Nuffield College
UNIVERSITY OF OXFORD

# ECONOMICS DISCUSSION PAPERS

Paper No: 2025-W03

## Least Trimmed Squares: Cointegration and outliers

By Vanessa Berenguer-Rico and Bent Nielsen

June 2025

# Least Trimmed Squares: Cointegration and outliers

Vanessa Berenguer-Rico[*‡] & Bent Nielsen[†‡]

2 June 2025

**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 analyze 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 the UK economic models of Ball and Burns (1968), Ball et al. (1975). 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. (1978) 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

[*]Mansfield College & Department of Economics, University of Oxford. Address for correspondence: Mansfield College, Oxford OX1 3TF, UK. E-mail: vanessa.berenguer-rico@economics.ox.ac.uk.

[†]Nuffield College & Department of Economics, University of Oxford. Address for correspondence: Nuffield College, Oxford OX1 1NF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk.

[‡]Support from Aarhus Center for Econometrics (ACE) funded by the Danish National Foundation grant DNRF186 is gratefully acknowledged.

estimator, this procedure is not robust in general and it has to be used with care (Welsh and Ronchetti, 2002).

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 (Atkinson and Riani, 2000) with an implementation to structural time series in Riani (2004), and Impulse Indicator Saturation (Hendry et al., 2008) which is aimed at ADL models and implemented in OxMetrics (Doornik, 2009), as Gets in R (Pretis et al., 2018) and in the Eviews software. Asymptotic analysis of these methods has focused on the situation without outliers (Johansen and Nielsen, 2009, 2016a,b) and with little emphasis on cointegration. In time series, it is common to distinguish between additive and innovative outliers (Fox, 1972), both of which can be relevant in cointegrated models (H. B. Nielsen, 2004). Methods based on extreme value theory can detect a finite number of additive outliers in a first order autoregression (Burridge and Taylor, 2006). 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 (Croux and Joossens, 2008).

The least trimmed squares (LTS) estimator (Rousseeuw, 1984) 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. Recently, LTS has 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 (Berenguer-Rico et al., 2023). 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 (Berenguer-Rico and Nielsen, 2024).

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 (Granger, 1986; Engle and Granger, 1987).

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 analyzed 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 (Engle and Granger, 1987; Johansen, 1988, 1995) 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 (Engle et al., 1983; Johansen, 1992), the hypothesis of no cointegration can be tested using Dickey-Fuller type distributions with a t-statistic (Banerjee et al., 1998) or a likelihood ratio statistic (Harbo et al., 1998); tests on coefficients in the cointegrating vector have standard normal inference (Phillips, 1988; Johansen, 1992); and tests for lag length (Nielsen, 2006).

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 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

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

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

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

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

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

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

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

## 2.2 Least Trimmed Squares estimation

LTS estimation was suggested by Rousseeuw (1984). The esimator divides the data in two groups. There is a given number of $h$ good errors with indices in an unknown $h$-subset $\zeta$ of $1, \ldots, T$. The indices in $\zeta$ need not be consecutive. The remaining $T-h$ indices in $\zeta^c$ are the outliers. The LTS estimator finds the $h$-subsample with the smallest residual sum of squares (Rousseeuw and van Driessen, 2000). Thus, the LTS estimator is defined as follows. Given a $h$-index set $\zeta$, the OLS estimators are

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

where $\sum_{t\in\zeta} x_t x_t'$ is assumed invertible for any choice of $\zeta$. 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 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 (Rousseeuw and van Driessen, 2000, 2006).

# 3 The vector autoregression

For inference, we set up a joint vector autoregressive model for the variables. We find the Granger-Johansen representation.

## 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 vocubulary of Johansen (1995) and refer to e.g. $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}^*$ 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 (Johansen, 1992) and the notion of weak exogeneity (Engle et al., 1983). 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

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

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 $\Delta y_t$. This leads to the ADL equations (2.1), (2.3) with

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

The ADL errors are defined through

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

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

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

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 (3.2) for $z_t$ at a particular $t$, but not in the ADL equation (3.4) is of special interest. The ADL equation may then have a structural interpretation. This relates to the ideas of super-exogeneity (Engle et al., 1983; Hendry and Santos, 2010) and causal transmission (Bazinas and Nielsen, 2022). We will allow for this situation.

## 3.2 A new Granger-Johansen representation

We need a Granger-Johansen representation for each of the 'good' episodes. We extend the Johansen (1995, Theorem 4.2) representation to assess the exact autoregressive structures of the common trends and for the stationary components and how these depend on the initial observations. This generalizes univariate ideas in Nielsen (2001).

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

$$\boldsymbol{y}_{t-1}^* = \begin{pmatrix} \boldsymbol{\beta}'\boldsymbol{x}_{t-1}^* \\ \Delta\boldsymbol{x}_{t-1}^* \\ \vdots \\ \Delta\boldsymbol{x}_{t-k+1}^* \end{pmatrix}, \quad \widetilde{\boldsymbol{y}}_{t-1}^* = \begin{pmatrix} \Delta\boldsymbol{x}_t^* \\ \boldsymbol{y}_{t-1}^* \end{pmatrix}, \quad \overline{\boldsymbol{y}}_{t-1}^* = \begin{pmatrix} \Delta z_t^* \\ \boldsymbol{y}_{t-1}^* \end{pmatrix} = \big(0, I_{\dim \overline{\boldsymbol{y}}}\big)\widetilde{\boldsymbol{y}}_{t-1}^*, \tag{3.6}$$

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

$$\boldsymbol{y}_t^* = \boldsymbol{Y}\boldsymbol{y}_{t-1}^* + \boldsymbol{e}_{\boldsymbol{y}^*}\boldsymbol{A}\boldsymbol{\varepsilon}_t, \tag{3.7}$$

where

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

We assume that $\boldsymbol{Y}$ has eigenvalues with absolute values less than unity. This implies that (3.7) has a stationary solution when the errors are independent normal.

**Assumption 3.1.** ***Stationarity.*** $|eigen(\boldsymbol{Y})| < 1$.

The stationary condition implies, in particular, that the eigenvalues of $\boldsymbol{Y}$ differ from unity. This, in turn, is equivalent to the socalled I(1) condition by (Johansen, 1995, Theorem 4.2), see Nielsen (2009) 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, for instance, $(\boldsymbol{\beta}, \boldsymbol{\beta}_\perp)$ is invertible and $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = 0$. Let $\boldsymbol{\Psi} = I_p - \sum_{j=1}^{k-1}\boldsymbol{\Gamma}_j$. 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:

$$\boldsymbol{C} = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}_\perp'\boldsymbol{\Psi}\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}_\perp', \qquad \psi' = -\boldsymbol{\beta}_\perp'\boldsymbol{C}\Big(\boldsymbol{\Psi}\overline{\beta}, \boldsymbol{\Gamma}_1, \ldots, \sum_{j=1}^{k-1}\boldsymbol{\Gamma}_j\Big). \tag{3.8}$$

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

$$\widetilde{\boldsymbol{y}}_t^* = \widetilde{\boldsymbol{Y}}\widetilde{\boldsymbol{y}}_{t-1}^* + \widetilde{\boldsymbol{\varepsilon}}_t^*, \tag{3.9}$$

with

$$\widetilde{\boldsymbol{Y}} = \begin{pmatrix} 0 & \boldsymbol{\omega}' \\ 0 & \boldsymbol{Y} \end{pmatrix}, \qquad \widetilde{\boldsymbol{\varepsilon}}_t^* = \begin{pmatrix} \boldsymbol{A}\boldsymbol{\varepsilon}_t \\ \boldsymbol{e}_{\boldsymbol{y}^*}\boldsymbol{A}\boldsymbol{\varepsilon}_{t-1} \end{pmatrix} \tag{3.10}$$

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 have the following Granger-Johansen representation.

**Theorem 3.1.** ***Granger-Johansen representation***. *Consider the model equation (3.2) for $\underline{t} < t$ and suppose Assumption 3.1.*
(*a*) ***I(1) component***. *Define $\boldsymbol{C}, \boldsymbol{\nu}$ as in (3.8). Then*

$$\boldsymbol{\beta}_\perp' \boldsymbol{x}_t^* = \boldsymbol{\beta}_\perp' \boldsymbol{C}\boldsymbol{A} \sum_{s=\underline{t}+1}^{t} \boldsymbol{\varepsilon}_s + \boldsymbol{\psi}' \boldsymbol{y}_t^* + \boldsymbol{\beta}_\perp' \boldsymbol{x}_{\underline{t}}^* - \boldsymbol{\psi}' \boldsymbol{y}_{\underline{t}}^* \qquad \text{for } \underline{t} < t. \tag{3.11}$$

(*b*) ***I(0) component***. *Suppose $\boldsymbol{\varepsilon}_t$ are independent $\mathsf{N}_p(0, I_p)$ 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*) $\min_{\underline{t}+k<t\leq\bar{t}} \min \text{eigen} \, \mathsf{Var}(\widetilde{\boldsymbol{y}}_t^* \mid \widetilde{\boldsymbol{y}}_s^*, \underline{t} - k < s \leq t - k) > 0$.

Theorem 3.1 extends the representation of Johansen (1995, Theorem 4.2). 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 implicitely 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 $\zeta_T$ 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

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

***The outlier ADL errors*** must be extreme relative to the standard normal ADL good errors. Extreme value theory shows that $\max_{t\in\zeta} \varepsilon_t/\sqrt{2\log h} \to 1$ almost surely (DasGupta, 2008, 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 Fox (1972).

## 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

$$0 = \bar{t}_0 \leq \underline{t}_1 < \bar{t}_1 < \cdots < \underline{t}_g < \bar{t}_g < \cdots < \underline{t}_G < \bar{t}_G \leq \underline{t}_{G+1} = T \quad \text{for } g = 1, \ldots, G. \quad (4.4)$$

***The good VAR errors*** are assumed independent normal:

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

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

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

## 4.3 Conditions for boundedness

The outlier errors must be of polynomial order

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

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

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

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

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

$$h_{VAR} > 2T/3. \quad (4.8)$$

This assumption will be sufficient to prove boundedness of the LTS estimator $\hat{\beta}$ for the ADL regression that selects $h$ observations, where $h \geq h_{VAR}$. It shows that the LTS estimator remains bounded if up to 1/3 of the observations are outliers. This contrasts 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 by Rousseeuw (1994). 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

We must 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

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

The initial observations $|\boldsymbol{x}_{-\ell}|^2$ for $0 \leq \ell < k$ 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 vector of $\widetilde{\boldsymbol{y}}_t^*$ and $\boldsymbol{\beta}_\perp' \boldsymbol{x}_t^* / \sqrt{T}$ have increasing order statistics $q_1 \leq \cdots \leq q_T$ satisfying

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

The maximum of a normal I(1) series is of the same order as the series itself, whereas the maximum of a stationary, normal VAR diverges, but satisfies the condition (4.11) (Watts et al., 1982). 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

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

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 (2024), 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 $\zeta$ of $\hat{\sigma}_\zeta^2$.

**Theorem 5.1.** ***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****:* $\max_{\zeta\in\mathcal{M}_T}|\hat{\beta}_\zeta - \hat{\beta}_{\zeta_T}| = \mathrm{O}_\mathsf{P}(1)$.

With Theorem 5.1 we avoid compactness assumptions for the parameter space. The asymptotic boundedness requires that the proportion of outliers is less than 1/3, see (4.8). This is an asymptotic parallel to the finite sample breakdown point introduced by Donoho and Huber (1983) and analyzed for LTS by Rousseeuw and Leroy (1987, §3.4). For further discussion in the context of M-estimators, see Klooster and Nielsen (2025).

## 5.2 Consistent selection and expansions

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

**Theorem 5.2.** ***Consistent selection and expansions.*** *Consider the setup in Sections 4.1, 4.2, 4.3 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 by*** $\hat{\zeta}$. $\forall 0 < \eta < 1$: $\max_{\zeta\in\mathcal{M}_T}\#(\zeta\cap\xi_T^c)/h = \mathrm{O}_\mathsf{P}(h^{\eta-1})$.
(*b*) ***Expansion for*** $\hat{\sigma}^2$. $\max_{\zeta\in\mathcal{M}_T} h^{1/2}|\hat{\sigma}^2_\zeta - \hat{\sigma}^2_{\xi_T}| = \mathrm{o}_\mathsf{P}(1)$.
(*c*) ***Expansion for*** $\hat{\beta}$. *Then*

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

*The square root matrices are defined through joint diagonalization, see Remark B.2 in the Appendix.*

Theorem 5.2 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 $\zeta_T$ 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 by Doornik et al. (1998).

## 5.3 Asymptotic distributions

Theorem 5.2 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.

To illustrate the different inferences one may want to draw from ADL analysis, we restate the constant level model (2.1).

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

The inference results include:

1. The hypothesis of no cointegration, $\alpha = 0$, can be tested using Dickey-Fuller type distributions with a t-statistic as in Banerjee et al. (1998) or a LR-statistic as in Harbo et al. (1998) on the good observations selected by LTS.

2. Hypotheses on the cointegration parameter $\kappa$ can be tested using standard normal inference as in Johansen (1992) on the good observations selected by LTS.

It should be noted that those results require weak exogeneity. An autoregression arises when removing $z_t$. This gives the model equation

$$\Delta y_t = \alpha(y_{t-1} - \nu_c) + \sum_{j=1}^{k-1} \gamma_j' \Delta y_{t-j} + \sigma \varepsilon_t \quad \text{for } t = 1, \ldots, T. \qquad (5.1)$$

The unit root hypothesis $\alpha = 0$ can be investigated by

3. The augmented Dickey-Fuller t or F test, see Dickey and Fuller (1979, 1981) on the good observations selected by LTS.

Further, one can test

4. A restriction on lag length, such as $\gamma_{k-1} = 0$ in either (2.1) or (5.1) can be tested using normal inference, see Nielsen (2006) for a derivation, on the good observations selected by LTS.

## 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

$$y_t = \beta' x_t + \sigma \varepsilon_t, \qquad (5.2)$$

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 3.1 applies with an empty I(1) component. *Second*, the boundedness result in Theorem 5.1 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 5.2 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-test for the cointegrating coefficient. Tests on other parameters and variations of the simulation design are reported in the supplementary material. The code was written in Matlab with LTS estimation done using the `mlts.m` code (Agullo et al., 2008)

We consider the data generating process

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

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

where $\varepsilon_1, \ldots, \varepsilon_T, \eta_1, \ldots, \eta_T$ are independent. We set $\omega = 0.5$, $\kappa = 1$, $\nu = 1$, $\sigma_\varepsilon = \sigma_\eta = 1$, $z_0 = 0$ and $y_0 = \nu_c$. The adjustment coefficient is either $\alpha = -1$ or $\alpha = -0.2$.

We study the performance of two-sided, 5% level t-tests for the cointegration coefficient $\kappa$ when computed using OLS and LTS. For this, 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 $\Delta y_t$ 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 $\sigma_\varepsilon^2$ are degrees of freedom corrected. We test the hypothesis $\mathsf{H}_0 : \kappa = 1$ indirectly using $t_{\kappa,s} = (\hat{\kappa}_s - 1)/\mathsf{s.e.}(\hat{\kappa}_s)$ where the standard errors $\mathsf{s.e.}(\hat{\kappa}_s) = \mathsf{s.e.}(\hat{\psi}_s/\hat{\alpha}_s)$ vary with s and are obtained using the $\delta$-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 $\mathsf{s.e.}(\hat{\kappa}_s) = (s_s^2 \hat{D}_s' M_s^{-1} \hat{D}_s)^{1/2}$.

We vary the sample size and the magnitude of the outliers as follows. Let $\zeta_T$ indicate the good observations while $\zeta_T^c$ 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 $\mathrm{o}(\sqrt{T}/\log T)$ 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. The number of repetitions is $10^4$ giving a Monte Carlo standard deviation of 0.002.

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

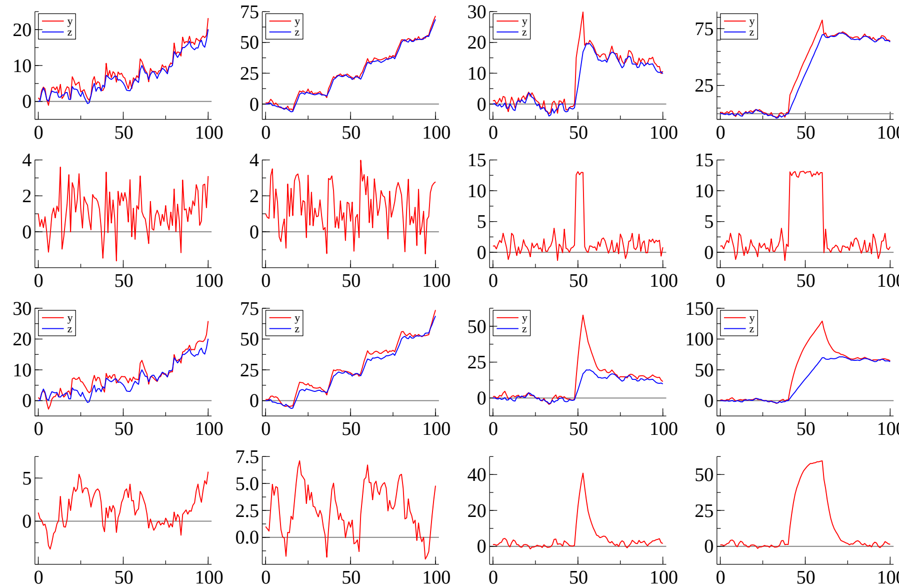


Figure 1: Examples of data generating processes.

Table 1: Outliers distributed in 5 episodes.

| | $T-h$ | $\sqrt{T}/2$ | $\sqrt{T}$ | $2\sqrt{T}$ | $\sqrt{T}/2$ | $\sqrt{T}$ | $2\sqrt{T}$ |
|---|---|---|---|---|---|---|---|
| method | $T$ | $\alpha=-1$ | | | $\alpha=-0.2$ | | |
| 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 |

In Table 1, the $T-h$ outliers occur in $G=5$ episodes and in both $\varepsilon_t$ and $\eta_t$. 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

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

Table 2: Outliers in one central episode

| method | $T-h$ / $T$ | $\sqrt{T}/2$ $\alpha=-1$ | $\sqrt{T}$ | $2\sqrt{T}$ | $\sqrt{T}/2$ $\alpha=-0.2$ | $\sqrt{T}$ | $2\sqrt{T}$ |
|---|---|---|---|---|---|---|---|
| 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 |

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 $\mathrm{o}(\sqrt{T}/\log T)$ 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.

# 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 (Aron et al., 2012). We use the R (R Core Team, 2024) package robustbase for LTS estimation and PcGive (Doornik and Hendry, 2022) 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.

[1]Household data released from 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.

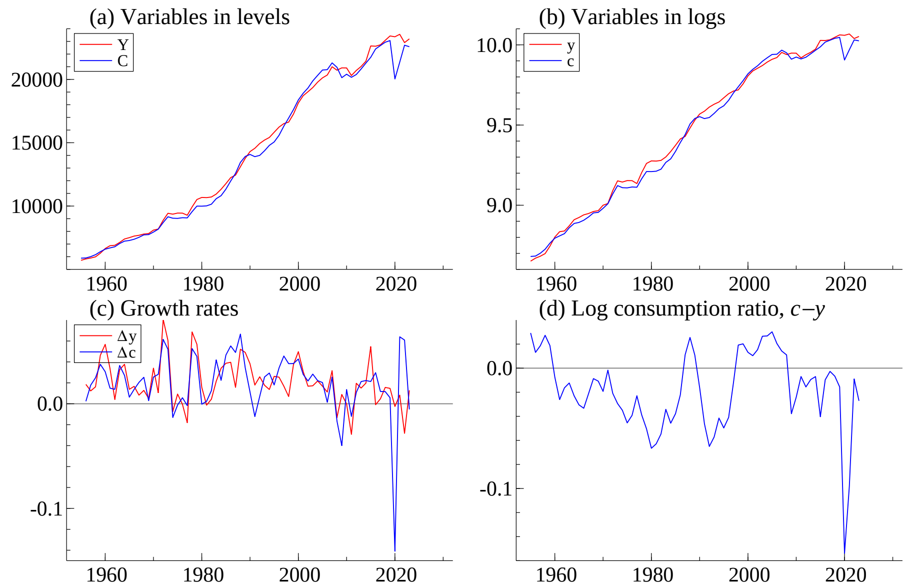


Figure 2: UK individual income and consumption expenditure.

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

$$\underset{(s.e.)}{\Delta\widehat{c}_t} = -\underset{(0.113)}{0.402}c_{t-1} + \underset{(0.142)}{0.460}y_{t-1} - \underset{(0.00130)}{0.00014}\,t \tag{7.1}$$
$$- \underset{(0.510)}{0.492} + \underset{(0.158)}{0.796}\Delta y_t + \underset{(0.132)}{0.223}\Delta c_{t-1} - \underset{(0.182)}{0.179}\Delta y_{t-1},$$
$$\widehat{\sigma} = 0.0241, \qquad \widehat{\ell} = 158.021, \qquad T = 67,$$
$$\mathsf{F}_{ar1-2}(2,58) = 2.46 \quad [p = 0.09], \qquad \chi^2_{normal}(2) = 39.7 \quad [p = 0.00],$$
$$\mathsf{F}_{arch1}(1,65) = 0.22 \quad [p = 0.64], \qquad \mathsf{F}_{hetero}(12,54) = 1.15 \quad [p = 0.34].$$

The OLS estimates are somewhat surprising. The consumption to income 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 $\mathsf{F}_{ar1-2}$ for residual second order autocorrelation (Godfrey, 1978), $\mathsf{F}_{arch1}$ for first order autocorrelated conditional heteroskedasticity (Engle, 1982), $\chi^2_{normal}$ for non-normality using the Doornik and Hansen (2008) version of the cumulant based test developed in 1880 by Thiele, and $\mathsf{F}_{hetero}$ for heteroskedasticity (White, 1980). These papers do not cover the cointegration setting. Cointegration is considered by Nielsen (2006) for $\mathsf{F}_{ar1-2}$ and Kilian and Demiroglu (2000) for $\chi^2_{normal}$. 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),

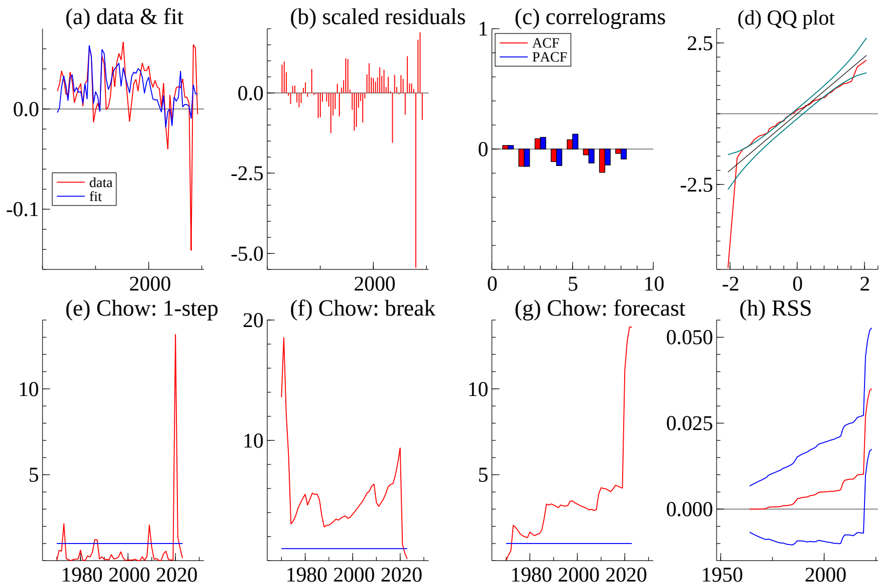


Figure 3: Mis-specification graphics for model estimated by OLS.

three versions of the Chow (1960) 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 (Engler and Nielsen, 2009), the 1-step Chow test (Nielsen and Whitby, 2015) and for the RSS plot (Nielsen and Sohkanen, 2011). 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 $T - h = 4$ and provide evidence in favour of this choice below. LTS finds outliers for 2009, 2020, 2021, 2023 matching the financial crisis and the pandemic. 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 4 impulse indicators, that is,

$$\begin{aligned}\underset{(s.e.)}{\Delta\widehat{c}_t} &= -\underset{(0.063)}{0.198}c_{t-1} + \underset{(0.078)}{0.178}y_{t-1} + \underset{(0.00079)}{0.00054}\,t \\ &+ \underset{(0.291)}{0.177} + \underset{(0.086)}{0.577}\Delta y_t + \underset{(0.107)}{0.491}\Delta c_{t-1} - \underset{(0.112)}{0.317}\Delta y_{t-1} \\ &- \underset{(0.014)}{0.043}I_{2009} - \underset{(0.014)}{0.148}I_{2020} + \underset{(0.022)}{0.089}I_{2021} - \underset{(0.016)}{0.063}I_{2023}, \\ \widehat{\sigma} &= 0.0127, \quad \widehat{\ell} = 203.342, \quad T = 67, \\ \mathsf{F}_{ar1-2}(2,54) &= 0.40 \quad [p = 0.67], \qquad \chi^2_{normal}(2) = 0.72 \quad [p = 0.70], \\ \mathsf{F}_{arch1}(1,65) &= 0.00 \quad [p = 0.96], \qquad \mathsf{F}_{hetero}(12,50) = 2.03 \quad [p = 0.04].\end{aligned} \tag{7.2}$$

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 §5.3. Outlier selection is considered by Berenguer-Rico and Nielsen (2023) for $\chi^2_{normal}$ and

Berenguer-Rico and Wilms (2021) for $\mathsf{F}_{hetero}$. The theory suggests that it is plausible that all tests are valid with LTS estimation using the assumptions in §4. None of the mis-specification test reject the model.

Figure 4 gives mis-specification graphics, which support the model.

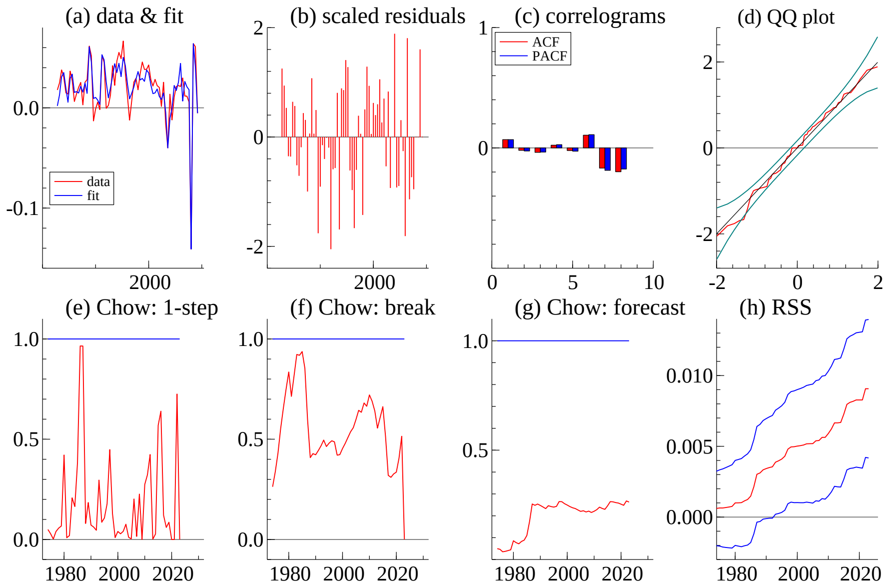


Figure 4: Mis-specification graphics for model estimated by LTS.

The theory presented in Section 5.3 shows that the $t$-statistics formed by dividing coefficients in (7.2) by their standard errors have the same asymptotic distributions as in models without outliers and outlier selection.

The hypothesis of no cointegration is tested by the coefficient on $c_{t-1}$ giving t-statistic $-0.198/0.063 = -3.14$. This is close to the 10% critical value of $-3.39$ (Banerjee et al., 1998) thus rejecting the hypothesis of no cointegration. Alternatively, a likelihood ratio test could be used (Harbo et al., 1998).

The equilibrium correction form of the model is

$$\underset{(s.e.)}{\Delta\widehat{c}_t} = -\underset{(0.063)}{0.198}\Big\{c_{t-1} - \underset{(0.180)}{0.897}y_{t-1} - \underset{(0.0042)}{0.0027}\,t\Big\} + \cdots . \tag{7.3}$$

The t-statistic for homogeneity in the consumption function is $(1-0.897)/0.180 = 0.57$, 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:

$$\underset{(s.e.)}{\Delta\widehat{y}_t} = \underset{(0.089)}{0.259}(c_{t-1} - y_{t-1}) - \underset{(0.00014)}{0.00028}\, t + \underset{(0.007)}{0.030} + \underset{(0.160)}{0.073}\Delta y_{t-1} + \underset{(0.160)}{0.177}\Delta c_{t-1}$$

$$- \underset{(0.021)}{0.004} I_{2009} - \underset{(0.020)}{0.012} I_{2020} + \underset{(0.033)}{0.062} I_{2021} - \underset{(0.024)}{0.004} I_{2023},$$

$$\widehat{\sigma} = 0.0196, \qquad \widehat{\ell} = 173.097, \qquad T = 67,$$

$$\mathsf{F}_{ar1-2}(2, 56) = 0.65 \quad [0.52], \qquad \chi^2_{normal}(2) = 4.13 \quad [0.25],$$

$$\mathsf{F}_{arch1}(1, 65) = 1.37 \quad [0.25], \qquad \mathsf{F}_{hetero}(8, 54) = 1.29 \quad [0.28].$$

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 $0.259/0.089 = 2.92$ which is large relative to a normal distribution. This is evidence against weak exogeneity and gives some doubt about the validity of the inferences in the single equation ADL. Thus, a system VAR analysis may be better. This is not unexpected given the usual difficulties in modelling consumption functions.

***Estimating the number of good observations***. Next, we consider two formal methods for estimating $h$, which give support to the chosen $h = 4$ above. 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. Nonetheless, both methods guided the choice of 4 outliers above.

The first method follows Berenguer-Rico et al. (2023), who suggest to estimate the model by LTS for different values of $h$, compute the normality test statistic and then minimize over $h$. For boundedness, we can have no more than $T/3$ outliers, see (4.8), so that we should search over $h \geq 2T/3 = 2 * 67/3 \approx 45$. For inference, the theory requires that there are no more than $\sqrt{T}/\log T$ outliers, see (4.12), while the simulation study shows good performance with as many as $\sqrt{T}/2$ outliers. Hence, we search over $h \geq T - \sqrt{T}/2 = 67 - \sqrt{67}/2 \approx 63$.

Table 3 reports normality test statistics for a wide range of values of $h$. Searching over $h \geq 63$ delivers a minimizer of 2.9 at $h = 63$ corresponding to 4 outliers, which supports the above analysis. Due to different implementations this local minimum value is different from the value in (7.2). Starting at $h = 45$ gives a minimizer of 0.6 at $h = 59$. The local minimum at $h = 59$ could perhaps be understood in the context of the scaled residuals and the QQ plot shown in Figure 4(b,d), where we see a slight left-skewness in the larger residuals. A global minimum of 0.1 is obtained at $h = 41$ but this value is outside the boundedness region for LTS. Thus, to achieve reliable inference, we search over $h \geq T - \sqrt{T}/2 \approx 63$, which delivers the 4 outliers considered in the above analysis.

The second method uses Impulse Indicator Saturation in PcGive (Doornik, 2009). The non-indicator variables in (7.2) were not selected over. Choosing a user defined gauge (0.5%) 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 the Brexit referendum.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $h$ | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 |
| $\chi^2_{normal,h}$ | 169.5 | 0.1 | 6.0 | 67.7 | 55.8 | 14.0 | 8.9 | 54.3 | 7.7 | 0.9 |
| $h$ | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 |
| $\chi^2_{normal,h}$ | 15.4 | 15.4 | 23.6 | 4.8 | 787.5 | 8.8 | 292.3 | 5.1 | 1.8 | 0.6 |
| $h$ | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | | |
| $\chi^2_{normal,h}$ | 117.0 | 21.1 | 20.0 | 3.0 | 8.8 | 20.5 | 12.6 | 116.6 | | |

Table 3: Normality test statistics for different values of $h$

# 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 (Rousseeuw and Leroy, 1987), a method based on the normal cumulants (Berenguer-Rico et al., 2023) and a bootstrap method (Heng and Lange, 2025). Related algorithms include the Forward Search (Atkinson and Riani, 2000) and the Impulse Indicator Saturation implemented as Autometrics in PcGive (Doornik, 2009), as Gets in R (Pretis et al., 2018) and in the Eviews software. Some asymptotic theory is available for data generating processes with normal errors and no outliers (Hendry et al., 2008; Johansen and Nielsen, 2009, 2016a,b). 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 (1988, 1995). We would then need a systems version of Least Trimmed Squares. One approach is to use the Minimum Covariance Determinant approach (Rousseeuw, 1985). Extensions are available to a VAR (Croux and Joossens, 2008) and to a VAR with different outliers in different equations (Raymaekers and Rousseeuw, 2024). Theory for these methods would be desirable.

# A General asymptotic theory for LTS

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

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

The $\beta$ 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 5.2 in the of this appendix.

The sequence of data generating processes has common $\beta$, $\sigma$. 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 $\zeta_T$ 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*** *$\varepsilon_t$ 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*

$$F_{Th}(a) = \max_{\zeta : \#\zeta = h} \sup_{\delta : |\delta| = 1} h^{-1} \sum_{t \in \zeta} 1_{(|x_{tT}' \delta| \leq a)}. \tag{A.2}$$

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

$$\lim_{(a,T) \to (0,\infty)} \mathsf{P}\{F_{Th}(a) > \xi\} = 0, \tag{A.3}$$

*that is $\forall \epsilon > 0$, $\exists (a_0, n_0) > 0$: $\forall a \leq a_0, T \geq T_0$ then $\mathsf{P}\{F_{Th}(a) > \xi\} < \epsilon$.*
($v$) ***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* $\forall 0 < \delta < 1$, $\exists 0 < r < 1$: $x_{(T - \lfloor T^r \rfloor)}^2 / x_{(T)}^2 \leq \delta\{1 + \mathrm{o}_{\mathsf{P}}(1)\}$.
($vii$) ***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 $h$ (Berenguer-Rico and Nielsen, 2024), but a distributional assumption seems necessary for estimating $h$ (Berenguer-Rico et al., 2023). 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 $\zeta$ (Johansen and Nielsen, 2019). Condition ($v$) has a trade-off with ($v$), ($vi$) which limit the magnitude of the regressors.

We quote the general LTS asymptotic theory (Berenguer-Rico and Nielsen, 2024). As before, let $\mathcal{M}_T$ denote the set of minimizers $\zeta$ of $\hat{\sigma}_\zeta^2$.

**Theorem A.1.** ***Boundedness***. *Suppose Assumption A.1($i, ii, iv$). Then the LTS estimator $\hat{\beta}$ for (A.1) 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 (A.1) satisfy:*
($a$) ***Consistent selection by $\hat{\zeta}$***. $\forall 0 < \beta < 1$: $\max_{\zeta \in \mathcal{M}_T} \#(\zeta \cap \zeta_T^c)/h = \mathrm{O}_{\mathsf{P}}(h^{\beta - 1})$.
($b$) ***Expansion for $\hat{\sigma}^2$***. $\max_{\zeta \in \mathcal{M}_T} h^{1/2} |\hat{\sigma}_\zeta^2 - \hat{\sigma}_{\zeta_n}^2| = \mathrm{o}_{\mathsf{P}}(1)$.
($c$) ***Expansion for $\hat{\beta}$***.

$$\max_{\zeta \in \mathcal{M}_T} \Big| \Big( \sum_{i \in \zeta} x_{tT} x_{tT}' \Big)^{1/2} (\hat{\beta}_\zeta - \beta) - \Big( \sum_{i \in \zeta_T} x_{tT} x_{tT}' \Big)^{1/2} (\hat{\beta}_{\zeta_T} - \beta) \Big| = \mathrm{o}_{\mathsf{P}}(1).$$

*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 (2024) provide a set of alternative conditions.

# B Proofs

## B.1 Proof of representation

*Proof of Theorem 3.1. 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

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

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

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

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

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

Pre-multiply by the inverse of $\boldsymbol{\alpha}_\perp'\boldsymbol{\Psi}\boldsymbol{\beta}_\perp$, which exists by Assumption 3.1. 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

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

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

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

Sum over $t$ to get the desired expression.

*Part $(b, i)$.* The normality assumption and assumption 3.1 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 $\boldsymbol{y}_{i+1}^*$, $\boldsymbol{y}_t^*$.

*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 $\boldsymbol{\Gamma}_j^\dagger = \boldsymbol{\Gamma}_j + \boldsymbol{\alpha\beta}'$ to get

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

This matches the formulation in Johansen (1988). 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

$$\boldsymbol{y}_t^\dagger = \boldsymbol{Y}^\dagger \boldsymbol{y}_{t-1}^\dagger + \boldsymbol{e}_p \boldsymbol{A}\boldsymbol{\varepsilon}_t. \tag{B.2}$$

Apply the autoregressive equation $k$ times to get

$$\boldsymbol{y}_t^\dagger = \begin{pmatrix} I_p & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & I_p & * \\ 0 & \cdots & 0 & \boldsymbol{\beta}' \end{pmatrix} \begin{pmatrix} \boldsymbol{A}\boldsymbol{\varepsilon}_t \\ \vdots \\ \vdots \\ \boldsymbol{A}\boldsymbol{\varepsilon}_{t-k+1} \end{pmatrix} + \boldsymbol{Y}^{\dagger k} \boldsymbol{y}_{t-k}^\dagger, \tag{B.3}$$

where $*$ represents quantities that are not of importance. As $\boldsymbol{\beta}'$ has full row rank by Assumption 3.1, so does the first matrix in (B.3). The vector of errors in (B.3) has an invertible covariance matrix whenever $t-k+1 > \underline{t}$. It is also independent of the $\sigma$-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, $\mathsf{Var}(\boldsymbol{y}_t^\dagger \mid \mathcal{G}_{t-k})$ is constant and invertible for $t \geq \underline{t}+k$.

We now concatenate $\boldsymbol{y}_t^\dagger$ with $\Delta x_{t+1}^*$. By the model equation (B.1), we have

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

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

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

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

Finally, the $\sigma$-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 $\Delta x_{t+1}^*$, or by $\widetilde{\boldsymbol{y}}_s^*$ for $\underline{t}-k < s < t-k$ by a linear, bijective transformation. □

## B.2 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

$$\Delta z_t, \boldsymbol{x}_{t-1}, \Delta \boldsymbol{x}_{t-1}, \ldots, \Delta \boldsymbol{x}_{t-k+1}, 1, \tag{B.4}$$

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 (B.4) 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 (3.6). The Granger-Johansen representation Theorem 3.1 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 (B.4) are a linear function of

$$x_{tT} = \big(\Delta z_t \,,\, \boldsymbol{y}_{t-1}^{*\prime} \,,\, T^{-1/2} \sum_{s=1}^{t-1} \boldsymbol{\varepsilon}_s' \,,\, 1\big)', \tag{B.5}$$

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 $\Delta y_t, \Delta z_t, \Delta \boldsymbol{y}_{t-1}^*$ concatenate as $\widetilde{\boldsymbol{y}}_{t-1}^*$ we get that $x_{tT}$ is a subvector of

$$\widetilde{x}_{tT} = \big(\widetilde{\boldsymbol{y}}_{t-1}^{*\prime} \,,\, T^{-1/2} \sum_{s=1}^{t-1} \boldsymbol{\varepsilon}_s' \,,\, t/T \,,\, 1\big)', \tag{B.6}$$

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

$$\overline{x}_{tT} = \big\{0, I_{(k+1)p+r+1}\big\} \widetilde{x}_{tT}. \tag{B.7}$$

## B.3 Conditions for boundedness

We check the boundedness Assumption A.1$(iv)$ for $x_{tT}$ defined in (B.5). This is a subvector of $\widetilde{x}_{tT}$ in (B.6). 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*

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

*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 $\delta$ is a unit vector of length $\dim x$, then $\widetilde{\delta} = S\delta$ is a unit vector of length $\dim \widetilde{x}$. Therefore,

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

Divide by $h$ and take maximum over $h$-sets $\zeta$ to get that $F_{Th}^x(a) \le F_{Th}^{\widetilde{x}}(a)$. □

**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 $F_{tT}^{\widetilde{x}}$ function for $\widetilde{x}_{tT}$ vanishes within a good episode indexed by $t = 1, \dots, T$. We combine the separate arguments for stationary processes, random walks and linear trends in Johansen and Nielsen (2019).

**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) $\exists C > 0$: $\max_{q_T < t \le T} \sup_{\delta_u : |\delta_u|=1} \sup_{\nu \in \mathbb{R}} \mathsf{P}(|u_t'\delta_u + \nu| < \alpha) \le \alpha C$,*
*(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) $\exists C > 0$: $\max_{q_T < t \le T} \sup_{\delta_v : |\delta_v|=1} \sup_{\nu \in \mathbb{R}} \mathsf{P}(|v_t'\delta_v/\sqrt{t} + \nu| < \alpha) \le \alpha C$.*
*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

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

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

*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 $A \ge 4$,

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

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

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

Take supremum and then expectation to bound

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

This is small for large $A$ since $u_t = \mathrm{O}_{\mathsf{P}}(1)$ uniformly in $t$ while $\max_{t \le T} |v_t/\sqrt{T}| = \mathrm{O}_{\mathsf{P}}(1)$ 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 $\delta$ as

$$\delta = \big(\delta_u' \cos\psi \cos\phi \cos\theta\,,\ \delta_v' \sin\psi \cos\phi \cos\theta\,,\ \sin\phi \cos\theta\,,\ -\sin\theta\big)', \tag{B.8}$$

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 $\delta$ and analyze 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 (B.8) 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 Johansen and Nielsen (2019, Lemma 3.1), we find for $a \leq 1/2$ and $|\theta| \leq \pi/2$ that $|-\sin\theta + z_{tT}'\delta_{\theta=0}\cos\theta| \leq a$ implies $1/\cos\theta \leq 2(1+|z_{tT}'\delta_{\theta=0}|) \leq 2(1+A)$.

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

$$\sup_\delta R_{T\delta}(a,0) \leq \max_{\ell \leq L} \big\{ R_{T\delta_\ell}(a,0) + \sup_{\delta:|\delta-\delta_\ell|\leq\epsilon} \big| R_{T\delta}(a,0) - R_{T\delta_\ell}(a,0) \big| \big\}. \tag{B.9}$$

*Oscillation term in (B.9).* By the triangle inequality,

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

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

$$\begin{aligned} \big| R_{T\delta}(a,0) - R_{T\delta_\ell}(a,0) \big| &\leq \frac{1}{T} \sum_{t>q_T} \big\{ 1_{(|z_{tT}'\delta_\ell - a|\leq\epsilon A, |z_{tT}|\leq A)} + 1_{(|z_{tT}'\delta_\ell + a|\leq\epsilon A, |z_{tT}|\leq A)} \big\} \\ &= R_{T\delta_\ell}(\epsilon A, a) + R_{T\delta_\ell}(\epsilon A, -a). \end{aligned} \tag{B.10}$$

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

$$R_{T\delta}(\alpha,\mu) \leq \frac{1}{T} \sum_{t>q_T} 1_{(|z_{tT}'\delta_\ell - \mu|\leq\alpha)} = S_{T\delta}(\alpha,\mu). \tag{B.11}$$

Return to the chaining inequality (B.9), apply the bounds (B.10), (B.11) 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 (B.11) noting $T - q_T \leq T$ to get, for any $\alpha \geq 0$, $\mu \in \mathbb{R}$,

$$\begin{aligned} \mathsf{P}\big\{ \max_{\ell\leq L} S_{T\delta_\ell}(\alpha,\mu) > \eta \big\} &\leq \frac{L}{T\eta} \max_{\ell\leq L} \sum_{t>q_T} \mathsf{P}\big( \big| z_{tT}'\delta_\ell - \mu \big| \leq \alpha \big) \\ &\leq \frac{L}{\eta} \max_{\ell\leq L} \max_{t>q_T} \mathsf{P}\big( \big| z_{tT}'\delta_\ell - \mu \big| \leq \alpha \big). \end{aligned} \tag{B.12}$$

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

$$\frac{z_{tT}'\delta - \mu}{\sin\phi\cos\theta} = \frac{t}{T} + \nu_1 \quad \text{with} \quad \nu_1 = \Big( u_t'\delta_u \cos\psi + \frac{1}{\sqrt{T}} v_t'\delta_v \sin\psi \Big) \tan\phi - \frac{\sin\theta + \mu}{\sin\psi\cos\theta}.$$

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

$$\big( \big| z_{tT}'\delta - \mu \big| \leq \alpha \big) \subset \Big( \Big| \frac{z_{tT}'\delta - \mu}{\sin\phi\cos\theta} \Big| \leq \widetilde{\alpha}_1 \Big) = \Big( \Big| \frac{t}{T} + \nu_1 \Big| \leq \widetilde{\alpha}_1 \Big).$$

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 $(\alpha, T) \to (0, \infty)$,

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

*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

$$\frac{z_{tT}'\delta - \mu}{\cos\psi\cos\phi\cos\theta} = u_t'\delta_u + \nu_2 \quad \text{with} \quad \nu_2 = \frac{1}{\sqrt{T}}v_t'\delta_v \tan\psi + \frac{t\tan\phi}{T\cos\psi} - \frac{\sin\theta + \mu}{\cos\psi\cos\phi\cos\theta}.$$

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

$$\big(\big|z_{tT}'\delta - \mu\big| \leq \alpha\big) \subset \Big(\Big|\frac{z_{tT}'\delta - \mu}{\cos\psi\cos\phi\cos\theta}\Big| \leq \widetilde{\alpha}_2\Big) = \big(\big|u_t'\delta_u + \nu_2\big| \leq \widetilde{\alpha}_2\big).$$

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

$$\mathsf{P}\big(\big|z_{tT}'\delta - \mu\big| \leq \alpha\big) \leq \max_{q_T < t \leq T} \sup_{\delta_u : |\delta_u| = 1} \sup_{\nu_2 \in \mathbb{R}} \mathsf{P}\big(\big|u_t'\delta_u + \nu_2\big| \leq \widetilde{\alpha}_2\big) \leq \widetilde{\alpha}_2 C. \tag{B.13}$$

Combine the inequalities (B.12), (B.13) with the definition of $\widetilde{\alpha}_2$ to get

$$\mathsf{P}\big\{\max_{\ell \leq L} S_{T\delta_\ell}(\alpha, \mu) > \eta\big\} \leq \frac{L}{\eta}\widetilde{\alpha}_2 C = \frac{L}{\eta}\alpha 4\big(1 + A\big)C \to 0, \tag{B.14}$$

as $(\alpha, T) \to (0, \infty)$ since $\eta, L, A, C$ 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

$$\frac{z_{tT}'\delta - \mu}{\sin\psi\cos\phi\cos\theta} = \frac{1}{\sqrt{T}}v_t'\delta_v + \nu_3 \quad \text{with} \quad \nu_3 = u_t'\delta_u \cot\psi + \frac{t\tan\phi}{T\sin\psi} - \frac{\sin\theta + \mu}{\sin\psi\cos\phi\cos\theta},$$

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

$$\big(\big|z_{tT}'\delta - \mu\big| \leq \alpha\big) \subset \Big(\Big|\frac{z_{tT}'\delta - \mu}{\sin\psi\cos\phi\cos\theta}\Big| \leq \widetilde{\alpha}_3\Big) = \big(\big|v_t'\delta_v/\sqrt{T} + \nu_3\big| \leq \widetilde{\alpha}_3\big).$$

Taking probability, normalizing by $\sqrt{T/t}$ and applying condition $(ii, b)$, we find a $C > 0$ exists such that, uniformly in $q_T < t \leq T$, $\delta_v : |\delta_v| = 1$, $\nu_3 \in \mathbb{R}$

$$\mathsf{P}\big(\big|z_{tT}'\delta - \mu\big| \leq \alpha\big) \leq \mathsf{P}\Big(\Big|\frac{v_t'\delta_v}{\sqrt{T}} + \nu_3\Big| \leq \widetilde{\alpha}_3\Big) = \mathsf{P}\Big(\Big|\frac{v_t'\delta_v}{\sqrt{t}} + \nu_3\Big| \leq \widetilde{\alpha}_3\frac{\sqrt{T}}{\sqrt{t}}\Big) \leq \widetilde{\alpha}_3 C\frac{\sqrt{T}}{\sqrt{t}}. \tag{B.15}$$

Combine the inequalities (B.12), (B.15) 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

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

as $(\alpha, T) \to (0, \infty)$ since $\eta, L, A, C$ 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*

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

*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 Berenguer-Rico and Nielsen (2024, eq. 4.2). As $F_{Th}$ sums over $\zeta$, bound $\zeta \subset \zeta_T \cup \zeta_T^c$ and then bound the indicators on $\zeta_T^c$ 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, $\underline{t}_g < t \leq \overline{t}_g$ 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 $h_g = \overline{t}_g - \underline{t}_g$ noting $h_g \leq h_{VAR}$ and where $F_g(a) = \sup_{\delta:|\delta|=1} h_g^{-1} \sum_{t=\underline{t}_g+1}^{\overline{t}} 1_{(|\delta' x_{tT}|\leq a)}$.

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

$$F_{Th}(a) \leq \frac{h - h_{VAR}}{h} + \sum_{g=1}^{G} 1_{(h_g/\sqrt{T}\text{ diverges})} F_g(a) + \mathrm{o}(1). \tag{B.16}$$

*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 $\max_{\underline{t}_g+q_T<t\leq\overline{t}_g} \mathsf{P}(|\widetilde{\boldsymbol{y}}_t^*| > C) \leq \epsilon$. Now, $\widetilde{\boldsymbol{y}}_t^*$ satisfies the VARMA equation (3.9). The Granger-Johansen representation Theorem 3.1$(b, i)$ 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 $\underline{t}_g < t \leq \overline{t}_g$. Further, by the VARMA equation, $\widetilde{\boldsymbol{y}}_{\underline{t}_g}^* = \mathrm{O_P}(1 + \max_{t\notin\zeta_{VAR,T}} |\boldsymbol{\varepsilon}_t|)$. This is $\mathrm{O_P}(T^c)$ by (4.6). Further, $\rho = \max|\mathrm{eigen}(\widetilde{\boldsymbol{Y}})| < 1$ by Assumption 3.1. Now, let $q_T = T^{1/4}$ so that $q_T/h_g \leq q_T/T \to 0$, but $\log(\|\widetilde{\boldsymbol{Y}}\|^{q_T}\widetilde{\boldsymbol{y}}_{\underline{t}_g}^*) = T^{1/4}\log\|\widetilde{\boldsymbol{Y}}\| + c\mathrm{O_P}(\log T) \to -\infty$, so that $\widetilde{\boldsymbol{Y}}^{q_T}\widetilde{\boldsymbol{y}}_{\underline{t}_g}^* = \mathrm{o_P}(1)$. It follows that $\widetilde{\boldsymbol{y}}_t^* = \mathrm{O_P}(1)$ uniformly in $\underline{t}_g + q_T < t \leq \overline{t}_g$ and the run-in period $q_T$ is vanishing. Thus, condition $(i, a)$ is satisfied.

Condition $(i, b)$. The Granger-Johansen representation Theorem 3.1$(b, ii)$ shows that

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

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 $v_t = \sum_{t=\underline{t}_g+1}^{t} \boldsymbol{\varepsilon}_t$. Let $T_g = \overline{t}_g - \underline{t}_g$ 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 (Billingsley, 1968).

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 $\mathsf{P}(|v'_{t-\underline{t}_g}\delta_v/(t-\underline{t}_g)^{1/2}|+\nu<\alpha)\le\alpha C$ uniformly in $t>\underline{t}_g$ and $\nu\in\mathbb{R}$. 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 (B.16) to finish the proof. $\square$

*Proof of Theorem 5.1.* 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 $\xi$ 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

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

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

**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*

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

*Let data be generated by $\alpha=-1$, $\nu_c=0$, $\sigma=1$ 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*

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

*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 $a\ge 0$ that*

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

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

## B.4 Conditions for consistent selection and expansions

**Lemma B.5.** *Consider the setup in Section 4. Then*
$(a)$ $\max_{1\le t\le T}|\widetilde{x}_{tT}|^2\le\max_{1\le t\le T}|\widetilde{\boldsymbol{y}}_t^*|^2+\max_{1\le t\le T}|T^{-1/2}\sum_{s=1}^T\boldsymbol{\varepsilon}_s|^2+2.$
$(b)$ $\max_{1\le t\le T}|\widetilde{\boldsymbol{y}}_t^*|=\mathrm{O}_{\mathsf{P}}(\sqrt{\log T}).$
$(c)$ $\max_{1\le t\le T}|T^{-1/2}\sum_{s=1}^T\boldsymbol{\varepsilon}_s|=\mathrm{O}_{\mathsf{P}}(1)+(\#\zeta_{VAR,T})\mathrm{O}_{\mathsf{P}}\{\sqrt{(\log T)/T}\}.$

*Proof of Lemma B.5.* $(a)$ The vector $\widetilde{x}_{tT}$ has components $\widetilde{\boldsymbol{y}}^*_{t-1}$ and $T^{-1/2}\sum_{s=1}^{t-1}\boldsymbol{\varepsilon}_s$ as well as $t/T$ and 1. 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 $\sum_{s=0}^{\infty}\|\widetilde{\boldsymbol{Y}}\|^s\max_{1\le t\le T}|\widetilde{\boldsymbol{\varepsilon}}^*_t|$. The geometric sum is convergent and the VAR errors are $\mathrm{O}_{\mathsf{P}}(\sqrt{\log T})$ due to (3.10), (4.10). Thus, $\max_{1\le t\le T}|\widetilde{\boldsymbol{y}}^*_t| = \mathrm{O}_{\mathsf{P}}(\sqrt{\log T})$.

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

$$\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\boldsymbol{\varepsilon}_s = \frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\boldsymbol{\varepsilon}_s 1_{(s\in\zeta_{VAR,T})} + \frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\boldsymbol{\varepsilon}_s 1_{(s\notin\zeta_{VAR,T})}. \tag{B.17}$$

The first term in (B.17) is a standard, normal random walk and converges to a Brownian motion when embedded in $C[0,1]$ with the uniform metric (Billingsley, 1968). In particular, the maximum of its absolute value is $\mathrm{O}_{\mathsf{P}}(1)$. We bound the second term in (B.17). It has at most $\#\zeta^c_{VAR,T}$ elements, which have order $\max_{t\le T}|\boldsymbol{\varepsilon_t}| = \mathrm{O}_{\mathsf{P}}(\sqrt{\log T})$ 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* $\widetilde{x}_{tT}$*,* $\overline{x}_{tT} = (0, I_q)\widetilde{x}_{tT}$ *defined in (B.6), (B.7) satisfy*

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

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

*Proof of Lemma B.6.* $(a)$ We have $\zeta_{VAR,T}\subset\zeta_T$ by assumption so that $\zeta^c_T\subset\zeta^c_{VAR,T}$ and $T^{-1}\sum_{t\in\zeta^c_T}|\widetilde{x}_{tT}|^2 \le T^{-1}\sum_{t\in\zeta^c_{VAR,T}}|\widetilde{x}_{tT}|^2 = \mathcal{S}_T$. Lemma B.5 shows $\max_{t\le T}|\widetilde{x}_{tT}|^2 = \mathrm{O}_{\mathsf{P}}(\log T)\{1+(\#\zeta^c_{VAR,T})^2/T\}$. Thus, $\mathcal{S}_T \le T^{-1}(\#\zeta^c_{VAR,T})\mathrm{O}_{\mathsf{P}}(\log T)\{1+(\#\zeta^c_{VAR,T})^2/T\}$. This vanishes when $\#\zeta^c_{VAR,T} = \mathrm{o}\{T^{2/3}/(\log T)^{1/3}\}$, which is implied by (4.12).

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

**Lemma B.7.** *Consider the sequence of data generating process of Section 4. Let* $\widetilde{\boldsymbol{y}}^\dagger_g$ *denote stationary solution of the normal VARMA equation (3.9) for* $\underline{t}_g < t \le \overline{t}_g$*. Let*

$$\widetilde{x}^\dagger_{tT} = \begin{pmatrix} \sum_{s=0}^{t-\underline{t}_g-1}\widetilde{\boldsymbol{Y}}^s\widetilde{\boldsymbol{\varepsilon}}^*_{t-s} + \widetilde{\boldsymbol{Y}}^{t-\underline{t}_g}\widetilde{\boldsymbol{y}}^\dagger_g \\ \frac{1}{\sqrt{T}}\sum_{s=1}^{t-1}\boldsymbol{\varepsilon}_s 1_{(s\in\zeta_{VAR,T})} \\ t/T \\ 1 \end{pmatrix}, \qquad \overline{x}^\dagger_{tT} = \big(0, I_{q+p+2}\big)\widetilde{x}^\dagger_{tT}.$$

*Then* $\widetilde{x}_{tT}$*,* $\overline{x}_{tT} = (0, I_q)\widetilde{x}_{tT}$ *defined in (B.6), (B.7) satisfy*

$(a)$ $\displaystyle\frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\big|\widetilde{x}_{tT} - \widetilde{x}^\dagger_{tT}\big|^2 = \mathrm{o}_{\mathsf{P}}(1).$

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

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

*Proof of Lemma B.7.* $(a)$ Write $\widetilde{x}_{tT} = \widetilde{x}_{tT}^{\dagger} + (\widetilde{x}_{tT} - \widetilde{x}_{tT}^{\dagger})$. Recalling the definition of $\widetilde{x}_{tT}$ and $\widetilde{\boldsymbol{y}}_t^*$ from (B.6), (3.9) we find for $\underline{t}_g < t \leq \overline{t}_g$ that $\widetilde{x}_{tT} - \widetilde{x}_{tT}^{\dagger} = (z_{1t}', z_{2t}', 0, 0)'$ where $z_{1t} = \widetilde{\boldsymbol{Y}}^{t-\underline{t}_g}(\widetilde{\boldsymbol{y}}_{\underline{t}_g}^* - \widetilde{\boldsymbol{y}}_g^{\dagger})$ 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

$$\Big\|\frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\big(\widetilde{x}_{tT} - \widetilde{x}_{tT}^{\dagger}\big)\big(\widetilde{x}_{tT} - \widetilde{x}_{tT}^{\dagger}\big)'\Big\| \leq \frac{1}{T}\sum_{t\in\zeta_{VAR,T}}\big|\widetilde{x}_{tT} - \widetilde{x}_{tT}^{\dagger}\big|^2 = Z_{1T} + Z_{2T}, \quad \text{(B.18)}$$

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

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

For $Z_{2T}$, which is an average 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, $|z_{2t}| \leq T^{-1/2}(\#\zeta_{VAR,T}^c)\max_{t\leq T}|\boldsymbol{\varepsilon}_t|$ uniformly in $t$. Here, we have $\max_{t\leq T}|\boldsymbol{\varepsilon}_t| = \mathrm{O_P}(\sqrt{\log T})$ by (4.10). Combine these bounds to get $Z_{2T} = \mathrm{O_P}\{(\#\zeta_{VAR,T}^c)^2\log T/T\}$. This term vanishes when vanishes when $\#\zeta_{VAR,T}^c = \mathrm{o_P}(\sqrt{T/\log T})$ which is implied by (4.12).

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

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

With this and the triangle inequality, we can bound

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

We apply this bound with $v_t = \widetilde{x}_{tT}^{\dagger}$, $w_t = \widetilde{x}_{tT} - \widetilde{x}_{tT}^{\dagger}$ 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 must argue that $T^{-1/2}\sum_{t\in\zeta_{VAR,T}}(\overline{x}_{tT} - \overline{x}_{tT}^{\dagger})\varepsilon_t = \mathrm{o_P}(1)$. This is a sum of martingale differences since $\varepsilon_t$ is standard normal and independent of $\Delta z_t$ and $\boldsymbol{x}_{t-s}$ for $s > 0$. By Lai and Wei (1982, Lemma 2) we get, for any $\delta > 0$ that

$$\sum_{t\in\zeta_{VAR,T}}\big(\overline{x}_{tT} - \overline{x}_{tT}^{\dagger}\big)\varepsilon_t \overset{a.s.}{=} \mathrm{O}(1) + \mathrm{o}\big\{(\log M)^{1+\delta}\big\} \quad \text{where} \quad M = \sum_{t\in\zeta_{VAR,T}}\big|\overline{x}_{tT} - \overline{x}_{tT}^{\dagger}\big|^2.$$

Here, $M$ vanishes by part (a). Normalize by $\sqrt{T}$ to get the desired result. □

**Lemma B.8.** *Consider the sequence of data generating process of Section 4. Let* $\widetilde{\boldsymbol{y}}_g^\dagger$ *denote stationary solution of the normal VARMA equation (3.9) for* $\underline{t}_g < t \le \overline{t}_g$. *Let* $W$ *be standard p-dimensional Brownian motion so that* $B = \sigma^{-1}(1, -\omega)\boldsymbol{A}W$ *is a standard univariate Brownian motion. Concatenate* $W_u, u, 1$ *as* $F_u$. *Let* $\widetilde{\boldsymbol{y}}_g^\dagger$ *denote stationary solution of the normal VARMA equation (3.9) for* $\underline{t}_g < t \le \overline{t}_g$. *Let* $\Sigma_{yy}$ *be the variance of* $(0, I_{r+kp-1})\widetilde{\boldsymbol{y}}_g^\dagger$. *Suppose* $h_{VAR} = \#\zeta_{VAR,T} \to \infty$ *as* $T \to \infty$. *Then*

$$\Big(\frac{1}{h_{VAR}} \sum_{t\in\zeta_{VAR,T}} \overline{x}_{tT}^\dagger \overline{x}_{tT}^{\dagger\prime}, \frac{1}{\sqrt{h_{VAR}}} \sum_{t\in\zeta_{VAR,T}} \overline{x}_{tT}^\dagger \varepsilon_t\Big) \overset{\mathsf{D}}{\to} \Big[\begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \int_0^1 F_u F_u' du \end{pmatrix}, \begin{pmatrix} N \\ \int_0^1 F_u dB_u \end{pmatrix}\Big], \tag{B.19}$$

*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 (1988) prove this for a univariate autoregression without deterministic terms. Chan (1989) extends this to include deterministic terms. Johansen (1995, Appendix B) extends this to VARs. □

*Proof of Theorem 5.2.* We apply Theorem A.1 and must check Assumption A.1. Parts $(i, ii, iv)$ were checked for Theorem 5.1.

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

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

Assumption A.1$(vi, b)$, first part. We have $\max_{t\le T}|x_{tT}|^2$ is bounded by $\mathrm{O}_\mathsf{P}(\log T) + \max_{1\le t\le T}|T^{-1/2}\sum_{s=1}^T \boldsymbol{\varepsilon}_t|^2$ by Lemma B.5$(a, b)$. The random walk term is $\mathrm{O}_\mathsf{P}(\log T)$ 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 $\mathrm{O}_\mathsf{P}(1)$ by Lemma B.5 using that $\#\zeta_{VAR,T}^c = T - h_{VAR} = \mathrm{O}(\sqrt{T})$ 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 (Watts et al., 1982). 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

$$S_{\zeta_T} = \Big(\frac{1}{\sqrt{h}}\sum_{i\in\zeta_T} \varepsilon_t x_{tT}'\Big)\Big(\frac{1}{h}\sum_{i\in\zeta_T} x_{tT}x_{tT}'\Big)^{-1}\Big(\frac{1}{\sqrt{h}}\sum_{i\in\zeta_T} x_{tT}\varepsilon_t\Big), \tag{B.20}$$

where $h = \#\zeta_T$. We show $S_{\zeta_T} = \mathrm{O}_\mathsf{P}(1)$. As $x_{tT}$ is a subset of $\overline{x}_{tT}$, which is a subset of $\widetilde{x}_{tT}$, the desired result follows by replacing $\zeta_T$ 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 $\overline{x}_{tT}^\dagger$, $\widetilde{x}_{tT}^\dagger$ using Lemma B.7, and using the convergence in Lemma B.8.

We can now apply Theorem A.1 to get

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

Finally, $x_{tT}=B_Tx_t$ 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 $\zeta_T$ by $\zeta_{VAR,T}$ by use of Lemma B.6. □

**Remark B.2.** *In Theorem 5.2, the square roots of the matrices $M=\sum_{t\in\zeta}x_tx'_t$ and $N=\sum_{t\in\zeta_T}x_tx'_t$ 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 $\Lambda$ such that $N=SS'$ and $M=S(I_{\dim x}+\Lambda)S'$ (Johansen, 1995, Lemma A.5). The elements $\lambda$ of $\Lambda$ 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*

$$\begin{aligned}D_\zeta&=\Big(\sum_{i\in\zeta}x_tx'_t\Big)^{1/2}\big(\hat\beta_\zeta-\beta\big)-\Big(\sum_{i\in\zeta_T}x_tx'_t\Big)^{1/2}\big(\hat\beta_{\zeta_T}-\beta\big)\\&=\Big(\sum_{i\in\zeta}x_tx'_t\Big)^{1/2}\Big(\sum_{i\in\zeta}x_tx'_t\Big)^{-1}\sum_{i\in\zeta}x_t\varepsilon_t-\Big(\sum_{i\in\zeta_T}x_tx'_t\Big)^{1/2}\Big(\sum_{i\in\zeta_T}x_tx'_t\Big)^{-1}\sum_{i\in\zeta_T}x_t\varepsilon_t\\&=S'\big(SS'\big)^{-1}\sum_{t\in\zeta}x_t\varepsilon_t-\big(I_{\dim x}+\Lambda\big)^{1/2}S'\big\{S\big(I_{\dim x}+\Lambda\big)S'\big\}^{-1}\sum_{t\in\zeta_T}x_t\varepsilon_t\end{aligned}$$

*and then eliminate terms to get*

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

*In the proof of Theorem 5.2 a rotated, normalized version of the regressors is used as in $x_{tT}=B_Tx_t$. We then get that $\sum_{t\in\zeta}x_{tT}x'_{tT}=B_TMB'_T=B_TS(I_{\dim x}+\Lambda)S'B'_T$ and $\sum_{t\in\zeta_T}x_{tT}x'_{tT}=B_TNB'_T=B_TSS'B'_T$. Argueing as above, we find*

$$D_{\zeta,T}=\Big(\sum_{i\in\zeta}x_{tT}x'_{tT}\Big)^{-1/2}\sum_{i\in\zeta}x_{tT}\varepsilon_t-\Big(\sum_{i\in\zeta_T}x_{tT}x'_{tT}\Big)^{-1/2}\sum_{i\in\zeta_T}x_{tT}\varepsilon_t=D_\zeta.$$

# References


Agullo, J., Croux, C., and Van Aelst, S. (2008). The multivariate least-trimmed squares estimator. *J. Multivariate Anal.*, 99:311–338.

Aron, J., Duca, J. V., Muellbauer, J., Murata, K., and Murphy, A. (2012). Credit, housing collateral, and consumption: Evidence from Japan, the U.K. and the U.S. *Review of Income and Wealth*, 58:397–423.

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

Ball, R. J., Boatwright, B. D., Burns, T., Lobban, P. W. M., and Miller, G. W. (1975). The London Business School quarterly econometric model of the UK economy. In Renton, G. A., editor, *Modelling the Economy*, pages 3–37. Heinemann Educational Books.

Ball, R. J. and Burns, T. (1968). An econometric approach to short-run analysis of the U.K. economy, 1955-66. *Operational Research Quarterly*, 19:225–256.

Banerjee, A., Dolado, J. J., and Mestre, R. (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *J. Time Series Anal.*, 19:267–283.

Bazinas, V. and Nielsen, B. (2022). Causal transmission in reduced-form models. *Econometrics*, 10:14.

Berenguer-Rico, V., Johansen, S., and Nielsen, B. (2023). A model where the least trimmed squares estimator is maximum likelihood. *J. Roy. Statist. Soc. Ser. B*, 85:886–912.

Berenguer-Rico, V. and Nielsen, B. (2023). Normality testing after outlier removal. *Econom. Stat.* To appear.

Berenguer-Rico, V. and Nielsen, B. (2024). Least trimmed squares: Nuisance parameter free asymptotics. *Econometric Theory*. To appear.

Berenguer-Rico, V. and Wilms, I. (2021). Heteroscedasticity testing after outlier removal. *Econometric Rev.*, 40:51–85.

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

Burridge, P. and Taylor, A. R. M. (2006). Additive outlier detection via extreme-value theory. *J. Time Series Anal.*, 27:685–701.

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

Chan, N. (1989). Asymptotic inference for unstable autoregressive time series with drifts. *J. Statist. Plann. Inference*, 23:301–312.

Chan, N. and Wei, C. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.*, 16:367–401.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28:591–605.

Croux, C. and Joossens, K. (2008). Robust estimation of the vector autoregressive model by a least trimmed squares procedure. In Brito, P., editor, *COMPSTAT 2008*, Proceedings in Computational Statistics, pages 489–501. Physica Verlag, Heidelberg.

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

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, 88:661–692.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.*, 74:427–431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49:1057–1072.

Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. pages 157–184. Wadsworth, Belmont, CA.

Doornik, J. A. (2009). Autometrics. In Castle and Shephard (2009), pages 88–121.
Doornik, J. A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70:927–939.
Doornik, J. A. and Hendry, D. F. (2022). *Empirical Econometric Modelling – PcGive 16*, volume 1. Timberlake, London.
Doornik, J. A., Hendry, D. F., and Nielsen, B. (1998). Inference in cointegrating models: UK M1 revisited. *Journal of Economic Surveys*, 12:533–572.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1108.
Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–276.
Engle, R. F., Hendry, D. F., and Richard, J. F. (1983). Exogeneity. *Econometrica*, 51:277–304.
Engler, E. and Nielsen, B. (2009). The empirical process of autoregressive residuals. *Econom. J.*, 12:367–381.
Fox, A. J. (1972). Outliers in time series. *J. Roy. Statist. Soc. Ser. B*, 34:350–363.
Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, 46:1293–1301.
Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48:213–228.
Harbo, I., Johansen, S., Nielsen, B., and Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems. *J. Bus. Econom. Statist.*, 16:388–399.
Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in fully saturated regression. *Computational Statistics*, 23:317–335. Erratum *ibid.* pp. 337–339.
Hendry, D. F. and Santos, C. (2010). An automatic test of super exogeneity. In Bollerslev, T., Russell, J. R., and Watson, M. W., editors, *Volatility and Time Series Econometrics*. Oxford University Press, Oxford.
Heng, Q. and Lange, K. (2025). Bootstrap estimation of the proportion of outliers in robust regression. *Statistics and Computing*, 35.
Johansen, S. (1988). Statistical analysis of cointegration vectors. *J. Econom. Dynam. Control*, 12:231–54.
Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *J. Econometrics*, 52:389–402.
Johansen, S. (1995). *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model*. Oxford University Press, Oxford.
Johansen, S. and Nielsen, B. (2009). Saturation by indicators in regression models. In Castle and Shephard (2009), pages 1–36.
Johansen, S. and Nielsen, B. (2016a). Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli*, 22:1131–1183. Corrigendum (2019) 25, 3201.
Johansen, S. and Nielsen, B. (2016b). Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scand. J. Stat.*, 43:321–81.
Johansen, S. and Nielsen, B. (2019). Boundedness of M-estimators for multiple linear

regression in time series. *Econometric Theory*, 35:653–683.
Kilian, L. and Demiroglu, U. (2000). Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence. *J. Bus. Econom. Statist.*, 18:40–50.
Klooster, J. and Nielsen, B. (2025). On (in)consistency of m-estimators under contamination. Technical report, arXiv:2502.09145v1.
Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.*, 10:154–166.
Maechler, M. and 9 other authors (2024). *robustbase*. R package version 0.99-4-1.
Nielsen, B. (2001). The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica*, 69:211–219.
Nielsen, B. (2006). Order determination in general vector autoregressions. In Ho, H. C., Ing, C. K., and Lai, T. L., editors, *Time series and related topics: In memory of Ching-Zong Wei*, volume 52 of *Lecture Notes–Monograph Series*, pages 93–112. Institute of Mathematical Statistics, Beachwood, OH.
Nielsen, B. (2009). Test for cointegration rank in general vector autoregressions. Discussion paper 2009-w10, Nuffield College.
Nielsen, B. and Sohkanen, J. (2011). Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends. *Econometric Theory*, 27:913–927.
Nielsen, B. and Whitby, A. (2015). A joint Chow test for structural instability. *Econometrics*, 3:156–186.
Nielsen, H. B. (2004). Cointegration analysis in the presence of outliers. *Econom. J.*, 7:249–271.
Phillips, P. C. B. (1988). Reflections on econometric methodology. *Economic Record*, 64:344–359.
Pretis, F., Reade, J., and Sucarrat, G. (2018). General-to-specific (gets) regression modeling and indicator saturation for outliers and structural breaks. *Journal of Statistical Software*, 86.
R Core Team (2024). R: A language and environment for statistical computing.
Raymaekers, J. and Rousseeuw, P. J. (2024). The cellwise minimum covariance determinant estimator. *J. Amer. Statist. Assoc.*, 119:2610–2621.
Riani, M. (2004). Extensions of the Forward Search to time series. *Studies in Nonlinear Dynamics & Econometrics*, 8:article 2.
Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In Grossmann, W., Pflug, G., Vincze, I., and Wertz, W., editors, *Mathematical Statistics and Applications*, pages 283–297. Reidel, Dordrecht.
Rousseeuw, P. J. (1984). Least median of squares regressions. *J. Amer. Statist. Assoc.*, 79:871–880.
Rousseeuw, P. J. (1994). Unconventional features of positive-breakdown estimators. *Statist. Probab. Lett.*, 19:417–431.
Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. John Wiley & Sons, Hoboken, NJ.
Rousseeuw, P. J. and van Driessen, K. (2000). An algorithm for positive-breakdown regression based on concentration steps. In Gaul, W., Opitz, O., and Schader, M., editors, *Data Analysis: Scientific Modeling and Practical Application*, pages 335–346.

Springer Verlag.

Rousseeuw, P. J. and van Driessen, K. (2006). Computing LTS regression for large data sets. *Data Mining and Knowledge Discovery*, 12:29–45.

Watts, V., Rootzén, H., and Leadbetter, M. R. (1982). On limiting distributions of intermediate order statistics from stationary sequences. *Ann. Probab.*, 10:653–662.

Welsh, A. H. and Ronchetti, E. (2002). A journey in single steps: robust one-step M-estimation in linear regression. *J. Statist. Plann. Inference*, 103:287–310.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–838.