$ for some g such that $h_g/\sqrt{T}$ diverges. By assumption, the ADL and VAR errors are normal. We apply Lemma B.3.

Condition (i,a). The I(0) component of $\widetilde{x}_{tT}$ is the process $\widetilde{\boldsymbol{y}}^*_{t-1}$. We show that a q_T exists such that $q_T/T \to 0$ and $\forall \epsilon > 0$, $\exists C > 0$ such that $\max_{\underline{t}_g+q_T<t\leq\bar{t}_g} \mathsf{P}(|\widetilde{\boldsymbol{y}}^*_t| > C) \leq \epsilon$. Now, $\widetilde{\boldsymbol{y}}^*_t$ satisfies the VARMA equation (3.15). The Granger-Johansen representation Theorem 1(b,i) shows that VARMA equation has a stationary solution, $\widetilde{\boldsymbol{y}}^*_{STAT,t}$ say, such that $\widetilde{\boldsymbol{y}}^*_t = \widetilde{\boldsymbol{y}}^*_{STAT,t} + \widetilde{\boldsymbol{Y}}^{t-\underline{t}_g}(\widetilde{\boldsymbol{y}}^*_{\underline{t}_g} - \widetilde{\boldsymbol{y}}^*_{STAT,\underline{t}_g})$. By stationarity (and normality), the components $\widetilde{\boldsymbol{y}}^*_{STAT,t}$ and $\widetilde{\boldsymbol{Y}}^{t-\underline{t}_g}\widetilde{\boldsymbol{y}}^*_{STAT,\underline{t}_g}$ are bounded in probability for $\underline{t}_g < t \leq \bar{t}_g$. Further, by the VARMA equation, $\widetilde{\boldsymbol{y}}^*_{\underline{t}_g} = \mathrm{O}_\mathsf{P}(1 + \max_{t\notin\zeta_{VAR,T}} |\varepsilon_t|)$. This is $\mathrm{O}_\mathsf{P}(T^c)$ by (4.6). Further, $\rho = \max|\text{eigen}(\widetilde{\boldsymbol{Y}})| < 1$ by Assumption 1. Now, let $q_T = T^{1/4}$ so that $q_T/h_g \leq q_T/T \to 0$, but $\log(\|\widetilde{\boldsymbol{Y}}\|^{q_T}\widetilde{\boldsymbol{y}}^*_{\underline{t}_g}) = T^{1/4}\log\|\widetilde{\boldsymbol{Y}}\| + c\mathrm{O}_\mathsf{P}(\log T) \to -\infty$, so that $\widetilde{\boldsymbol{Y}}^{q_T}\widetilde{\boldsymbol{y}}^*_{\underline{t}_g} = \mathrm{o}_\mathsf{P}(1)$. It follows that $\widetilde{\boldsymbol{y}}^*_t = \mathrm{O}_\mathsf{P}(1)$ uniformly in $\underline{t}_g + q_T < t \leq \bar{t}_g$ and the run-in period q_T is vanishing. Thus, condition (i,a) is satisfied.

Condition (i,b). The Granger-Johansen representation Theorem 1(b,ii) shows that

$$\min_{\underline{t}+k<t\leq\bar{t}} \min \text{eigen Var}\big(\widetilde{\boldsymbol{y}}^*_t | \widetilde{\boldsymbol{y}}^*_s, \underline{t} - k < s \leq t - k\big) > 0.$$

In particular, the variance bounded applies for $t > \underline{t} + q_T$. Under normality, this implies the densities are bounded and in turn condition (i,b) follows.

Condition (ii,a). For each episode we have $v_t = \sum_{t=\underline{t}_g+1}^{t}\varepsilon_t$. Let $T_g = \bar{t}_g - \underline{t}_g$ and $u \in [0,1]$. The normalized time series $v_{\lfloor uT_g\rfloor}/T_g^{1/2}$ converges to a Brownian motion on $D[0,1]$ with the Skorokhod metric. The supremum is a continuous mapping. The Continuous Mapping Theorem gives the desired bounded [37].

Condition (ii,b). For any unit vector, we get that $v'_{t-\underline{t}_g}\delta_v/(t-\underline{t}_g)^{1/2}$ is standard normal. Therefore, a $C > 0$ exists such that $\mathsf{P}(|v'_{t-\underline{t}_g}\delta_v/(t-\underline{t}_g)^{1/2}| + v < \alpha) \leq \alpha C$ uniformly in $t > \underline{t}_g$ and $v \in \mathbb{R}$. The condition follows.

Lemma B.3 now shows that $F_g(a) = \mathrm{o}_\mathsf{P}(1)$ as $(a,T) \to (0,\infty)$ for each g as h_g diverges. Insert in (B16) to finish the proof. □

Proof of Theorem 2. We check the Assumption A.1 (i,ii,iv) used in Theorem A.1.

Assumption A.1 (i,ii) are satisfied by assumption, see Section 4.1.

Assumption A.1 (iv). We find a ξ such that $0 < \xi < 2 - \lambda^{-1}$ and $\mathsf{P}\{F_{Th}(a) > \xi\} \to 0$ as $(a,T) \to (0,\infty)$. First, $2 - \lambda^{-1} > 1/2$ if and only if $\lambda > 2/3$, which is required in (4.8). Second, applying Lemma B.4 with the present assumptions gives

$$F_{Th}(a) \leq \frac{T - h_{VAR}}{h} + \mathrm{o}_\mathsf{P}(1).$$

We have $T - h_{VAR} < T/3$ by (4.8) and $h \geq h_{VAR} > 2T/3$, see Section 4.3. As $(1/3)/(2/3) = 1/2$, then $F_{Th}(a) < 1/2 + \mathrm{o}_\mathsf{P}(1)$. Thus, a ξ can be found as desired. □

Remark B.1. The bound $\lambda > 2/3$ in (4.8) is necessary to meet Assumption A.1(iv). Indeed, for a first order autogression where $p = k = 1$, the model Equation (2.1) is

$$\Delta y_t = \alpha(y_{t-1} - v_c) + \sigma\varepsilon_t \qquad \text{for } t = 1, \ldots, T.$$

Let data be generated by $\alpha = -1$, $v_c = 0$, $\sigma = 1$ so that $y_t = \varepsilon_t$. Let $h = \#\zeta_T = \lfloor\lambda n\rfloor$ and $\zeta_T = \zeta_{VAR,T} = (1,\ldots,h)$ with outliers $\varepsilon_t = \sqrt{2\log h}$ for $t \in \zeta_T^c$ so that $t > h$. Then

$$x_{tT} = \begin{pmatrix} \varepsilon_{t-1} \\ 1 \end{pmatrix} \text{ for any } t, \qquad x_{tT} = \begin{pmatrix} \sqrt{2\log h} \\ 1 \end{pmatrix} \text{ for } t > h.$$

We lower bound F_{tT}. Since the vector $\delta_* = (-1, \sqrt{2\log h})'$ is orthogonal to x_{tT} for $t \in \zeta_T^c$ so that $x'_{tT}\delta_* = 0$ and we get for any $a \geq 0$ that

$$F_{tT}(a) = \max_{\zeta:\#\zeta=h} \sup_{\delta:|\delta|=1} h^{-1}\sum_{t\in\zeta} 1_{(|x'_{tT}\delta|\leq a)} \geq h^{-1}$$
$$\sum_{t\in\zeta_T^c} 1_{(|x'_{tT}\delta_*|\leq a)} = \frac{T-h}{h} \to \frac{1-\lambda}{\lambda}.$$

Assumption A.1(iv) requires $(1-\lambda)/\lambda < 2 - 1/\lambda$, which is equivalent to $\lambda > 2/3$.

Conditions for Consistent Selection and Expansions

Lemma B.5. *Consider the setup in Section 4. Then*

a. $\max_{1\leq t\leq T} |\widetilde{x}_{tT}|^2 \leq \max_{1\leq t\leq T} |\widetilde{\boldsymbol{y}}^*_t|^2 + \max_{1\leq t\leq T} |T^{-1/2}\sum_{s=1}^{T}\varepsilon_s|^2 + 2$.

b. $\max_{1\leq t\leq T} |\widetilde{\boldsymbol{y}}^*_t| = \mathrm{O}_\mathsf{P}(\sqrt{\log T})$.

c. $\max_{1\leq t\leq T} |T^{-1/2}\sum_{s=1}^{T}\varepsilon_s| = \mathrm{O}_\mathsf{P}(1) + (\#\zeta_{VAR,T})\mathrm{O}_\mathsf{P}\{\sqrt{(\log T)/T}\}$.

Proof of Lemma B.5.

a. The vector $\widetilde{x}_{tT}$ has components $\widetilde{\boldsymbol{y}}^*_{t-1}$ and $T^{-1/2}\sum_{s=1}^{t-1}\varepsilon_s$ as well as t/T and 1, see (B6). The latter two are bounded by unity.

b. The I(0) component. For any t, we use the triangle inequality to bound $|\widetilde{\boldsymbol{y}}^*_t| = |\sum_{s=0}^{t}\widetilde{\boldsymbol{Y}}^s\widetilde{\boldsymbol{\varepsilon}}^*_{t-s}|$ by $\sum_{s=0}^{\infty}\|\widetilde{\boldsymbol{Y}}\|^s \max_{1\leq t\leq T}|\widetilde{\boldsymbol{\varepsilon}}^*_t|$. The geometric sum is convergent and the VAR errors are $\mathrm{O}_\mathsf{P}(\sqrt{\log T})$ due to (3.16), (4.10). Thus, $\max_{1\leq t\leq T} |\widetilde{\boldsymbol{y}}^*_t| = \mathrm{O}_\mathsf{P}(\sqrt{\log T})$.

c. The I(1) component. Expand this as

$$\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\varepsilon_s = \frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\varepsilon_s 1_{(s\in\zeta_{VAR,T})} + \frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\varepsilon_s 1_{(s\notin\zeta_{VAR,T})}. \tag{B17}$$

The first term in (B17) is a standard, normal random walk and converges to a Brownian motion when embedded in $C[0,1]$ with the uniform metric [37]. In particular, the maximum of its absolute value is $\mathrm{O}_\mathsf{P}(1)$. We bound the second term in (B17). It has at most $\#\zeta^c_{VAR,T}$ elements, which have order $\max_{t\leq T}|\varepsilon_t| = \mathrm{O}_\mathsf{P}(\sqrt{\log T})$ by (4.10). Overall, the I(1) component is $(\#\zeta^c_{VAR,T})\mathrm{O}_\mathsf{P}\{\sqrt{(\log T)/T}\}$. □

Lemma B.6. *Consider the sequence of data generating process of Section 4. Then $\widetilde{x}_{tT}$, $\bar{x}_{tT} = (0, I_q)\widetilde{x}_{tT}$ defined in (B6), (B7) satisfy*

a. $\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{tT}\widetilde{x}'_{tT} = \frac{1}{T}\sum_{t\in\zeta_T}\widetilde{x}_{tT}\widetilde{x}'_{tT} + \mathrm{o}(1) = \frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\widetilde{x}_{tT}\widetilde{x}'_{tT} + \mathrm{o}(1)$.

b. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{x}_{tT}\varepsilon_t = \frac{1}{\sqrt{T}}\sum_{t\in\zeta_T}\widetilde{x}_{tT}\varepsilon_t + \mathrm{o}_\mathsf{P}(1) = \frac{1}{\sqrt{T}}\sum_{t\in\zeta_{VAR,T}}\widetilde{x}_{tT}\varepsilon_t + \mathrm{o}_\mathsf{P}(1)$.

Proof of Lemma B.6.

a. We have $\zeta_{VAR,T} \subset \zeta_T$ by assumption so that $\zeta_T^c \subset \zeta_{VAR,T}^c$ and $T^{-1}\sum_{t\in\zeta_T^c}|\tilde{x}_{tT}|^2 \leq T^{-1}\sum_{t\in\zeta_{VAR,T}^c}|\tilde{x}_{tT}|^2 = \mathcal{S}_T$. Lemma B.5 shows $\max_{t\leq T}|\tilde{x}_{tT}|^2 = \mathrm{O_P}(\log T)\{1+(\#\zeta_{VAR,T}^c)^2/T\}$. Thus, $\mathcal{S}_T \leq T^{-1}(\#\zeta_{VAR,T}^c)\mathrm{O_P}(\log T)\{1+(\#\zeta_{VAR,T}^c)^2/T\}$. This vanishes when $\#\zeta_{VAR,T}^c = \mathrm{o}\{T^{2/3}/(\log T)^{1/3}\}$, which is implied by (4.12).

b. As before, it suffices to bound $\mathcal{T}_T = T^{-1/2}\sum_{t\in\zeta_{VAR,T}^c}|\tilde{x}_{tT}\varepsilon_t|$, which we can bound by $\mathcal{T}_T \leq T^{-1/2}(\#\zeta_{VAR,T}^c)(\max_{t\leq T}|\tilde{x}_{tT}|)(\max_{t\leq T}|\varepsilon_t|)$. As, the maxima are $\mathrm{O_P}(\sqrt{\log T})$ by Lemma B.5 and by (4.10), we get $\mathcal{T}_T \leq \mathrm{O_P}(\log T/\sqrt{T})\#\zeta_{VAR,T}^c = \mathrm{o_P}(1)$ as $\#\zeta_{VAR,T}^c = \mathrm{o}(\sqrt{T}/\log T)$ by condition (4.12). □

Lemma B.7. *Consider the sequence of data generating process of Section 4. Let $\tilde{\mathbf{y}}_g^\dagger$ denote a stationary solution of the normal VARMA Equation (3.15) for $\underline{t}_g < t \leq \bar{t}_g$. Let*

$$\tilde{x}_{tT}^\dagger = \begin{pmatrix} \sum_{s=0}^{t-\underline{t}_g-1}\tilde{\mathbf{Y}}^s\tilde{\boldsymbol{\varepsilon}}_{t-s}^* + \tilde{\mathbf{Y}}^{t-\underline{t}_g}\tilde{\mathbf{y}}_g^\dagger \\ \frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\boldsymbol{\varepsilon}_s 1_{(s\in\zeta_{VAR,T})} \\ t/T \\ 1 \end{pmatrix}, \qquad \bar{x}_{tT}^\dagger = (0_{q\times 1}, I_q)\tilde{x}_{tT}^\dagger,$$

where $q=(k+1)p+r+1$. Then $\tilde{x}_{tT}, \bar{x}_{tT} = (0, I_q)\tilde{x}_{tT}$ defined in (B6), (B7) satisfy

a. $\frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\left|\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger\right|^2 = \mathrm{o_P}(1)$.

b. $\frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\tilde{x}_{tT}\tilde{x}_{tT}' = \frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\tilde{x}_{tT}^\dagger\tilde{x}_{tT}^{\dagger\prime} + \mathrm{o_P}(1) + \mathrm{o}\left\{\left(\frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\left|\tilde{x}_{tT}^\dagger\right|^2\right)^{1/2}\right\}$.

c. $\frac{1}{\sqrt{T}}\sum_{t\in\zeta_{VAR,T}}\bar{x}_{tT}\varepsilon_t = \frac{1}{\sqrt{T}}\sum_{t\in\zeta_{VAR,T}}\bar{x}_{tT}^\dagger\varepsilon_t + \mathrm{o_P}(1)$.

Proof of Lemma B.7.

a. Write $\tilde{x}_{tT} = \tilde{x}_{tT}^\dagger + (\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger)$. Recalling the definition of $\tilde{x}_{tT}$ and $\tilde{\mathbf{y}}_t^*$ from (B6), (3.15) we find for $\underline{t}_g < t \leq \bar{t}_g$ that $\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger = (z_{1t}', z_{2t}', 0, 0)'$ where $z_{1t} = \tilde{\mathbf{Y}}^{t-\underline{t}_g}(\tilde{\mathbf{y}}_{\underline{t}_g}^* - \tilde{\mathbf{y}}_g^\dagger)$ and $z_{2t} = T^{-1/2}\sum_{s=1}^{t-1}\boldsymbol{\varepsilon}_s 1_{(s\in\zeta_{VAR,T}^c)}$. By the triangle inequality and the submultiplicativity of the spectral norm, we then get that

$$\left\|\frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\left(\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger\right)\left(\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger\right)'\right\| \leq \frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\left|\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger\right|^2 = Z_{1T}+Z_{2T}, \qquad \text{(B18)}$$

where $Z_{jT} = T^{-1}\sum_{t\in\zeta_{VAR,T}}\left|z_{jt}\right|^2$.

For Z_{1T}, bound $|z_{1t}|^2 \leq \|\tilde{\mathbf{Y}}\|^{2(t-\underline{t}_g)}(|\tilde{\mathbf{y}}_{\underline{t}_g}^*| + |\tilde{\mathbf{y}}_g^\dagger|)$. Here, g takes finitely many values, $\max_{1\leq t\leq T}|\tilde{\mathbf{y}}_t^*| = \mathrm{O_P}(\sqrt{\log T})$ by Lemma B.5(b), $|\tilde{\mathbf{y}}_g^\dagger| = \mathrm{O_P}(1)$, while the geometric series $\sum_{t\in\zeta_{VAR,T}}\|\tilde{\mathbf{Y}}\|^{2(t-\underline{t}_g)} = \sum_{g=1}^G\sum_{t=\underline{t}_g+1}^{\bar{t}_g}\|\tilde{\mathbf{Y}}\|^{2(t-\underline{t}_g)}$ is bounded by $G\sum_{t=0}^\infty\|\tilde{\mathbf{Y}}\|^{2t} < \infty$. Combine and normalize by T to get $Z_{1T} = \mathrm{O_P}(\sqrt{\log T}/T) = \mathrm{o_P}(1)$.

For Z_{2T}, we bound $Z_{2T} \leq T^{-1}(\#\zeta_{VAR,T})\max_{1\leq t\leq T}|z_{2t}|^2$. Bound $\#\zeta_{VAR,T} \leq T$, so that $Z_{2T} \leq \max_{1\leq t\leq T}|z_{2t}|^2$. Now, $|z_{2t}| \leq T^{-1/2}(\#\zeta_{VAR,T}^c)\max_{t\leq T}|\varepsilon_t|$ uniformly in t. Here, we have $\max_{t\leq T}|\varepsilon_t| = \mathrm{O_P}(\sqrt{\log T})$ by (4.10) and $\#\zeta_{VAR,T}^c = \mathrm{o_P}(\sqrt{T/\log T})$ by (4.12). Combine these bounds to get $\max_{1\leq t\leq T}|z_{2t}| = \mathrm{o_P}(1)$, so that $Z_{2T} = \mathrm{o_P}(1)$.

b. Write $\tilde{x}_{tT} = \tilde{x}_{tT}^\dagger + (\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger) = v_t + w_t$ say. For vectors v_i, w_i and $\|\cdot\|$ denoting the spectral norm, the triangle, sub-multiplicative and Cauchy-Schwarz inequalities give

$$\left\|\sum_{i=1}^n v_i w_i'\right\| \leq \sum_{i=1}^n\left\|v_i w_i'\right\| \leq \sum_{i=1}^n|v_i||w_i| \leq \left(\sum_{i=1}^n|v_i|^2\sum_{i=1}^n|w_i|^2\right)^{1/2}.$$

With this and the triangle inequality, we can bound

$$\sum_{i=1}^n(v_i+w_i)(v_i+w_i)' \leq \sum_{i=1}^n v_iv_i' + \sum_{i=1}^n w_iw_i' + 2\left(\sum_{i=1}^n|v_i|^2\sum_{i=1}^n|w_i|^2\right)^{1/2}.$$

We apply this bound with $v_t = \tilde{x}_{tT}^\dagger$, $w_t = \tilde{x}_{tT}-\tilde{x}_{tT}^\dagger$ and sum over $t\in\zeta_{VAR,T}$. Thus, it suffices that $T^{-1}\sum_{t\in\zeta_{VAR,T}}|w_t|^2 = \mathrm{o_P}(1)$, which was shown in part (a).

c. We argue that $\sum_{t\in\zeta_{VAR,T}}(\bar{x}_{tT}-\bar{x}_{tT}^\dagger)\varepsilon_t = \mathrm{o_P}(T^{1/2})$. Since $\bar{x} = (0, I_q)\tilde{x}$ and $\zeta_{VAR,T}$ has finitely many good episodes, see (4.4), and $\tilde{x}_{tT}-\tilde{x}_{tT}^\dagger = (z_{1t}', z_{2t}', 0, 0)$, see part (a), then it suffices to argue that $\sum_{t=\underline{t}_g+1}^{\bar{t}_g}z_{jt}\varepsilon_t = \mathrm{o_P}(T^{1/2})$ for each $1\leq g\leq G$, $j=1,2$.

For $j=1$, we have $\sum_{t=\underline{t}_g+1}^{\bar{t}_g}z_{1t}\varepsilon_t = (\sum_{t=\underline{t}_g+1}^{\bar{t}_g}\tilde{\mathbf{Y}}^{t-\underline{t}_g}\varepsilon_t)(\tilde{\mathbf{y}}_{\underline{t}_g}^* - \tilde{\mathbf{y}}_g^\dagger)$. The geometric average of the independent normal ε_t is $\mathrm{O_P}(1)$, while it was established in (a) that $\tilde{\mathbf{y}}^* = \mathrm{O_P}(\sqrt{\log T})$ and $\tilde{\mathbf{y}}_g^\dagger = \mathrm{O_P}(1)$. Thus, $\sum_{t=\underline{t}_g+1}^{\bar{t}_g}z_{1t}\varepsilon_t = \mathrm{O_P}(\sqrt{\log T}) = \mathrm{o_P}(\sqrt{T})$.

For $j=2$, we have $\sum_{t=\underline{t}_g+1}^{\bar{t}_g}z_{2t}\varepsilon_t = z_{2\underline{t}_g}\sum_{t=\underline{t}_g+1}^{\bar{t}_g}\varepsilon_t$. Here, $z_{2\underline{t}_g} = \mathrm{o_P}(1)$ by (a), while the sum of normals is $\mathrm{O_P}(\sqrt{T})$. Thus, $\sum_{t=\underline{t}_g+1}^{\bar{t}_g}z_{2t}\varepsilon_t = \mathrm{o_P}(\sqrt{T})$. □

Lemma B.8. *Consider the sequence of data generating process of Section 4. Let $\tilde{\mathbf{y}}_g^\dagger$ denote stationary solution of the normal VARMA Equation (3.15) for $\underline{t}_g < t \leq \bar{t}_g$. Let W be standard p-dimensional Brownian motion so that $B = \sigma^{-1}(1, -\omega)\mathbf{A}W$ is a standard univariate Brownian motion. Concatenate $W_u, u, 1$ as F_u. Let $\tilde{\mathbf{y}}_g^\dagger$ denote stationary solution of the normal VARMA Equation (3.15) for $\underline{t}_g < t \leq \bar{t}_g$. Let Σ_{yy} be the variance of $(0, I_{r+kp-1})\tilde{\mathbf{y}}_g^\dagger$. Suppose $h_{VAR} = \#\zeta_{VAR,T} \to \infty$ as $T\to\infty$. Then*

$$\left(\frac{1}{h_{VAR}}\sum_{t\in\zeta_{VAR,T}}\bar{x}_{tT}^\dagger\bar{x}_{tT}^{\dagger\prime}, \frac{1}{\sqrt{h_{VAR}}}\sum_{t\in\zeta_{VAR,T}}\bar{x}_{tT}^\dagger\varepsilon_t\right) \xrightarrow{\mathsf{D}} \left[\begin{pmatrix}\Sigma_{yy} & 0 \\ 0 & \int_0^1 F_uF_u'du\end{pmatrix}, \begin{pmatrix}N \\ \int_0^1 F_u dB_u\end{pmatrix}\right], \qquad \text{(B19)}$$

where N is $\mathsf{N}(0, \Sigma_{yy})$ and independent of W and hence also of F, B.

Proof of Lemma B.8. Chan and Wei [71] prove this for a univariate autoregression without deterministic terms and a zero initialization. The impact of starting in a stationary solution is negligible. Chan [72] extends this to include deterministic terms. Johansen [25, appendix B] extends this to VARs. □

Proof of Theorem 3. We apply Theorem A.1 and must check Assumption A.1. Parts (i, ii, iv) were checked for Theorem 2.

Assumption A.1(iii) is satisfied by (4.2) in Section 4.

Assumption A.1(v). We show that $T^{-1}\sum_{t=1}^T x_{tT}x_{tT}' = \mathrm{O_P}(1)$. As x_{tT} is a subvector of $\tilde{x}_{tT}$ in (B6), it suffices that $T^{-1}\sum_{t=1}^T\tilde{x}_{tT}\tilde{x}_{tT}' = \mathrm{O_P}(1)$. Lemma B.6(a) using that $T - h_{VAR} = \mathrm{o}\{T^{2/3}/(\log T)^{1/3}\}$ show that this sum equals $T^{-1}\sum_{t\in\zeta_{VAR,T}}\tilde{x}_{tT}\tilde{x}_{tT}'$, which is $\mathrm{O_P}(1)$ by Lemma B.7(b) as $T^{-1}\sum_{t\in\zeta_{VAR,T}}\tilde{x}_{tT}^\dagger\tilde{x}_{tT}^{\dagger\prime} = \mathrm{O_P}(1)$ by Lemma B.8.

Assumption A.1(vi, b), first part. We have $\max_{t\leq T}|x_{tT}|^2$ is bounded by $\mathrm{O_P}(\log T) + \max_{1\leq t\leq T}|T^{-1/2}\sum_{s=1}^T\varepsilon_t|^2$ by Lemma B.5(a, b). The random walk term is $\mathrm{O_P}(\log T)$ by Lemma B.5(c) as $\#\zeta_{VAR,T} = T - h_{VAR} = \mathrm{O}(\sqrt{T})$ by condition (4.12).

Assumption A.1(vi, b), second part concerns the intermediate order statistics of x_{tT}. These relate to the intermediate order statistics of $\boldsymbol{y}_t^*$, since the random walk part, the linear trend and the constant of x_{tT} are $\mathrm{O_P}(1)$ by Lemma B.5using that $\#\zeta^c_{VAR,T} = T - h_{VAR} = \mathrm{O}(\sqrt{T})$ by condition (4.12). Thus, the intermediate order statistics of x_{tT} relate to those of $\boldsymbol{y}_t^*$. Again, this can be split in a normal, stationary VARMA part and the outlier part. For the normal, stationary VARMA part the intermediate extreme decline as required [42]. For the outlier part, the desired behaviour of the intermediate extremes of the outliers must be assumed as done in (4.11).

Assumption A.1(vii). Let $\hat{\beta}_{\zeta_T,T}$ denote the LTS estimator for regression on x_{tT}. Write $S_{\zeta_T} = (\hat{\beta}_{\zeta_T,T} - \beta)'(\sum_{i\in\zeta_T} x_{tT}x_{tT}')(\hat{\beta}_{\zeta_T,T} - \beta)$ as

$$S_{\zeta_T} = \left(\frac{1}{\sqrt{h}}\sum_{i\in\zeta_T}\varepsilon_t x_{tT}'\right)\left(\frac{1}{h}\sum_{i\in\zeta_T}x_{tT}x_{tT}'\right)^{-1}\left(\frac{1}{\sqrt{h}}\sum_{i\in\zeta_T}x_{tT}\varepsilon_t\right), \tag{B20}$$

where $h = \#\zeta_T$. We show $S_{\zeta_T} = \mathrm{O_P}(1)$. As x_{tT} is a subset of $\overline{x}_{tT}$, which is a subset of $\widetilde{x}_{tT}$, the desired result follows by replacing ζ_T by $\zeta_{VAR,T}$ using Lemma B.6 requiring $T - h_{VAR} = \mathrm{o}(\sqrt{T}/\log T)$, replacing $\overline{x}_{tT}, \widetilde{x}_{tT}$ with $\overline{x}^\dagger_{tT}, \widetilde{x}^\dagger_{tT}$ using Lemma B.7, and using the convergence in Lemma B.8.

We can now apply Theorem A.1 to get

$$\max_{\zeta\in\mathcal{M}_T}\left|\left(\sum_{i\in\zeta}x_{tT}x_{tT}'\right)^{1/2}\left(\hat{\beta}_{\zeta,T} - \beta\right) - \left(\sum_{i\in\zeta_T}x_{tT}x_{tT}'\right)^{1/2}\left(\hat{\beta}_{\zeta_T,T} - \beta\right)\right| = \mathrm{o_P}(1).$$

Finally, $x_{tT} = B_T x_t$ for some invertible matrix B_T. The above expression is invariant to rotations as explained in Remark B.2. Thus, we can replace x_{tT} with x_t as desired. We can also replace ζ_T by $\zeta_{VAR,T}$ by use of Lemma B.6. ◻

Remark B.2. In Theorem 3, the square roots of the matrices $M = \sum_{t\in\zeta} x_t x_t'$ and $N = \sum_{t\in\zeta_T} x_t x_t'$ must be found through joint diagonalization. As M, N are symmetric and positive semi-definite, there exists an invertible matrix S and a diagonal matrix Λ such that $N = SS'$ and $M = S(I_{\dim x} + \Lambda)S'$ [25, lemma A.5]. The elements λ of Λ solve the equation $\det\{(1+\lambda)N - M\} = 0$ where $1 + \lambda > 0$ with corresponding eigenvectors v, such that $(1+\lambda)Nv = Mv$ and where the v's are the columns of $V = (S')^{-1}$. We define the right square roots $N^{1/2} = S'$ and $M^{1/2} = (I_{\dim x} + \Lambda)^{1/2}S'$. In particular, we can write

$$\begin{aligned}
D_\zeta &= \left(\sum_{i\in\zeta}x_t x_t'\right)^{1/2}\left(\hat{\beta}_\zeta - \beta\right) - \left(\sum_{i\in\zeta_T}x_t x_t'\right)^{1/2}\left(\hat{\beta}_{\zeta_T} - \beta\right) \\
&= \left(\sum_{i\in\zeta}x_t x_t'\right)^{1/2}\left(\sum_{i\in\zeta}x_t x_t'\right)^{-1}\sum_{i\in\zeta}x_t\varepsilon_t \\
&\quad - \left(\sum_{i\in\zeta_T}x_t x_t'\right)^{1/2}\left(\sum_{i\in\zeta_T}x_t x_t'\right)^{-1}\sum_{i\in\zeta_T}x_t\varepsilon_t \\
&= S'\left(SS'\right)^{-1}\sum_{t\in\zeta}x_t\varepsilon_t - \left(I_{\dim x} + \Lambda\right)^{1/2}S' \\
&\quad \left\{S\left(I_{\dim x} + \Lambda\right)S'\right\}^{-1}\sum_{t\in\zeta_T}x_t\varepsilon_t
\end{aligned}$$

and then eliminate terms to get

$$D_\zeta = S^{-1}\sum_{t\in\zeta}x_t\varepsilon_t - \left(I_{\dim x} + \Lambda\right)^{-1/2}S^{-1}\sum_{t\in\zeta_T}x_t\varepsilon_t.$$

In the proof of Theorem 3 a rotated, normalized version of the regressors is used as in $x_{tT} = B_T x_t$. We then get that $\sum_{t\in\zeta} x_{tT}x_{tT}' = B_T M B_T' = B_T S(I_{\dim x} + \Lambda)S'B_T'$ and $\sum_{t\in\zeta_T} x_{tT}x_{tT}' = B_T N B_T' = B_T SS'B_T'$. Arguing as above, we find

$$\begin{aligned}
D_{\zeta,T} &= \left(\sum_{i\in\zeta}x_{tT}x_{tT}'\right)^{-1/2}\sum_{i\in\zeta}x_{tT}\varepsilon_t \\
&\quad - \left(\sum_{i\in\zeta_T}x_{tT}x_{tT}'\right)^{-1/2}\sum_{i\in\zeta_T}x_{tT}\varepsilon_t = D_\zeta.
\end{aligned}$$