CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2008-9

An analysis of the indicator saturation estimator as a robust regression estimator

Søren Johansen and Bent Nielsen

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Aarhus School of Business
University of Aarhus
Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF
COPENHAGEN

An analysis of the indicator saturation estimator as a robust regression estimator

Søren Johansen
Department of Economics, University of Copenhagen
and CREATES, University of Aarhus
and
Bent Nielsen*
Department of Economics, University of Oxford

February 4, 2008

Abstract

An algorithm suggested by Hendry (1999) for estimation in a regression with more regressors than observations, is analyzed with the purpose of finding an estimator that is robust to outliers and structural breaks. This estimator is an example of a one-step M-estimator based on Huber's skip function. The asymptotic theory is derived in the situation where there are no outliers or structural breaks using empirical process techniques. Stationary processes, trend stationary autoregressions and unit root processes are considered.

Keywords: Empirical processes, Huber's skip, indicator saturation, M-estimator, outlier robustness, vector autoregressive process.

JEL Classification: C32

*The first author gratefully acknowledges support from Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation. The second author received financial support from ESRC grant RES-000-27-0179. The figure was constructed using R (R Development Core Team, 2006). The authors would like to thank David Cox and Mette Ejrnæs for some useful comments on an earlier version of the paper.

1 Introduction

In an analysis of US food expenditure Hendry (1999) used an indicator saturation approach. The annual data spanned the period 1931-1989 including the great depression, World War II, and the oil crises. These episodes, covering 25% of the sample, could potentially result in outliers. An indicator saturation approach was adopted by forming zero-one indicators for these observation. Condensing the outcome, this large number of indicators could be reduced to just two outliers with an institutional interpretation.

The suggestion for outlier detection divides the sample in two sets and saturates first one set and then the other with indicators. The indicators are tested for significance using the parameter estimates from the other set and the corresponding observation is deleted if the test statistic is significant. The estimator is the least squares estimator based upon the retained observations. A formal version of this estimator is the indicator saturation estimator. This was analyzed recently by Hendry, Johansen and Santos (2008), who derived the asymptotic distribution of the estimator of the mean in the case of i.i.d. observations.

The purpose of the present paper is to analyse the indicator saturation algorithm as a special case of a general procedure considered in the literature of robust statistics. We consider the regression model $y_t = \beta' x_t + \varepsilon_t$ where ε_t are i.i.d. $(0, \sigma^2)$, and a preliminary estimator $(\hat{\beta}, \hat{\sigma}^2)$, which gives residuals $r_t = y_t - \hat{\beta}' x_t$. Let $\hat{\omega}_t^2$ be an estimate of the variance of r_t. Examples are $\hat{\omega}_t^2 = \hat{\sigma}^2$ which is constant in t and $\hat{\omega}_t^2 = \hat{\sigma}^2\{1 - x_t'(\sum_{s=1}^T x_s x_s')^{-1} x_t\}$ which varies with t. From this define the normalized residuals $v_t = r_t/\hat{\omega}_t$. The main result in Theorem 3.1 is an asymptotic expansion of the least squares estimator for (β, σ^2) based upon those observation for which $\underline{c} \leq v_t \leq \bar{c}$.

This expansion is then applied to find asymptotic distributions for various choices of preliminary estimator, like least squares and the split least squares considered in the indicator saturation approach. Asymptotic distributions are derived under stationary and trend stationary autoregressive processes and some results are given for unit root processes.

We do not give any results on the behavior of the estimators in the presence of outliers, but refer to further work which we intend to do in the future.

1.1 The relation to the literature on robust statistics

Detections of outliers is generally achieved by robust statistics in the class of M-estimators, or L-estimators, see for instance Huber (1981). An M-estimator of the type considered here is found by solving

$$\sum_{t=1}^{T} (y_t - \beta' x_t) x_t' 1_{(\sigma\underline{c} \leq y_t - \beta' x_t \leq \sigma\bar{c})} = 0, \tag{1.1}$$

supplemented with an estimator of variance of the residual. The objective function is known as Huber's skip function and has the property that it is not differentiable in β, σ^2. The solution may not be unique and the calculation can be difficult due to the

lack of differentiability, see Koenker (2005). A more tractable one-step estimator can be found from a preliminary estimator $(\hat{\beta}, \hat{\sigma})$ and choice of $\hat{\omega}_t^2$, by solving

$$\sum_{t=1}^{T}(y_t - \beta' x_t)x_t' 1_{(\hat{\omega}_t \underline{c} \leq y_t - \hat{\beta}' x_t \leq \hat{\omega}_t \bar{c})} = 0, \tag{1.2}$$

which is just the least squares estimator where some observations are removed as outliers according to a test based on the preliminary estimator. Note that the choice of the quantiles requires that we know the density f.

An alternative method is to order the residuals $r_t = y_t - \beta' x_t$ and eliminate the smallest $T\alpha_1$ and largest $T\alpha_2$ observations, and then use the remaining observations to calculate the least squares estimators. This is an L-estimator, based upon order statistics. A one-step estimator is easily calculated if a preliminary estimator is used to define the residuals. One can consider the M- and L-estimators as the estimators found by iterating the one step procedure described.

Rather than discarding outliers they could be capped at the quantile c as in the Winsorized least squares estimator solving $\sum_{t=1}^{T} r_t x_t' \min(1, c\hat{\omega}_t/|r_t|) = 0$, see Huber (1981, page 18). While the treatment of the outliers must depend on the substantive context, we focus on the skip estimator in this paper. A related estimator is the least trimmed squares estimator by Rousseeuw (1984) which minimizes $\sum_{i=1}^{h} r_i^2$ after having discarded the largest $T - h = T(\alpha_1 + \alpha_2)$ values of r_i^2.

The estimator we consider in our main result is the estimator (1.2), and we apply the main result to get the asymptotic distribution of the estimators for stationary processes, trend stationary processes, and some unit root processes for different choices of preliminary estimator.

One-step estimators have been considered before. The paper by Bickel (1975) has a one-step M-estimator of a different kind as the minimization problem is approximated using a linearization of the derivative of the objective function around a preliminary estimator. The estimator considered by Ruppert and Carroll (1980), however, is a one-step estimator of the kind described above, although of the L-type, see also Yohai and Maronna (1976).

The focus in the robustness literature has been on deterministic regressors satisfying $T^{-1}\sum_{t=1}^{T} x_t x_t' \to \Sigma > 0$, whereas we prove results for stationary and trend stationary autoregressive processes. We also allow for a non-symmetric error distribution.

We apply the theory of empirical processes using tightness arguments similar to Bickel (1975). The representation in our main result Theorem 3.1 generalizes the representations in Ruppert and Carroll (1980) to stochastic regressors needed for time series analysis.

As an example of the relation between the one-step estimator we consider and the general theory of M-estimators, consider the representation we find in Theorem 3.1 for the special case of i.i.d. observations with a symmetric distribution with mean μ, so that $x_t = 1$. In this case we find

$$T^{1/2}(\breve{\mu} - \mu) = (1-\alpha)^{-1}\{T^{-1/2}\sum_{t=1}^{T}\varepsilon_t 1_{(\underline{c}\sigma \leq \varepsilon_t \leq \sigma\bar{c})} + 2c\mathsf{f}(c)T^{1/2}(\hat{\mu} - \mu)\} + \mathsf{o}_{\mathsf{P}}(1).$$

If we iterate this procedure we could end up with an estimator, μ^*, which satisfies

$$T^{1/2}(\mu^* - \mu) = (1-\alpha)^{-1}\{T^{-1/2}\sum_{t=1}^{T}\varepsilon_t 1_{(\underline{c}\sigma \le \varepsilon_t \le \sigma\overline{c})} + 2c\mathsf{f}(c)T^{1/2}(\mu^* - \mu)\} + \mathsf{o_P}(1),$$

so that

$$\begin{aligned} T^{1/2}(\mu^* - \mu) &= \{1-\alpha-2c\mathsf{f}(c)\}^{-1}T^{-1/2}\sum_{t=1}^{T}\varepsilon_t 1_{(\underline{c}\sigma \le \varepsilon_t \le \sigma\overline{c})} + \mathsf{o_P}(1) \\ &\xrightarrow{\mathsf{D}} \mathsf{N}[0, \sigma^2 \frac{\tau_2^c}{\{1-\alpha-2c\mathsf{f}(c)\}^2}], \end{aligned}$$

which is the limit distribution conjectured by Huber (1964) for the M-estimator (1.1). It is also the asymptotic distribution of the least trimmed squares estimator, see Rousseeuw and Leroy (1987, p. 180), who rely on Yohai and Maronna (1976) for the i.i.d case.

1.2 The structure of the paper

The one-step estimators are described in detail in §2, and in §3 we find the asymptotic expansion of the estimators under general assumptions on the regressor variables, but under the assumption that the data generating process is given by the regression model without indicators. The situation where the initial estimator is a least square estimator is analysed for stationary processes in §4.1. The situation where the initial estimator is an indicator saturated estimator is then considered for stationary process in §4.2 and for trend stationary autoregressive processes and unit root processes in §5. Finally, §6 contains the proof of the main theorem, which involves techniques for empirical processes, whereas proofs for special cases are given in §7.

2 The one-step M-estimators

At first the statistical model is set up. Subsequently, the considered one-step estimators are introduced.

2.1 The regression model

As a statistical model consider the regression model

$$y_t = \beta' x_t + \sum_{i=1}^{T}\gamma_i 1_{(i=t)} + \varepsilon_t \qquad t = 1, \ldots, T, \tag{2.1}$$

where x_t is an m-dimensional vector of regressors and the conditional distribution of the errors, ε_t, given $(x_1, \ldots x_t, \varepsilon_1, \ldots, \varepsilon_{t-1})$ has density $\sigma^{-1}\mathsf{f}(\sigma^{-1}\varepsilon)$, so that $\sigma^{-1}\varepsilon_t$ are i.i.d. with density f. Thus, the density of y_t given the past should be a member of

a location-scale family such as the family of univariate normal distributions. When working with other distributions, such as the t-distribution the degrees of freedom should be known. We denote expectation and variance given $(x_1, \dots x_t, \varepsilon_1, \dots, \varepsilon_{t-1})$ by E_{t-1} and Var_{t-1}.

The parameter space of the model is given by $\beta, (\gamma_1, \dots, \gamma_T), \sigma^2 \in \mathbb{R}^m \times \mathbb{R}^T \times \mathbb{R}_+$. The number of parameters is therefore larger than the sample length. We want to make inference on the parameter of interest β in this regression problem with T observations and m regressors, where we consider the γ_is as nuisance parameters. The least squares estimator for β is contaminated by the γ_is and we therefore seek to robustify the estimator by introducing two critical values $\underline{c} < \overline{c}$ chosen so that

$$\tau_0^c = \int_{\underline{c}}^{\overline{c}} \mathsf{f}(v)dv = 1 - \alpha \quad \text{and} \quad \tau_1^c = \int_{\underline{c}}^{\overline{c}} v\mathsf{f}(v)dv = 0. \tag{2.2}$$

It is convenient to introduce as a general notation

$$\tau_n = \int_{\mathbb{R}} u^n \mathsf{f}(u)du, \qquad \tau_n^c = \int_{\underline{c}}^{\overline{c}} u^n \mathsf{f}(u)du, \tag{2.3}$$

for $n \in \mathbb{N}_0$, for the moments and truncated moments of f. A smoothness assumption to the density is needed.

Assumption A *The density* f *has continuous derivative* f' *and satisfies the condition*

$$\sup_{v \in \mathbb{R}} \{(1 + v^4)\mathsf{f}(v) + (1 + v^2)|\mathsf{f}'(v)|\} < \infty,$$

with moments $\tau_1 = 0$, $\tau_2 = 1$, $\tau_4 < \infty$.

2.2 Two one-step M-estimators

Two estimators are presented based on algorithms designed to eliminate observations with large values of $|\gamma_i|$. Both estimators are examples of one-step M-estimators. They differ in the choice of initial estimator. The first is based on a standard least squares estimator, while the second is based on the indicator saturation argument.

2.2.1 The robustified least squares estimator

The robustified least squares estimator is a one-step M-estimator with initial estimator given as the least squares estimator $(\hat{\beta}, \hat{\sigma}^2)$. From this, construct the t-ratios for testing $\gamma_i = 0$ as

$$v_t = (y_t - \hat{\beta}' x_t)/\hat{\omega}_t, \tag{2.4}$$

where $\hat{\omega}_t^2$ could simply be chosen as $\hat{\sigma}^2$ or as $\hat{\sigma}\{1 - x_t'(\sum_{s=1}^T x_s x_s')^{-1} x_t\}$ by following the usual finite sample formula for the distribution of residuals for fixed regressors.

We base the estimator on those observations that are judged insignificantly different from the predicted value $\hat{\beta}'x_t$, and define the robustified least squares estimator as the one-step M-estimator

$$\breve{\beta}_{LS} = \{\sum_{t=1}^{T} x_t x_t' 1_{(\underline{c}\leq v_t\leq \overline{c})}\}^{-1} \sum_{t=1}^{T} x_t y_t 1_{(\underline{c}\leq v_t\leq \overline{c})}, \tag{2.5}$$

$$\breve{\sigma}^2_{LS} = (\frac{\tau_2^c}{1-\alpha})^{-1}\{\sum_{t=1}^{T} 1_{(\underline{c}\leq v_t\leq \overline{c})}\}^{-1} \sum_{t=1}^{T} (y_t - \breve{\beta}_{LS}' x_t)^2 1_{(\underline{c}\leq v_t\leq \overline{c})}. \tag{2.6}$$

It will be shown that $\{\sum_{t=1}^{T} 1_{(\underline{c}\leq v_t\leq \overline{c})}\}^{-1} \sum_{t=1}^{T} (y_t - \breve{\beta}_{LS}' x_t)^2 1_{(\underline{c}\leq v_t\leq \overline{c})} \xrightarrow{\mathsf{P}} \sigma^2\tau_2^c/(1-\alpha)$, which justifies the bias correction in the expression for $\breve{\sigma}^2_{LS}$.

Obviously the denominators can be zero, but in this case also the numerator is zero and we can define $\breve{\beta}_{LS} = 0$ and $\breve{\sigma}^2_{LS} = 0$.

2.2.2 The indicator saturation estimator

Based on the idea of Hendry (1999) the indicator saturated estimator is defined as follows:

1. We split the data in two sets $\mathcal{I}_1$ and $\mathcal{I}_2$ of T_1 and T_2 observations respectively, where $T_j T^{-1} \to \lambda_j > 0$ for $T \to \infty$.

2. We calculate the ordinary least squares estimator for (β, σ^2) based upon the sample $\mathcal{I}_j$,

$$\hat{\beta}_j = (\sum_{t\in\mathcal{I}_j} x_t x_t')^{-1} \sum_{t\in\mathcal{I}_j} x_t y_t, \qquad \hat{\sigma}_j^2 = \frac{1}{T_j}\sum_{t\in\mathcal{I}_j} (y_t - \hat{\beta}_j' x_t)^2, \tag{2.7}$$

and define the t-ratios for testing $\gamma_i = 0$:

$$v_t = 1_{(t\in\mathcal{I}_2)}(y_t - \hat{\beta}_1' x_t)/\hat{\omega}_{t,1} + 1_{(t\in\mathcal{I}_1)}(y_t - \hat{\beta}_2' x_t)/\hat{\omega}_{t,2}, \tag{2.8}$$

where $\hat{\omega}^2_{t,j}$ could be chosen as $\hat{\sigma}_j^2\{1 + x_t'(\sum_{s\notin I_j} x_s x_s')^{-1} x_t\}$ for fixed regressors.

3. We then compute robustified least squares estimators $\tilde{\beta}$ and $\tilde{\sigma}^2$ by (2.5) and (2.6) based on v_t given by (2.8).

4. Based on the estimators $\tilde{\beta}$ and $\tilde{\sigma}^2$ define the t-ratios for testing $\gamma_i = 0$:

$$\tilde{v}_t = (y_t - \tilde{\beta}' x_t)/\tilde{\omega}_t, \tag{2.9}$$

where $\tilde{\omega}_t^2$ could be chosen as $\tilde{\sigma}^2$. It is less obvious how to choose a finite sample correction since the second round initial estimator $(\tilde{\beta}, \tilde{\sigma}^2)$ is not based upon least squares.

5. Finally, compute the indicator saturated estimators $\breve{\beta}_{Sat}$ and $\breve{\sigma}^2_{Sat}$ as the robustified least squares estimators (2.5) and (2.6) based on $\tilde{v}_t$ given by (2.9).

3 The main asymptotic result

Asymptotic distributions will be derived under the assumption that in (2.1) the indicators are not needed because $\gamma_i = 0$ for all i, that is, $(y_t - \beta' x_t)/\sigma$ are i.i.d. with density

f. The main result, given here shows that in the analysis of one-step M-estimators the indicators $1_{(\underline{c}\leq v_t\leq \overline{c})}$, based on the normalized residual $v_t = (y_t - \hat{\beta}' x_t)/\hat{\omega}_t$, can be replaced by $1_{(\underline{c}\sigma\leq \varepsilon_t<\overline{c}\sigma)}$ combined with correction terms. This shows how the limit distributions of the initial estimators $\hat{\beta}$ and $\hat{\sigma}^2$ influence the limit distribution of the robustified estimators. The result is the basis for any further asymptotic analysis and can be applied both for stationary and trend stationary regressors, and for unit root processes, but not for explosive processes.

It is convenient to define product moments of the retained observations for any two processes u_t and w_t as $S_{uw} = \sum_{t=1}^{T} u_t w_t' 1_{(\underline{c}\leq v_t\leq \overline{c})}$, so that the robustified estimators (2.5) and (2.6) become

$$\breve{\beta} = S_{xx}^{-1} S_{xy}, \tag{3.1}$$

$$\breve{\sigma}^2 = (1-\alpha)(\tau_2^c S_{11})^{-1}(S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}). \tag{3.2}$$

The estimator $\hat{\omega}_t^2$ for the variance of residual r_t can be chosen from a wide range of estimators including $\hat{\sigma}^2$ and $\hat{\sigma}^2\{1 - x_t'(\sum_{s=1}^{T} x_s x_s')^{-1} x_t\}$. These estimators do, however, have to satisfy the following condition.

Assumption B *The estimator $\hat{\omega}_t^2$ is chosen so* $\max_{1\leq t\leq T} T^{1/2}|\hat{\omega}_t^2 - \hat{\sigma}^2| = \mathsf{o_P}(1)$.

We can now formulate the main result which shows how the product moments S_{uv} depend on the truncation points $\underline{c}$ and $\overline{c}$ and the initial estimators $\hat{\beta}$ and $\hat{\sigma}^2$.

Theorem 3.1 *Consider model (2.1), where $\gamma_i = 0$ for all i, and there exists some estimators $(\hat{\beta}, \hat{\sigma}^2)$ and non-stochastic normalization matrices $N_T \to 0$, so that*
(i) The initial estimators satisfy
(a) $T^{1/2}(\hat{\sigma}^2 - \sigma^2)$, $(N_T^{-1})'(\hat{\beta} - \beta) = \mathrm{O_P}(1)$,
(b) $\hat{\omega}_t^2$ satisfies Assumption B.
(ii) The regressors satisfy, jointly,
(a) $N_T \sum_{t=1}^{T} x_t x_t' N_T' \xrightarrow{\mathsf{D}} \Sigma \overset{a.s.}{>} 0$,
(b) $T^{-1/2} N_T \sum_{t=1}^{T} x_t \xrightarrow{\mathsf{D}} \mu$,
(c) $\max_{t\leq T} \mathsf{E}|T^{1/2} N_T x_t|^4 = \mathrm{O}(1)$.
(iii) The density f *satisfies Assumption A, and $\underline{c}$ and $\overline{c}$ are chosen so that $\tau_1^c = 0$.*
Then it holds

$$T^{-1} S_{11} \xrightarrow{\mathsf{P}} 1-\alpha, \tag{3.3}$$

$$N_T S_{xx} N_T' \xrightarrow{\mathsf{D}} (1-\alpha)\Sigma, \tag{3.4}$$

$$T^{-1/2} N_T S_{x1} \xrightarrow{\mathsf{D}} (1-\alpha)\mu. \tag{3.5}$$

For $\xi_n^c = (\overline{c})^n f(\overline{c}) - (\underline{c})^n f(\underline{c})$ *and* $\tau_2^c = \int_{\underline{c}}^{\overline{c}} v^2 \mathsf{f}(v) dv$ *we find the expansions*

$$N_T S_{x\varepsilon} = N_T \sum_{t=1}^{T} \{x_t \varepsilon_t 1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)} + \xi_1^c x_t x_t'(\hat{\beta} - \beta) + \xi_2^c (\hat{\sigma} - \sigma) x_t\} + \mathrm{o_P}(1), \quad (3.6)$$

$$S_{\varepsilon\varepsilon} = \sum_{t=1}^{T} \{\varepsilon_t^2 1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)} + \sigma \xi_2^c (\hat{\beta} - \beta)' x_t + \sigma \xi_3^c (\hat{\sigma} - \sigma)\} + \mathrm{o_P}\left(T^{1/2}\right), \quad (3.7)$$

$$S_{11} = \sum_{t=1}^{T} \{1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)} + \xi_0^c (\hat{\beta} - \beta)' x_t / \sigma + \xi_1^c (\hat{\sigma}/\sigma - 1)\} + \mathrm{o_P}\left(T^{1/2}\right). \quad (3.8)$$

Combining the expressions for the product moments gives expressions for the one-step M-estimators of the form (3.1), (3.2). The expressions give a linearization of these estimators in terms of the initial estimators. For particular initial estimators explicit expressions for the limiting distributions are then derived in the subsequent sections.

Corollary 3.2 *Suppose the assumptions of Theorem 3.1 are satisfied. Then*

$$\begin{aligned}(1-\alpha)\Sigma(N_T^{-1})'(\breve{\beta} - \beta) = N_T \sum_{t=1}^{T} x_t \varepsilon_t 1_{(\underline{c}\sigma \le \varepsilon_t \le \sigma\overline{c})} \\ + \xi_1^c \Sigma (N_T^{-1})'(\hat{\beta} - \beta) + \xi_2^c T^{1/2} (\hat{\sigma} - \sigma)\mu + \mathrm{o_P}(1), \quad (3.9)\end{aligned}$$

$$\begin{aligned}\tau_2^c T^{1/2}(\breve{\sigma}^2 - \sigma^2) = T^{-1/2} \sum_{t=1}^{T} (\varepsilon_t^2 - \sigma^2 \frac{\tau_2^c}{1-\alpha}) 1_{(\underline{c}\sigma \le \varepsilon_t \le \sigma\overline{c})} \\ + \sigma \zeta_2^c \mu' (N_T^{-1})'(\hat{\beta} - \beta) + \sigma \zeta_3^c T^{1/2}(\hat{\sigma} - \sigma) + \mathrm{o_P}(1), \quad (3.10)\end{aligned}$$

where $\zeta_n^c = \xi_n^c - \xi_{n-2}^c \tau_2^c/(1-\alpha)$. *It follows that*

$$\{(N_T^{-1})'(\breve{\beta} - \beta), T^{1/2}(\breve{\sigma}^2 - \sigma^2)\} = \mathrm{O_P}(1), \quad (3.11)$$

so that $(\breve{\beta}, \breve{\sigma}^2) \xrightarrow{\mathsf{P}} (\beta, \sigma^2)$.

The proofs of Theorem 3.1 and Corollary 3.2 are given in §6. It involves a series of steps. In §6.1 a number of inequalities are given for the indicator functions appearing in S_{xx} and $S_{x\varepsilon}$, and in §6.2 we show some limit results which take care of the remainder terms in the expansions. The argument involves weighted empirical processes with weights $x_t x_t'$, $x_t \varepsilon_t$, ε_t^2 and 1 appearing in the numerator and denominators of $\breve{\beta}$ and $\breve{\sigma}^2$. Weighted empirical processes have been studied by Koul (2002), but with conditions on the weights that would be too restrictive for this study. Finally, the threads are pulled together in §6.3.

The assumptions (ii, a) and (ii, b) are satisfied in a wide range of models. The assumption (ii, c) is slightly more restrictive: It permits classical stationary regressions as well as stationary autoregressions in which case $N_T = T^{-1/2}$ and trend stationary processes with a suitable choice of N_T. It also permits unit root processes where

$N_T = T^{-1}$, as well as processes combining stationary and unit root phenomena. The assumption (ii, c) does, however, exclude exponentially growing regressors. As an example let $x_t = 2^t$. In that case $N_T = 2^{-T}$ and $\max_{t\leq T} T^{1/2}2^{-T}2^t = T^{1/2}$ diverges. Likewise, explosive autoregressions are excluded.

Similarly, the assumption (i, b), referring to Assumption B, is satisfied for a wide range of situations. If $\hat{\omega}_t^2 = \hat{\sigma}^2$ it is trivially satisfied. If $\hat{\omega}_t^2 = \hat{\sigma}^2\{1-x_t'(\sum_{s=1}^T x_s x_s')^{-1}x_t\}$ as in the computation of the robustified least squares estimator the assumption is satisfied when the regressors x_t have stationary, unit root, or polynomial components, but not if the regressors are explosive. This is proved by first proving (ii, a, c) and then combining these conditions.

The assumption that $\tau_1^c = 0$ is important. If it had been different from zero then $\varepsilon_t 1_{(\underline{c}\sigma\leq\varepsilon\leq\sigma\bar{c})}$ would not have zero mean and the conclusion (3.11) would in general fail because $N_T S_{x\varepsilon}$ would diverge.

4 Asymptotic distributions in the stationary case

We now apply Theorem 3.1 and Corollary 3.2 to the case of stationary regressors with finite fourth moment where we can choose $N_T = T^{-1/2}I_m$. With this choice the assumptions $(ii)(a, b, c)$ of Theorem 3.1 are satisfied by the Law of Large Numbers for stationary processes with finite fourth moments.

The stationary case covers a wide range of standard models:

(i) The classical regression model, where x_t is stationary with finite fourth moment.

(ii) Stationary autoregression of order k. We let $y_t = X_t$ and $x_t = (X_{t-1} \dots X_{t-k})'$. An intercept could, but need not, be included as in the equation

$$X_t = \sum_{j=1}^{k} \alpha_j X_{t-j} + \mu + \varepsilon_t.$$

(iii) Autoregressive distributed lag models of order k. For this purpose consider a p-dimensional stationary process X_t partitioned as $X_t = (y_t, z_t')'$. This gives the model equation for y_t given the past $(X_s, s \leq t-1)$ and z_t

$$y_t = \sum_{j=1}^{k} \alpha_j' X_{t-j} + \beta' z_t + \mu_y + \varepsilon_t.$$

Here, the regressor z_t could be excluded to give the equation of a vector autoregression.

4.1 Asymptotic distribution of the robustified least squares estimator

In this section we denote the least squares estimators by $(\hat{\beta}, \hat{\sigma}^2)$ and we let $(\breve{\beta}_{LS}, \breve{\sigma}_{LS}^2)$ be the robustified least squares estimators based on these, as given by (2.4), (3.1), and (3.2). We find the asymptotic distribution of these estimators with a proof in §7.

Theorem 4.1 *Consider model* (2.1) *with* $\gamma_i = 0$ *for all* i. *We assume that* x_t *is a stationary process with mean* μ, *variance* Σ, *and finite fourth moment so we can take* $N_T = T^{-1/2} I_m$, *and that* $\hat{\omega}_t^2$ *satisfies Assumption B. The density* f *satisfies Assumption A, and* $\underline{c}$ *and* $\bar{c}$ *are chosen so that* $\tau_1^c = 0$. *Then*

$$T^{1/2} \begin{pmatrix} \breve{\beta}_{LS} - \beta \\ \breve{\sigma}^2_{LS} - \sigma^2 \end{pmatrix} \xrightarrow{\mathsf{D}} \mathsf{N}_{m+1}\{0, \begin{pmatrix} \Omega_\beta & \Omega_c \\ \Omega_c' & \Omega_\sigma \end{pmatrix}\},$$

where

$$\begin{aligned} \Omega_\beta &= \sigma^2(\eta_\beta \Sigma^{-1} + \kappa_\beta \Sigma^{-1}\mu\mu'\Sigma^{-1}), \\ \Omega_c &= \sigma^3(\eta_c \Sigma^{-1}\mu + \kappa_c \Sigma^{-1}\mu\mu'\Sigma^{-1}\mu), \\ \Omega_\sigma &= 2\sigma^4(\eta_\sigma + \kappa_\sigma \mu'\Sigma^{-1}\mu), \end{aligned}$$

and

$$\begin{aligned} (1-\alpha)^2\eta_\beta &= \tau_2^c\,(1 + 2\xi_1^c) + (\xi_1^c)^2\,, \\ (1-\alpha)^2\kappa_\beta &= \xi_2^c\{\frac{1}{4}\xi_2^c(\tau_4 - 1) + \xi_1^c\tau_3 + \tau_3^c\}, \\ (1-\alpha)\tau_2^c\eta_c &= \zeta_2^c(\tau_2^c + \xi_1^c) + \frac{\xi_2^c}{2}\{\tau_4^c - \frac{(\tau_2^c)^2}{1-\alpha}\} + \frac{\xi_2^c\zeta_3^c}{4}(\tau_4 - 1) \\ &\quad + (1 + \xi_1^c)\tau_3^c + \frac{\zeta_3^c}{2}(\tau_3^c + \xi_1^c\tau_3). \\ (1-\alpha)\tau_2^c\kappa_c &= \frac{(\zeta_2^c)^2}{2}\tau_3^c \\ 2(\tau_2^c)^2\eta_\sigma &= \{\tau_4^c - \frac{(\tau_2^c)^2}{1-\alpha}\}(1 + \zeta_3^c) + \frac{(\zeta_3^c)^2}{4}(\tau_4 - 1) \\ 2(\tau_2^c)^2\kappa_\sigma &= \zeta_2^c(\zeta_2^c + 2\tau_3^c + \zeta_3^c\tau_3). \end{aligned}$$

For a given f, α, $\underline{c}$, *and* $\bar{c}$, *the coefficients* η *and* κ *are known. The parameters* (σ^2, Σ, μ) *are estimated by* $\breve{\sigma}^2_{LS}$, *see* (3.11), $N_T S_{xx} N_T/(1-\alpha)$, *see* (3.4), *and* $T^{-1/2}N_T S_{x1}/(1-\alpha)$, *see* (3.5), *respectively, so that, for instance,*

$$(\breve{\Sigma}^{-1}\eta + \breve{\Sigma}^{-1}\breve{\mu}\breve{\mu}'\breve{\Sigma}^{-1}\kappa)^{-1/2}\breve{\sigma}_{LS}^{-1}T^{1/2}(\breve{\beta}_{LS} - \beta) \xrightarrow{\mathsf{D}} \mathsf{N}_m\,(0, I_m)\,.$$

The case where f is symmetric is of special interest. The critical value is then $c = -\underline{c} = \bar{c}$ and $\tau_3 = \tau_3^c = 0$ and $\xi_0^c = \xi_2^c = 0$ so $\zeta_2^c = 0$, whereas $\xi_1^c = 2c\mathsf{f}\,(c)$ and $\xi_3^c = 2c^3\mathsf{f}(c)$ so $\zeta_3^c = \{c^2 - \tau_2^c/(1-\alpha)\}2c\mathsf{f}(c)$. It follows that $\kappa_\beta = \kappa_\sigma = \kappa_c = \eta_c = 0$. Theorem 4.1 then has the following Corollary.

Corollary 4.2 *If* f *is symmetric and the assumptions of Theorem 4.1 hold, then*

$$T^{1/2} \begin{pmatrix} \breve{\beta}_{LS} - \beta \\ \breve{\sigma}^2_{LS} - \sigma^2 \end{pmatrix} \xrightarrow{\mathsf{D}} \mathsf{N}_{m+1}\{0, \begin{pmatrix} \sigma^2\eta_\beta\Sigma^{-1} & 0 \\ 0 & 2\sigma^4\eta_\sigma \end{pmatrix}\},$$

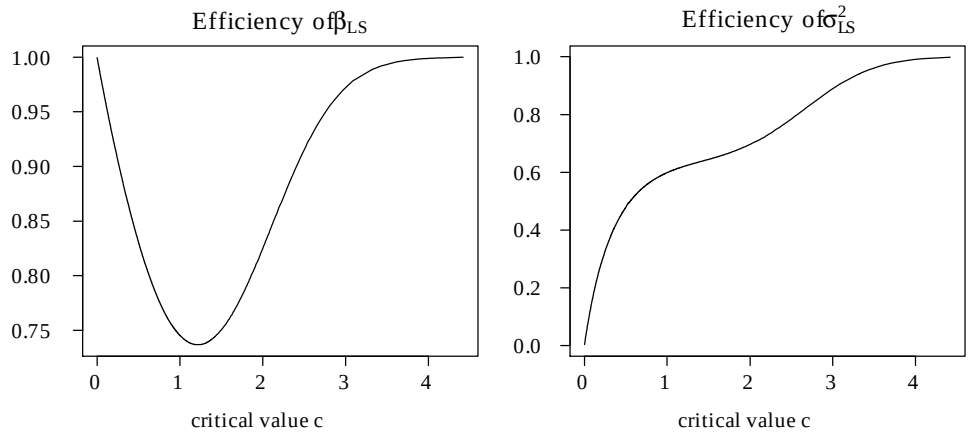

Figure 1: The efficiency of the estimators $\breve{\beta}_{LS}$ and $\breve{\sigma}^2_{LS}$ with respect to the least squares estimators $\hat{\beta}$ and $\hat{\sigma}^2$, respectively, for f equal to the Gaussian density.

where, with $\xi_1^c = 2c\mathsf{f}(c)$ *and* $\zeta_3^c = \{c^2 - \tau_2^c/(1-\alpha)\}2c\mathsf{f}(c)$, *it holds*

$$\begin{aligned}(1-\alpha)^2\eta_\beta &= \tau_2^c(1+2\xi_1^c) + (\xi_1^c)^2,\\ 2(\tau_2^c)^2\eta_\sigma &= \{\tau_4^c - \frac{(\tau_2^c)^2}{1-\alpha}\}(1+\zeta_3^c) + \frac{(\zeta_3^c)^2}{4}(\tau_4 - 1).\end{aligned}$$

Corollary 4.2 shows that the efficiency of the indicator saturated estimator $\breve{\beta}_{LS}$ with respect to the least squares estimator $\hat{\beta}$ is

$$\text{efficiency}(\hat{\beta}, \breve{\beta}_{LS}) = \{\mathsf{asVar}(\breve{\beta}_{LS})\}^{-1}\{\mathsf{asVar}(\hat{\beta})\} = \eta_\beta^{-1}.$$

Likewise the efficiency of $\breve{\sigma}_{LS}$ is $\text{efficiency}(\hat{\sigma}^2, \breve{\sigma}^2_{LS}) = \eta_\sigma^{-1}$. In the symmetric case the efficiency coefficients do not depend on the parameters of the process, only on the reference density f and the chosen critical value $c = \bar{c} = -\underline{c}$. They are illustrated in Figure 1.

4.2 The indicator saturated estimator

The indicator saturated estimator, $\breve{\beta}_{Sat}$, is a one-step M-estimator iterated twice. Its properties are derived from Theorem 3.1. We first prove two representations corresponding to (3.9) and (3.10) for the first round estimators $\tilde{\beta}$, $\tilde{\sigma}^2$ based on the least squares estimators $\hat{\beta}_j$ and $\hat{\sigma}_j$. Secondly, the limiting distributions of these first round estimators are found. Finally, the limiting distributions of the second round estimators $\breve{\beta}_{Sat}$, $\breve{\sigma}_{Sat}$ are found.

Theorem 4.3 *Suppose* $\gamma_i = 0$ *for all* i *in model (2.1), and that* x_t *is stationary with mean* μ, *variance* Σ, *and finite fourth moment, and that* $\hat{\omega}^2_{t,1}$ *and* $\hat{\omega}^2_{t,2}$ *satisfy Assumption*

B. The density f *satisfies Assumption A, and* $\underline{c}$ *and* $\bar{c}$ *are chosen so that* $\tau_1^c = 0$*. Then, for* $j = 1, 2$ *it holds, with* $\lambda_1 + \lambda_2 = 1$ *and* $\lambda_j > 0$*, that*

$$T^{-1}\sum_{t\in\mathcal{I}_j} x_t \xrightarrow{\mathsf{P}} \lambda_j\mu, \qquad T^{-1}\sum_{t\in\mathcal{I}_j} x_t x_t' \xrightarrow{\mathsf{P}} \lambda_j\Sigma. \tag{4.1}$$

Defining $\zeta_n^c = \xi_n^c - \xi_{n-2}^c\tau_2^c\sigma^2/(1-\alpha)$ *and the function* $h_t = (\lambda_1/\lambda_2)1_{\{t\in\mathcal{I}_2\}} + (\lambda_2/\lambda_1)1_{\{t\in\mathcal{I}_1\}}$*. Then it holds that*

$$\begin{aligned}(1-\alpha)\Sigma T^{1/2}(\tilde{\beta}-\beta) &= T^{-1/2}\sum_{t=1}^{T}[x_t\{\varepsilon_t 1_{(\underline{c}\sigma\le\varepsilon_t\le\sigma\bar{c})} + h_t\xi_1^c\varepsilon_t\} \\ &\quad + \frac{\xi_2^c}{2}\mu h_t(\varepsilon_t^2/\sigma - \sigma)] + \mathsf{o_P}(1), \qquad (4.2)\\ \tau_2^c T^{1/2}(\tilde{\sigma}^2-\sigma^2) &= T^{-1/2}\sum_{t=1}^{T}\{(\varepsilon_t^2 - \sigma^2\frac{\tau_2^c}{1-\alpha})1_{(\underline{c}\sigma\le\varepsilon_t\le\sigma\bar{c})} \\ &\quad + \sigma\zeta_2^c\mu'\Sigma^{-1}x_t\varepsilon_t h_t + \sigma\frac{\zeta_3^c}{2}(\varepsilon_t^2/\sigma-\sigma)h_t\} + \mathsf{o_P}(1). \qquad (4.3)\end{aligned}$$

The asymptotic distribution of the first-round estimators $\tilde{\beta}, \tilde{\sigma}^2$ can now be deduced. For simplicity only $\tilde{\beta}$ is considered.

Theorem 4.4 *Suppose* $\gamma_i = 0$ *for all* i *in model* (2.1)*, and that* x_t *is stationary with mean* μ*, variance* Σ*, and finite fourth moment, and that* $\hat{\omega}_{t,1}^2$ *and* $\hat{\omega}_{t,2}^2$ *satisfy Assumption B. The density* f *satisfies Assumption A, and* $\underline{c}$ *and* $\bar{c}$ *are chosen so that* $\tau_1^c = 0$*. Then*

$$T^{1/2}(\tilde{\beta}-\beta) \xrightarrow{\mathsf{D}} \mathsf{N}_m\left\{0, \sigma^2(\eta\Sigma^{-1} + \kappa\Sigma^{-1}\mu\mu'\Sigma^{-1})\right\}, \tag{4.4}$$

where

$$\begin{aligned}(1-\alpha)^2\eta &= \tau_2^c(1+2\xi_1^c) + (\xi_1^c)^2(\frac{\lambda_2^2}{\lambda_1} + \frac{\lambda_1^2}{\lambda_2}),\\ (1-\alpha)^2\kappa &= \xi_2^c[\{\frac{1}{4}\xi_2^c(\tau_4-1) + \xi_1\tau_3\}(\frac{\lambda_2^2}{\lambda_1} + \frac{\lambda_1^2}{\lambda_2}) + \tau_3^c].\end{aligned}$$

We note that the result of Hendry, Johansen, and Santos (2008) is a special case of Theorem 4.4. They were concerned with the situation of estimating the mean in an i.i.d. sequence where $\Sigma = 1$. Due to the relatively simple setup their proof could avoid the empirical process arguments used here.

In the special case where $\lambda_1 = \lambda_2 = 1/2$ then the limiting expression for $\tilde{\beta}$ is exactly the same as that for the robustified least squares estimator $\breve{\beta}_{LS}$, in that $\eta = \eta_\beta$ and $\kappa = \kappa_\beta$.

We finally analyse the situation where we first find the least squares estimators in the two subsets $\mathcal{I}_1$ and $\mathcal{I}_2$, then construct $\tilde{\beta}$ and finally find the robustified least squares estimator $\breve{\beta}_{Sat}$ based upon $\tilde{\beta}$. For simplicity we consider only the symmetric case.

Theorem 4.5 *Suppose $\gamma_t = 0,\ t = 1, \ldots, T$ in model (2.1), and that x_t is stationary with mean μ, variance Σ, and finite fourth moment, and that $\hat{\omega}^2_{t,j}$ and $\tilde{\omega}^2_t$ satisfy Assumption B. The symmetric density f satisfies Assumption A, and $\underline{c}$ and $\bar{c}$ are chosen so that $\tau^c_1 = 0$. Then*

$$T^{1/2}(\breve{\beta}_{Sat} - \beta) \xrightarrow{\mathsf{D}} \mathsf{N}_m(0, \sigma^2 \Sigma^{-1} \eta_{Sat}),$$

where

$$(1-\alpha)^4 \eta_{Sat} = (1 - \alpha + \xi^c_1)\tau^c_2\{(1 - \alpha + \xi^c_1) + 2(\xi^c_1)^2\} + (\xi^c_1)^4(\frac{\lambda^2_1}{\lambda_2} + \frac{\lambda^2_2}{\lambda_1}). \tag{4.5}$$

The assumption to the residual variance estimators is satisfied in a number of situations. If $\hat{\omega}^2_{t,j} = \hat{\sigma}^2_j$ and $\tilde{\omega}^2_t = \tilde{\sigma}^2$ then Assumption B is trivially satisfied. If $\hat{\omega}^2_{t,j} = \hat{\sigma}^2_j\{1 + x'_t(\sum_{s \notin \mathcal{I}_j} x_s x'_s)^{-1} x_t\}$ then Assumption B is satisfied due to the difference in the order of magnitude of x_t and $\sum_{s \notin \mathcal{I}_j} x_s x'_s$.

5 Asymptotic distribution for trending autoregressive processes

We first discuss the limit distribution of the least squares estimator in a trend stationary k-th order autoregression, and then apply the results to the indicator saturated estimator. Finally, the unit root case is discussed.

5.1 Least squares estimation in an autoregression

The asymptotic distribution of the least squares estimator is derived for a trend stationary autoregression. Consider a time series $y_{1-k}, \ldots, y_T$. The model for y_t has a deterministic component d_t. These satisfy the autoregressive equations

$$\begin{aligned} y_t &= \sum_{i=1}^{k} \gamma_i y_{t-i} + \varphi d_{t-1} + \varepsilon_t, \\ d_t &= D d_{t-1}, \end{aligned} \tag{5.1}$$

where $\varepsilon_t \in \mathbb{R}$ are independent, identically distributed with mean zero and variance σ^2, whereas $d_t \in \mathbb{R}^\ell$ are deterministic terms. The autoregression (5.1) is of the form (2.1) with $x'_t = (y_{t-1}, \ldots, y_{t-k}, d'_t)$ and $\beta' = (\gamma_1, \ldots, \gamma_k, \phi)$, so $m = k + \ell$. The least squares estimator is denoted $(\hat{\beta}, \hat{\sigma}^2)$.

The deterministic terms are defined in terms of the matrix D which has characteristic roots on the complex unit circle, so d_t is a vector of terms such as a constant, a linear trend, or periodic functions like seasonal dummies. For example,

$$D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{with} \qquad d_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

will generate a constant and a dummy for a bi-annual frequency. The deterministic term d_t is assumed to have linearly independent coordinates, which is formalised as follows.

Assumption C $|\text{eigen}\,(D)| = 1$ *and* $\text{rank}\,(d_1, \ldots, d_\ell) = \ell$.

It is convenient to introduce the companion form

$$Y_{t-1} = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-k} \end{pmatrix}, \quad A = \begin{Bmatrix} (\gamma_1, \ldots, \gamma_{k-1}) & \gamma_k \\ I_{k-1} & 0 \end{Bmatrix}, \quad \Phi = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}, \quad e_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix},$$

so that $Y_t = AY_{t-1} + \Phi d_{t-1} + e_t$. Focusing on the stationary case where $|\text{eigen}\,(A)| < 1$ so A and D have no eigenvalues in common, Nielsen (2005, §3) shows that

$$Y_t = Y_t^* + \Psi d_t \qquad \text{where} \qquad Y_t^* = AY_{t-1}^* + e_t,$$

and Ψ is the unique solution of the linear equation $\Phi = \Psi D - A\Psi$.

A normalization matrix N_T is needed. To construct this let

$$M_T = (\sum_{t=1}^{T} d_{t-1} d_{t-1}')^{-1/2},$$

so that $M_T \sum_{t=1}^{T} d_{t-1} d_{t-1}' M_T' = I_\ell$. Equivalently, a block diagonal normalisation, N_D, could be chosen if D, without loss of generality, were assumed to have a Jordan structure as in Nielsen (2005, §4). Theorem 4.1 of that paper then implies that

$$T^{-1/2} M_T \sum_{t=1}^{T} d_{t-1} \to \mu_D,$$

for some vector μ_D. For the entire vector of regressors, $x_t = (Y_{t-1}', d_{t-1}')'$, define

$$N_T = \begin{pmatrix} T^{-1/2} & 0 \\ 0 & M_T \end{pmatrix} \begin{pmatrix} I_k & -\Psi \\ 0 & I_\ell \end{pmatrix}. \tag{5.2}$$

Theorem 5.1 *Let y_t be the trend stationary process given by* (5.1) *so* $|\text{eigen}(A)| < 1$, *with finite fourth moment and deterministic component satisfying Assumption C. Then, with* $\Sigma_Y = \sum_{t=0}^{\infty} A^t \Omega\,(A^t)'$ *and* $\Sigma_D = I_\ell$ *and* $\mu_D = \lim_{T\to\infty} T^{-1/2} M_T \sum_{t=1}^{T} d_t$ *it holds*

$$N_T \sum_{t=1}^{T} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix}' N_T' \quad \xrightarrow{\mathsf{P}} \quad \Sigma \stackrel{\text{def}}{=} \begin{pmatrix} \Sigma_Y & 0 \\ 0 & \Sigma_D \end{pmatrix}, \tag{5.3}$$

$$T^{-1/2} N_T \sum_{t=1}^{T} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix} \quad \xrightarrow{\mathsf{P}} \quad \mu \stackrel{\text{def}}{=} \begin{pmatrix} 0 \\ \mu_D \end{pmatrix}, \tag{5.4}$$

$$\max_{1\le t\le T} |M_T d_t| \quad = \quad \mathrm{O}(T^{-1/2}), \tag{5.5}$$

$$N_T \sum_{t=1}^{T} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix} \varepsilon_t' \quad \xrightarrow{\mathsf{D}} \quad N_m(0, \sigma^2 \Sigma). \tag{5.6}$$

In particular, it holds

$$(N_T^{-1})'(\hat{\beta}-\beta) \xrightarrow{\mathsf{D}} \mathsf{N}_m(0,\sigma^2\Sigma^{-1}), \tag{5.7}$$

$$T^{1/2}(\hat{\sigma}^2-\sigma^2) = T^{-1/2}\sum_{t=1}^{T}(\varepsilon_t^2-\sigma^2)+\mathsf{o_P}(1)=\mathsf{O_P}(1). \tag{5.8}$$

A conclusion from the above analysis is that the normalization by N_T involving the parameter separates the asymptotic distribution into independent components. This will be exploited to simplify the analysis of the indicator saturated estimator below.

5.2 Indicator saturation in a trend stationary autoregression

We now turn to the indicator saturated estimator in the trend stationary autoregression, although only the first round estimator $\tilde{\beta}$ is considered. As before this estimator will consist of a numerator and a denominator term, each of which is a sum of two components. The main result in Theorem 3.1 can then be applied to each of these components.

Theorem 5.2 *Let y_t be the trend stationary process given by (5.1) so $|\text{eigen}(A)|<1$, with finite fourth moment, deterministic component satisfying Assumption C, and $\hat{\omega}_{t,j}^2$ satisfies Assumption B. Suppose the density f satisfies Assumption A, and the truncation points are chosen so that $\tau_1^c=0$. Finally, assume that*

$$\lim_{T\to\infty} M_T\sum_{t\in\mathcal{I}_j} d_t d_t' M_T = \Sigma_{D,j}>0, \tag{5.9}$$

$$\lim_{T\to\infty} T^{-1/2}M_T\sum_{t\in\mathcal{I}_j} d_t = \mu_{D,j}, \tag{5.10}$$

where $\Sigma_{D,1}+\Sigma_{D,2}=I_m$ and $\mu_{D,1}+\mu_{D,2}=\mu$ and define

$$\mu_j=\begin{pmatrix}0\\ \mu_{D,j}\end{pmatrix}, \qquad \Sigma_j=\begin{pmatrix}\lambda_j\Sigma_Y & 0\\ 0 & \Sigma_{D,j}\end{pmatrix}.$$

Then it holds

$$(\widetilde{\beta}-\beta)N_T^{-1}\xrightarrow{\mathsf{D}}\mathsf{N}_m(0,\sigma^2\Sigma^{-1}\Phi\Sigma^{-1}), \tag{5.11}$$

where

$$\begin{aligned}(1-\alpha)^2\Phi &= \tau_2^c(1+2\xi_1^c)\Sigma+(\xi_1^c)^2(\Sigma_2\Sigma_1^{-1}\Sigma_2+\Sigma_1\Sigma_2^{-1}\Sigma_1)\\ &+\tau_3^c\frac{\xi_2^c}{2}(\mu_2\mu_1'+\mu_1\mu_2')(\frac{1}{\lambda_1}+\frac{1}{\lambda_2})+(\tau_4-1)(\frac{\xi_2^c}{2})^2(\frac{\mu_2\mu_2'}{\lambda_1}+\frac{\mu_1\mu_1'}{\lambda_2})\\ &+\tau_3\frac{\xi_1^c\xi_2^c}{2}(\frac{\mu_2\mu_1'\Sigma_1^{-1}\Sigma_2+\Sigma_2\Sigma_1^{-1}\mu_1\mu_2'}{\lambda_1}+\frac{\mu_1\mu_2'\Sigma_2^{-1}\Sigma_1+\Sigma_1\Sigma_2^{-1}\mu_2\mu_1'}{\lambda_2}).\end{aligned}$$

A closer look at the expression for Φ shows that it is block diagonal. The variance for the autoregressive components is $(1-\alpha)^2\Phi_Y = \Sigma_Y\{\tau_2^c(1+2\xi_1^c)+(\xi_1^c)^2(\lambda_2^2\lambda_1^{-1}+\lambda_1^2\lambda_2^{-1})\}$. The somewhat complicated limiting covariance matrix for the deterministic terms, Φ_D, simplifies in two important special cases highlighted in the next Corollary. This covers the case where the reference density f is symmetric so $\xi_2^c = 0$ and the terms involving μ_j disappear. Alternatively, the proportionality $\Sigma_{D,j} = \lambda_j I_\ell$ and $\mu_{D,j} = \lambda_j\mu_D$ would also simplify the covariance. In §5.3 it is shown how this proportionality can be achieved by choosing the index sets in a particular way.

Corollary 5.3 *If* f *is symmetric then* $\xi_2^c = 0$ *so*

$$(1-\alpha)^2\Phi = \tau_2^c(1+2\xi_1^c)\Sigma + (\xi_1^c)^2(\Sigma_2\Sigma_1^{-1}\Sigma_2 + \Sigma_1\Sigma_2^{-1}\Sigma_1).$$

If $\Sigma_{D,j} = \lambda_j I_\ell$ *and* $\mu_{D,j} = \lambda_j\mu_D$ *then* $\Sigma_j = \lambda_j\Sigma$ *and* $\mu_j = \lambda_j\mu$ *so* $\Phi = \eta_\beta\Sigma + \kappa_\beta\mu\mu'$, *where the constants* η_β, κ_β *were defined in Theorem 4.1.*

5.3 Choice of index sets in the non-stationary case

Corollary 5.3 showed that the limiting distribution for the trend stationary case reduces to that of the strictly stationary case in the presence of proportionality, that is, if $\Sigma_{D,j} = \lambda_j I_\ell$ and $\mu_{D,j} = \lambda_j\mu_D$. This can be achieved if the index sets are chosen carefully. The key is that the index sets are, up to an approximation, alternating and dense in $[0,1]$, so that for any $0 \le u \le v \le 1$,

$$\frac{1}{T}\sum_{t\in\mathrm{int}(Tu)+1}^{\mathrm{int}(Tv)} 1_{(t\in\mathcal{I}_j)} \to \lambda_j\,(v-u)\,, \tag{5.12}$$

where $\lambda_1 + \lambda_2 = 1$. The alternating nature of the sets allows information to be accumulated in a proportional fashion over the two sub-samples, even though the process at hand is trend stationary. Two schemes for choosing the index sets are considered. First, a random scheme which is, perhaps, most convenient in applications, and, secondly, a deterministic scheme. The random scheme is not far from what has been applied in some Monte Carlo simulation experiments made by David Hendry in similar situations.

5.3.1 Random index sets

We will consider one particular index set which is alternating in a random way. Generate a series of independent Bernoulli variables, $\varsigma_1, \ldots, \varsigma_T$ taking the values 1 and 2 so that

$$\mathsf{P}\,(\varsigma_t = 1) = \lambda_1, \qquad \mathsf{P}\,(\varsigma_t = 2) = \lambda_2, \qquad \text{so} \qquad \lambda_1 + \lambda_2 = 1$$

for some $0 \le \lambda_1, \lambda_2 \le 1$. Then form the index sets

$$\mathcal{I}_1 = (t : \varsigma_t = 1) \qquad \text{and} \qquad \mathcal{I}_2 = (t : \varsigma_t = 2)\,.$$

The index sequence has to be independent of the generating process for the data, so that the data can be analysed conditionally on the index sets. In the following we will comment on examples of deterministic processes and unit root processes.

Consider the trend stationary model in (5.1). Since the index sets are constructed by independent sampling then

$$\begin{aligned}
\mathsf{E}(N_T \sum_{t\in\mathcal{I}_j} x_t x_t' N_T') &= \mathsf{E}\{N_T \sum_{t=1}^{T} (x_t x_t') N_T'\}\mathsf{E}1_{(\varsigma_t=j)} = \mathsf{E}\{N_T \sum_{t=1}^{T} x_t x_t' N_T'\}\lambda_j \to \lambda_j \Sigma, \\
\mathsf{E}(T^{-1/2} N_T \sum_{t\in\mathcal{I}_j} x_t) &= \mathsf{E}(T^{-1/2} N_T \sum_{t=1}^{T} x_t)\mathsf{E}1_{(\varsigma_t=j)} = \mathsf{E}(T^{-1/2} N_T \sum_{t=1}^{T} x_t)\lambda_j \to \lambda_j \mu.
\end{aligned}$$

5.3.2 Alternating index sets

It is instructive also to consider an index set, which is alternating in a deterministic way. That is

$$\mathcal{I}_1 = (t \text{ is odd}) \qquad \text{and} \qquad \mathcal{I}_2 = (t \text{ is even}).$$

This index set satisfies the property (5.12) with $\lambda_1 = \lambda_2 = 1/2$.

Consider the trend stationary model in (5.1) where the eigenvalues of the deterministic transition matrix D are all at one, so only polynomial trends are allowed. For simplicity restrict the calculations to a bivariate deterministic terms and let T be even, so with

$$d_t = \begin{pmatrix} 1 \\ t \end{pmatrix}, \qquad Q_T = \begin{pmatrix} 1 & 0 \\ 0 & T^{-1} \end{pmatrix},$$

the desired proportionality then follows, in that

$$\begin{aligned}
T^{-1} Q_T \sum_{t\in\mathcal{I}_j} d_t d_t' Q_T &= T^{-1} Q_T \sum_{t=0}^{T/2-1} d_{2t+j} d_{2t+j}' Q_T \to \frac{1}{2}\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}, \\
T^{-1} Q_T \sum_{t\in\mathcal{I}_j} d_t &= T^{-1} Q_T \sum_{t=0}^{T/2-1} d_{2t+j} \to \frac{1}{2}\begin{pmatrix} 1 \\ 1/2 \end{pmatrix}.
\end{aligned}$$

The proportionality will, however, fail if the process has a seasonal component with the same frequency as the alternation scheme. If for instance $d_t = (-1)^t$ and T even then it holds that

$$\mu_{D,1} = T^{-1} \sum_{t\in\mathcal{I}_1} (-1)^t = -\frac{1}{2}, \qquad \mu_{D,2} = T^{-1} \sum_{t\in\mathcal{I}_2} (-1)^t = \frac{1}{2}, \qquad \mu = T^{-1} \sum_{t=1}^{T} (-1)^t = 0,$$

so $\mu_{D,j} \neq \lambda_j \mu$, and proportionality does not hold. The proportionality will only arise when information is accumulated proportionally over the two index sets, either by choosing them randomly or by constructing them to be out of sync with the seasonality, for instance by choosing the first index set as every third observation.

5.4 A few results for unit root processes.

Consider the first order autoregression

$$X_t = \beta X_{t-1} + \varepsilon_t, \tag{5.13}$$

where $\beta = 1$ gives the unit root situation, and we assume for simplicity that f is symmetric so $\xi_2^c = 0$ and the term involving k_t falls away. The Functional Central Limit Theorem shows that

$$T^{-1/2} \sum_{t=1}^{\text{int}(Tu)} \left\{ \begin{array}{c} \varepsilon_t 1_{(t\in\mathcal{I}_1)} \\ \varepsilon_t 1_{(t\in\mathcal{I}_2)} \\ \varepsilon_t 1_{(|\varepsilon_t|<c)} \end{array} \right\} \xrightarrow{\mathsf{D}} \left(\begin{array}{c} w_{1u} \\ w_{2u} \\ w_u^c \end{array} \right) = W_u,$$

where W_u is a Brownian motion with variance matrix

$$\widetilde{\Omega} \stackrel{\text{def}}{=} \sigma^2 \left(\begin{array}{ccc} \lambda_1 & 0 & \lambda_1\tau_2^c \\ 0 & \lambda_2 & \lambda_2\tau_2^c \\ \lambda_1\tau_2^c & \lambda_2\tau_2^c & \tau_2^c \end{array} \right).$$

From the decomposition

$$\sum_{t\in\mathcal{I}_j} X_{t-1}^2 = \sum_{t=1}^T X_{t-1}^2 1_{(t\in\mathcal{I}_j)} = \sum_{t=1}^T X_{t-1}^2 \lambda_j + \sum_{t=1}^T X_{t-1}^2 \left\{ 1_{(t\in\mathcal{I}_j)} - \lambda_j \right\},$$

it is seen that the first term is of order T^2, whereas the second term has mean zero and variance $\lambda_1\lambda_2 E(\sum_{t=1}^T X_{t-1}^4)$; it is therefore of order $T^{3/2}$. It follows that

$$\frac{1}{T^2}(\sum_{t\in\mathcal{I}_1} X_{t-1}^2, \sum_{t\in\mathcal{I}_2} X_{t-1}^2, \sum_{t=1}^T X_{t-1}^2) \xrightarrow{\mathsf{D}} (\lambda_1, \lambda_2, 1) \int_0^1 w_u^2 du,$$

where $w_u = w_{1u} + w_{2u}$ is the Brownian motion generated by the cumulated ε_t. The information accumulated over each of the two sub-samples are therefore proportional to $\int_0^1 w_u^2 du$. It follows from Theorem 3.1, that the first round indicator saturated estimator satisfies

$$T(\widetilde{\beta} - 1) \xrightarrow{\mathsf{D}} \frac{\int_0^1 w_u d\left\{ w_u^c + 2c\mathsf{f}(c) \left(\lambda_1^{-1}\lambda_2 w_{1u} + \lambda_2^{-1}\lambda_1 w_{2u} \right) \right\}}{(1-\alpha)^2 \int_0^1 w_u^2 du}.$$

When $c \to \infty$ then $w_u^c \xrightarrow{\mathsf{D}} w_u$ while $c\mathsf{f}(c) \to 0$ and $\alpha \to 0$ giving the usual Dickey-Fuller distribution,

$$T(\hat{\beta} - 1) \xrightarrow{\mathsf{D}} \frac{\int_0^1 w_u dw_u}{\int_0^1 w_u^2 du}.$$

While the limiting distribution is now different from the stationary case, the relevant modification corresponds to the usual modification of normal distributions into Dickey-Fuller-type distributions when moving from the stationary to the non-stationary case.

Nearly the same arguments apply as with random index sets. In this case the definition of the Brownian motions becomes

$$T^{-1/2}\sum_{t=1}^{\mathrm{int}(Tu/2)}\left\{\begin{array}{c}\varepsilon_{2t-1}\\ \varepsilon_{2t}\\ \varepsilon_t 1_{(|\varepsilon_t|<c)}\end{array}\right\}\xrightarrow{\mathsf{D}}\left(\begin{array}{c}w_{1u}\\ w_{2u}\\ w_u^c\end{array}\right)=W_u.$$

6 Proof of main result

The results of Theorem 3.1 concern the matrices

$$N_T S_{xx} N_T' = \sum_{t=1}^{T} N_T x_t x_t' N_T' 1_{(\underline{c}\le v_t\le \overline{c})}, \qquad N_T S_{x\varepsilon} = \sum_{t=1}^{T} N_T x_t \varepsilon_t 1_{(\underline{c}\le v_t\le \overline{c})}.$$

For $N_T S_{xx} N_T'$ the main idea in the proof is to approximate $\hat{\omega}_t v_t = \varepsilon_t - (\hat{\beta}-\beta)' x_t$ by ε_t and the indicator $1_{(\underline{c}\le v_t\le\overline{c})}$ by $1_{(\underline{c}\sigma\le\varepsilon_t\le\overline{c}\sigma)}$, because the limit of the approximation $\sum_{t=1}^{T} N_T x_t x_t' N_T' 1_{(\underline{c}\sigma\le\varepsilon_t\le\overline{c}\sigma)}$ is easy to find. It turns out that the approximation involves terms from the preliminary estimator of β and σ. In the proof of Theorem 3.1 this replacement is justified using techniques for empirical processes and in particular Koul (2002, Theorem 7.2.1, p.298).

We define the normalised regressors $x_{Tt} = T^{1/2} N_T x_t$ and the estimation errors $\hat{a}_{Tt} = \hat{\omega}_t - \sigma$, $\hat{a}_T = \hat{\sigma} - \sigma$ and $\hat{b}_T = T^{-1/2}(N_T^{-1})'(\hat{\beta}-\beta)$. Then $T^{1/2}(\hat{a}_T, \hat{b}_T) = \mathrm{O}_{\mathsf{P}}(1)$ and $T^{1/2}\max_{1\le t\le T}|\hat{a}_{Tt} - \hat{a}_T| = T^{1/2}\max_{1\le t\le T}|\hat{\omega}_t - \hat{\sigma}| = \mathrm{o}_{\mathsf{P}}(1)$ by assumption (i) of Theorem 3.1. Note that

$$\hat{\omega}_t v_t = \varepsilon_t - (\hat{\beta}-\beta)' x_t = \varepsilon_t - \{T^{-1/2}(N_T^{-1})'(\hat{\beta}-\beta)\}'(T^{1/2} N_T x_t) = \varepsilon_t - \hat{b}_T' x_{Tt}, \quad (6.1)$$

so that

$$(\underline{c}\le v_t\le\overline{c}) = \{\underline{c}\,(\sigma+\hat{a}_{Tt}) \le \varepsilon_t - \hat{b}_T' x_{Tt} \le \overline{c}\,(\sigma+\hat{a}_{Tt})\}.$$

We define $u = (a, b')'$ and

$$I_t(u) = I_t(a,b) = 1_{\{\underline{c}(\sigma+a)\le\varepsilon_t - b' x_{Tt}\le\overline{c}(\sigma+a)\}} - 1_{(\underline{c}\sigma\le\varepsilon_t\le\overline{c}\sigma)}, \quad (6.2)$$

and find for the denominator $N_T S_{xx} N_T'$

$$\begin{aligned} N_T S_{xx} N_T' &= T^{-1}\sum_{t=1}^{T} x_{Tt} x_{Tt}' 1_{(\underline{c}\le v_t\le\overline{c})} = T^{-1}\sum_{t=1}^{T} x_{Tt} x_{Tt}' 1_{(\underline{c}\sigma\le\varepsilon_t\le\overline{c}\sigma)} \quad (6.3)\\ &\quad + T^{-1}\sum_{t=1}^{T} x_{Tt} x_{Tt}'\{I_t(\hat{a}_{Tt},\hat{b}_T) - I_t(\hat{a}_T,\hat{b}_T)\} + T^{-1}\sum_{t=1}^{T} x_{Tt} x_{Tt}' I_t(\hat{a}_T,\hat{b}_T) \end{aligned}$$

We then have to show that $\hat{a}_{tT}$ is so close to $\hat{a}_T$ that the second term tends to zero, and if we can show that $T^{-1}\sum_{t=1}^{T} x_{Tt} x_{Tt}' I_t(a,b)$ is tight as a process in (a,b) and because $T^{-1}\sum_{t=1}^{T} x_{Tt} x_{Tt}' I_t(0,0) = 0$, and $(\hat{a}_T, \hat{b}_T) = \mathrm{O}_{\mathsf{P}}(T^{1/2})$, we find that the last

term tends to zero. Finally we find from the Law of Large Numbers the probability limit of the first term.

Similarly we find for $N_T S_{x\varepsilon}$

$$\begin{aligned} N_T S_{x\varepsilon} &= T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t 1_{(\underline{c}\le v_t\le \bar{c})} = T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t 1_{(\underline{c}\sigma\le \varepsilon_t\le \bar{c}\sigma)} \\ &\quad + T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t\{I_t(\hat{a}_{tT},\hat{b}_T) - I_t(\hat{a}_T,\hat{b}_T)\} + T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t I_t(\hat{a}_T,\hat{b}_T) \end{aligned}$$

The limit of the second term will be shown to be zero because $\hat{a}_{Tt}$ is very close to $\hat{a}_T$. We get a contribution from the third term, which we decompose at the point (a,b) as

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t I_t(a,b) &= T^{-1/2}\sum_{t=1}^{T} x_{Tt}[\varepsilon_t I_t(a,b) - \mathsf{E}_{t-1}\{\varepsilon_t I_t(a,b)\}] \\ &\quad + T^{-1/2}\sum_{t=1}^{T} x_{Tt}\mathsf{E}_{t-1}\{\varepsilon_t I_t(a,b)\}. \end{aligned}$$

The first of these tends to zero, and for the second we find that a linear approximation to the smooth function $\mathsf{E}_{t-1}\{\varepsilon_t I_t(a,b)\}$ is $a\xi_2^c + b' x_{Tt}\xi_1^c$, and we therefore introduce the processes, for $\ell, m = 0, 1, 2$,

$$M_T^{\ell,m} = T^{-1/2}\sum_{t=1}^{T} g_m(x_{Tt})\,\varepsilon_t^\ell\{I_t(\hat{a}_{Tt},\hat{b}_T) - I_t(\hat{a}_T,\hat{b}_T)\} \tag{6.4}$$

$$W_T^{\ell,m}(a,b) = \frac{1}{T}\sum_{t=1}^{T} g_m(x_{Tt})\,\varepsilon_t^\ell I_t(a,b) \tag{6.5}$$

$$V_T^{\ell,m}(a,b) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_m(x_{Tt})\left\{\varepsilon_t^\ell I_t(a,b) - \sigma^{\ell-1}(a\xi_{\ell+1}^c + b' x_{Tt}\xi_\ell^c)\right\}, \tag{6.6}$$

where the function g_m is given as

$$g_0(x_{Tt}) = 1, \qquad g_1(x_{Tt}) = x_{Tt}, \qquad g_2(x_{Tt}) = x_{Tt}x_{Tt}'. \tag{6.7}$$

Lemma 6.4 below shows that $\sigma^{\ell-1}(a\xi_{\ell+1}^c + b' x_{Tt}\xi_\ell^c)$ is an approximation to the conditional mean of $\varepsilon_t^\ell I_t(a,b)$ given the past. Theorems 6.5, 6.6, and 6.7 below show that as $T\to\infty$ and if $T^{1/2}(\hat{a}_T,\hat{b}_T)$ is tight, then

$$M_T^{\ell,m} \xrightarrow{\mathsf{P}} 0, \qquad W_T^{\ell,m}(\hat{a}_T,\hat{b}_T) \xrightarrow{\mathsf{P}} 0 \qquad \text{and} \qquad V_T^{\ell,m}(\hat{a}_T,\hat{b}_T) \xrightarrow{\mathsf{P}} 0,. \tag{6.8}$$

Some equalities and expansions are established initially in §6.1. The remainder terms are analysed in §6.2. Finally, the threads are pulled together in a proof of Theorem 3.1 in §6.3.

6.1 Some initial inequalities and expansions

We define the indicator function $1_{(e\leq\varepsilon\leq f)}$ as

$$1_{(e\leq\varepsilon\leq f)} = 1_{(e\leq f)}\{1_{(\varepsilon\leq f)} - 1_{(\varepsilon\leq e)}\}.$$

We first prove an inequality for differences of such indicator functions.

Lemma 6.1 *For $e < f$, $e_0 < f_0$, and $\zeta \geq \max(|e - e_0|, |f - f_0|)$ we have*

$$|1_{(e\leq\varepsilon\leq f)} - 1_{(e_0\leq\varepsilon\leq f_0)}| \leq 1_{(|\varepsilon-e_0|\leq\zeta)} + 1_{(|\varepsilon-f_0|\leq\zeta)}.$$

Proof of Lemma 6.1. From $e = e_0+(e-e_0)$ and $|e-e_0| \leq \zeta$ we find $e_0-\zeta \leq e \leq e_0+\zeta$ and similarly $f_0 - \zeta \leq f \leq f_0 + \zeta$. Hence using the monotonicity in e and f, we find

$$1_{(e_0+\zeta\leq\varepsilon\leq f_0-\zeta)} \leq 1_{(e\leq\varepsilon\leq f)} \leq 1_{(e_0-\zeta\leq\varepsilon\leq f_0+\zeta)}.$$

Because the same inequalities hold for $1_{\{e_0\leq\varepsilon\leq f_0\}}$ we find

$$|1_{(e\leq\varepsilon\leq f)} - 1_{(e_0\leq\varepsilon\leq f_0)}| \leq 1_{(e_0-\zeta\leq\varepsilon\leq f_0+\zeta)} - 1_{(e_0+\zeta\leq\varepsilon\leq f_0-\zeta)} \leq 1_{(|\varepsilon-e_0|\leq\zeta)} + 1_{(|\varepsilon-f_0|\leq\zeta)},$$

where the last inequality is found by exploiting that $e_0 \leq f_0$ by assumption so

$$1_{(e_0-\zeta\leq\varepsilon\leq f_0+\zeta)} = 1_{(e_0-\zeta\leq f_0+\zeta)}\{1_{(\varepsilon\leq f_0+\zeta)} - 1_{(\varepsilon\leq e_0-\zeta}\} = 1_{(\varepsilon\leq f_0+\zeta)} - 1_{(\varepsilon\leq e_0-\zeta)},$$

whereas $1_{(e_0+\zeta>f_0-\zeta)}\{1_{(\varepsilon\leq e_0+\zeta)} - 1_{(\varepsilon\leq f_0-\zeta)}\} \geq 0$ so

$$-1_{(e_0+\zeta\leq\varepsilon\leq f_0-\zeta)} = 1_{(e_0+\zeta\leq f_0-\zeta)}\{1_{(\varepsilon\leq e_0+\zeta)} - 1_{(\varepsilon\leq f_0-\zeta)}\} \leq 1_{(\varepsilon\leq e_0+\zeta)} - 1_{(\varepsilon\leq f_0-\zeta)}.$$

■

Now, apply this result to the indicator function $I_t(u)$ introduced in (6.2). Note that $I_t(0) = 0$ and introduce the notation, for some $\delta > 0$, and $c = \max(|\underline{c}|, |\overline{c}|)$,

$$J_t(u, \delta) = 1_{\{|\varepsilon_t-\underline{c}(\sigma+a)-b'x_{Tt}|\leq\delta(c+|x_{Tt}|)\}} + 1_{\{|\varepsilon_t-\overline{c}(\sigma+a)-b'x_{Tt}|\leq\delta(c+|x_{Tt}|)\}}.$$

Lemma 6.2 *For $u = (a, b')'$, $u_0 = (a_0, b_0')'$ and $|u - u_0| \leq \delta$ we have*

$$|I_t(u) - I_t(u_0)| \leq J_t(u_0, \delta)$$

Proof of Lemma 6.2. The object of interest is

$$I_t(u) - I_t(u_0) = 1_{\{\underline{c}(\sigma+a)+b'x_{Tt}\leq\varepsilon_t\leq\overline{c}(\sigma+a)+b'x_{Tt}\}} - 1_{\{\underline{c}(\sigma+a_0)+b_0'x_{Tt}\leq\varepsilon_t\leq\overline{c}(\sigma+a_0)+b_0'x_{Tt}\}}.$$

The inequality follows from Lemma 6.1 by the choice $e = \underline{c}(\sigma + a) + b'x_{Tt}$, $e_0 = \underline{c}(\sigma + a_0) + b_0'x_{Tt}$, $f = \overline{c}(\sigma + a) + b'x_{Tt}$, $f_0 = \overline{c}(\sigma + a_0) + b_0'x_{Tt}$, and $\zeta = \delta(c + |x_{Tt}|)$. ■

Introduce the notation E_{t-1} for the expectation conditional on the information given by $(x_s, \varepsilon_s, s \leq t - 1, x_t)$.

Lemma 6.3 *For $\ell \in \mathbb{N}_0$, let $u = (a, b')'$, $u_0 = (a_0, b_0')'$ be random and $\mathsf{E}|\varepsilon_t|^\ell < \infty$. Then it holds with $c = \max(|\underline{c}|, |\overline{c}|)$ that*

$$\begin{aligned}\mathsf{E}_{t-1}\{1_{(|u-u_0|\le\delta)}|\varepsilon_t|^\ell |I_t(u) - I_t(u_0)|\} &\le \mathsf{E}_{t-1}|\varepsilon_t|^\ell J_t(u_0,\delta)\\ &\le 4\delta\sigma^{\ell-1}(c + |x_{Tt}|)\sup_{v\in\mathbb{R}}|v|^\ell \mathsf{f}(v).\end{aligned}$$

Proof of Lemma 6.3. The first inequality follows from Lemma 6.2. The function $J_t(u_0, \delta)$ is nonzero on two intervals of total length $4\delta(c + |x_{Tt}|)$, and the integrand $|\varepsilon_t|^\ell \mathsf{f}(\varepsilon_t/\sigma)/\sigma$ is bounded by $\sigma^{\ell-1}\sup_{v\in\mathbb{R}}|v|^\ell \mathsf{f}(v)$, so that the second inequality holds. ■

Finally, an approximation to the conditional expectation of $\varepsilon_t I_t(u)$ follows.

Lemma 6.4 *Let f have derivative f'. For $u = (a, b')'$ and $|u| \le \delta$ it holds for $\ell \in \mathbb{N}_0$*

$$\left|\mathsf{E}_{t-1}\left\{\varepsilon_t^\ell I_t(u)\right\} - \sigma^{\ell-1}(a\xi_{\ell+1}^c + b'x_{Tt}\xi_\ell^c)\right| \le 2\delta^2 \sup_{v\in\mathbb{R}}\{\ell|v|^{\ell-1}\mathsf{f}(v) + |v^\ell \mathsf{f}'(v)|\}(c^2 + |x_{Tt}|^2),$$

where $c = \max(|\underline{c}|, |\overline{c}|)$ and $\xi_\ell^c = (\overline{c})^\ell \mathsf{f}(\overline{c}) - (\underline{c})^\ell \mathsf{f}(\underline{c})$.

Proof of Lemma 6.4. Let $\psi(\varepsilon) = (\varepsilon/\sigma)^\ell \mathsf{f}(\varepsilon/\sigma)$. A second order Taylor expansion gives

$$\int_{c\sigma}^{c\sigma+h} \psi(\varepsilon)d\varepsilon = h\psi(c\sigma) + \frac{1}{2}h^2\psi'(\sigma c^*),$$

for c^* satisfying $|\sigma c - \sigma c^*| \le h$. Thus

$$\sigma^{1-\ell}\mathsf{E}_{t-1}\left\{\varepsilon_t^\ell I_t(u)\right\} = \int_{\underline{c}(\sigma+a)+b'x_{Tt}}^{\overline{c}(\sigma+a)+b'x_{Tt}} \psi(\varepsilon)d\varepsilon - \int_{\underline{c}\sigma}^{\overline{c}\sigma} \psi(\varepsilon)d\varepsilon = \overline{S} - \underline{S},$$

where

$$\begin{aligned}\overline{S} &= (\overline{c}a + b'x_{Tt})\psi(\overline{c}\sigma) + \frac{1}{2}(\overline{c}a + b'x_{Tt})^2\psi'(\sigma c_1^*),\\ \underline{S} &= (\underline{c}a + b'x_{Tt})\psi(\underline{c}\sigma) + \frac{1}{2}(\underline{c}a + b'x_{Tt})^2\psi'(\sigma c_2^*).\end{aligned}$$

Using $\psi(c\sigma) = c^\ell \mathsf{f}(c)$ the first order term of $\overline{S} - \underline{S}$ is

$$(\overline{c}a + b'x_{Tt})(\overline{c})^\ell \mathsf{f}(\overline{c}) - (\underline{c}a + b'x_{Tt})(\underline{c})^\ell \mathsf{f}(\underline{c}) = a\xi_{\ell+1}^c + b'x_{Tt}\xi_\ell^c.$$

Using $(|c|a + b'x_{Tt})^2 \le 2\delta^2(c^2 + |x_{Tt}|^2)$ the second order term is bounded by

$$2\delta^2(c^2 + |x_{Tt}|^2)\sup_{v\in\mathbb{R}}|\psi'(v)| \le 2\delta^2(c^2 + |x_{Tt}|^2)\sup_{v\in\mathbb{R}}\{\ell|v|^{\ell-1}\mathsf{f}(v) + |v^\ell \mathsf{f}'(v)|\}.$$

■

6.2 Some limit results

The first result on $M_T^{\ell,m}$ shows that we can replace the estimator, $\hat{\omega}_t^2$, of the variance of the residuals with $\hat{\sigma}^2$.

Theorem 6.5 *Let $\ell \in \mathbb{N}_0$ and $m \in \{0,1,2\}$. Suppose that*
(i) $\theta_T T^{1/2} \max_{1\leq t\leq T} |\hat{a}_{tT} - \hat{a}_T| = \mathrm{O_P}(1)$, for some $\theta_T \to \infty$
(ii) $\max_{t\leq T} \mathsf{E}\,|x_{Tt}|^3 = \mathrm{O}(1)$,
(iii) $\sup_v |v|^\ell \mathsf{f}(v) < \infty$ and $\mathsf{E}|\varepsilon_t|^\ell < \infty$. Then it holds for $T \to \infty$ that

$$M_T^{\ell,m} = \frac{1}{T^{1/2}} \sum_{t=1}^{T} g_m\left(x_{Tt}\right) \varepsilon_t^\ell \{I_t(\hat{a}_{tT}, \hat{b}_T) - I_t(\hat{a}_T, \hat{b}_T)\} \xrightarrow{\mathsf{P}} 0$$

Proof of Theorem 6.5. Due to condition (i), for all $\zeta > 0$ there exists a $U > 0$ so that for large T then $\mathsf{P}(\theta_T T^{1/2} \max_{1\leq t\leq T} |\hat{a}_{tT} - \hat{a}_T| \leq U) \geq 1 - \zeta$. Thus, with $\delta_T = UT^{-1/2}\theta_T^{-1}$, it suffices to show that $|M_T^{\ell,m}| 1_{(\max_{1\leq t\leq T} |\hat{a}_{tT} - \hat{a}_T| \leq \delta_T)} \xrightarrow{\mathsf{P}} 0$, and in turn by the Markov inequality it suffices to show $S = \mathsf{E}|M_T^{\ell,m}| 1_{(\max_{1\leq t\leq T} |\hat{a}_{tT} - \hat{a}_T| \leq \delta_T)} \to 0$.

Using the triangle inequality and taking iterated expectations it holds

$$S \leq \frac{1}{T^{1/2}} \sum_{t=1}^{T} \mathsf{E}|x_{Tt}|^m \mathsf{E}_{t-1}\{\varepsilon_t^\ell |I_t(\hat{a}_{tT}, \hat{b}_T) - I_t(\hat{a}_T, \hat{b}_T)| 1_{(\max_{1\leq t\leq T} |\hat{a}_{tT} - \hat{a}_T| \leq \delta_T)}\}.$$

Lemma 6.2 then shows

$$S \leq 4\delta_T T^{1/2} \sigma^{\ell-1} \sup_{v\in\mathbb{R}} \{|v|^\ell \mathsf{f}(v)\} a^{-1} T^{-1} \mathsf{E} \sum_{t=1}^{T} |x_{Tt}|^m (c + |x_{Tt}|).$$

This vanishes since $\delta_T T^{1/2} \to 0$ and the other terms are bounded. ∎

Theorem 6.6 *Let $\ell \in \mathbb{N}_0$ and $m \in \{0,1,2\}$. Suppose that*
(i) $(\hat{a}_T, \hat{b}_T) = \mathrm{O_P}(T^{-1/2})$,
(ii) $\max_{t\leq T} \mathsf{E}\,|x_{Tt}|^{m+1} = \mathrm{O}(1)$,
(iii) $\sup_v |v|^\ell \mathsf{f}(v) < \infty$ and $\mathsf{E}|\varepsilon_t|^\ell < \infty$.
Then it holds for $T \to \infty$ that

$$W_T^{\ell,m}(\hat{a}_T, \hat{b}_T) = \frac{1}{T} \sum_{t=1}^{T} g_m\left(x_{Tt}\right) \varepsilon_t^\ell I_t(\hat{a}_T, \hat{b}_T) \xrightarrow{\mathsf{P}} 0, \tag{6.9}$$

where g_m was defined in (6.7) *as* 1, x_{Tt}, $x_{Tt}x_{Tt}'$ *for* $m = 0,1,2$, *so that* $|g_m(x_{Tt})| \leq |x_{Tt}|^m$.

Proof of Theorem 6.6. Due to condition (i), for all $\zeta > 0$ there exists a $U > 0$ so that for large T then $\mathsf{P}\{|(\hat{a}_T, \hat{b}_T)| \leq T^{-1/2}U\} \geq 1 - \zeta$. Thus, it suffices to show that

$\sup_{|u|\leq T^{-1/2}U}|W_T^{\ell,m}(u)|\xrightarrow{\mathsf{P}}0$, and in turn by the Markov inequality it suffices to show that $\mathsf{E}\sup_{|u|\leq T^{-1/2}U}|W_T^{\ell,m}(u)|\to 0$.

Because $I_t(0)=0$ then $I_t(u)=I_t(u)-I_t(0)$. Lemma 6.2 then shows $|I_t(u)|\leq J_t\left(0,T^{-1/2}U\right)$ for $|u|\leq T^{-1/2}U$. Thus, using the triangle inequality it holds

$$\sup_{|u|\leq T^{-1/2}U}|W_T^{\ell,m}(u)|\leq\frac{1}{T}\sum_{t=1}^{T}|x_{Tt}|^m|\varepsilon_t|^\ell J_t(0,T^{-1/2}U).$$

Then take iterated expectations

$$S=\mathsf{E}\{\sup_{|u|\leq T^{-1/2}U}|W_T^{\ell,m}(u)|\}\leq\mathsf{E}\frac{1}{T}\sum_{t=1}^{T}|x_{Tt}|^m\mathsf{E}_{t-1}|\varepsilon_t|^\ell J_t(0,T^{-1/2}U). \tag{6.10}$$

Apply Lemma 6.3 with $\delta=T^{-1/2}U$ and find

$$\begin{aligned}S &\leq \mathsf{E}\frac{1}{T}\sum_{t=1}^{T}|x_{Tt}|^m\frac{4U\sigma^{\ell-1}}{T^{1/2}}(c+|x_{Tt}|)\sup_{v\in\mathbb{R}}|v|^\ell\mathsf{f}(v)\\ &= 4U\sigma^{\ell-1}\sup_{v\in\mathbb{R}}\{|v|^\ell\mathsf{f}(v)\}\frac{1}{T^{3/2}}\sum_{t=1}^{T}\{c\mathsf{E}\left(|x_{Tt}|^m\right)+\mathsf{E}(|x_{Tt}|^{m+1})\},\end{aligned}$$

which vanishes due to Assumptions (ii) and (iii). ■

Theorem 6.7 *Let $\ell\in\mathbb{N}_0$ and $m\in\{0,1\}$. Suppose that*
(i) $(\hat{a}_T,\hat{b}_T)=\mathrm{O}_\mathsf{P}(T^{-1/2})$,
(ii) $\max_{t\leq T}\mathsf{E}\left|x_{Tt}\right|^3=\mathrm{O}(1)$,
(iii) $\sup_{v\in\mathbb{R}}\left\{(\ell|v|^{\ell-1}+|v|^\ell+v^{2\ell})\mathsf{f}(v)+|v^\ell\mathsf{f}'(v)|\right\}<\infty$ *and* $\mathsf{E}|\varepsilon_t|^{2\ell}<\infty$.
Then it holds for $T\to\infty$ that

$$V_T^{\ell,m}(\hat{a}_T,\hat{b}_T)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_m(x_{Tt})\{\varepsilon_t^\ell I_t(\hat{a}_T,\hat{b}_T)-\sigma^{\ell-1}(\hat{a}_T\xi_{\ell+1}^c+\hat{b}_T'x_{Tt}\xi_\ell^c)\}\xrightarrow{\mathsf{P}}0.$$

Proof. As in the proof of Theorem 6.6, using condition (i), it suffices to show that $\sup_{|u|\leq T^{-1/2}U}|V_T^{\ell,m}(u)|\xrightarrow{\mathsf{P}}0$.

1. *Decompose* $V_T^{\ell,m}$ as a sum of martingale differences $\tilde{V}_T$ and a correction term $\overline{V}_T$ so $V_T^{\ell,m}(u)=\tilde{V}_T(u)+\overline{V}_T(u)$, where

$$\begin{aligned}\tilde{V}_T(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_m(x_{Tt})[\varepsilon_t^\ell I_t(u)-\mathsf{E}_{t-1}\{\varepsilon_t^\ell I_t(u)\}]\\ \overline{V}_T(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_m(x_{Tt})[\mathsf{E}_{t-1}\{\varepsilon_t^\ell I_t(u)\}-\sigma^{\ell-1}(\hat{a}_{tT}\xi_{\ell+1}^c+\hat{b}_T'x_{Tt}\xi_\ell^c)]\end{aligned}$$

It has to be shown that the supremum of each of these terms vanishes.

2. *The term* $\overline{V}_T(u)$. Using first the triangular inequality and then Lemma 6.4 with $\delta = T^{-1/2}U$ gives

$$\begin{aligned}\sup_{|u|\leq T^{-1/2}U}|\overline{V}_T(u)| &\leq \sup_{|u|\leq T^{-1/2}U}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|x_{Tt}|^m\left|\mathsf{E}_{t-1}\{\varepsilon_t^\ell I_t(u)\}-\sigma^{\ell-1}(a\xi_{\ell+1}^c+b'x_{Tt}\xi_\ell^c)\right| \\ &\leq 2U^2\sup_{v\in\mathbb{R}}\{\ell|v|^{\ell-1}\mathsf{f}(v)+|v^\ell\mathsf{f}'(v)|\}\frac{1}{T^{3/2}}\sum_{t=1}^{T}|x_{Tt}|^m\left(c^2+|x_{Tt}|^2\right) \\ &= \mathrm{O}_{\mathsf{P}}(T^{-1/2}),\end{aligned}$$

by Assumption (ii) and (iii), because $\max_{t\leq T}(\mathsf{E}\,|x_{Tt}|^m, \mathsf{E}\,|x_{Tt}|^{m+2})$ is bounded.

3. *The term* $\tilde{V}_T(u)$. For a given χ, to be chosen later, choose $|u_k|\leq UT^{-1/2}$, $k=1,\ldots,K$ and $B_k=(u:|u-u_k|\leq \chi T^{-1/2},\ |u|\leq UT^{-1/2})T^{-1/2}$ as a finite cover of $(u:|u|\leq UT^{-1/2})$. Thus, for any u we have $u\in B_k$ for some k. In particular, it holds for $u\in B_k$

$$|\tilde{V}_T(u)|\leq|\tilde{V}_T(u_k)|+|\tilde{V}_T(u)-\tilde{V}_T(u_k)|\leq\max_k|\tilde{V}_T(u_k)|+\max_k\sup_{u\in B_k}|\tilde{V}_T(u)-\tilde{V}_T(u_k)|.$$

4. *The term* $\max_k|\tilde{V}_T(u_k)|$. Because $\tilde{V}_T$ is a sum of martingale differences then

$$\mathsf{Var}\{\tilde{V}_T(u_k)\}=\frac{1}{T}\mathsf{E}\sum_{t=1}^{T}\left[g_m(x_{Tt})g_m(x_{Tt})'\mathsf{Var}_{t-1}\{\varepsilon_t^\ell I_t(u_k)\}\right].$$

From Lemma 6.2 with $u_0=0$, $I_t(0)=0$, and $|u_k|\leq UT^{-1/2}$ we have $\left\{\varepsilon_t^\ell I_t(u_k)\right\}^2\leq\varepsilon_t^{2\ell}J_t^2(0,UT^{-1/2})$. Further, by the inequality $(a+b)^2\leq 2(a^2+b^2)$ we have $J_t^2(0,UT^{-1/2})\leq 2J_t(0,UT^{-1/2})$, so that from Lemma 6.3 we find

$$\mathsf{E}_{t-1}\left\{\varepsilon_t^\ell I_t(u_k)\right\}^2\leq 2\mathsf{E}_{t-1}\varepsilon_t^{2\ell}J_t(0,UT^{-1/2})\leq 8\frac{U}{T^{1/2}}\sigma^{2\ell-1}(c+|x_{Tt}|)\sup_{v\in\mathbb{R}}|v|^{2\ell}\mathsf{f}(v). \quad (6.11)$$

Since $\mathsf{Var}_{t-1}\{\varepsilon_t^\ell I_t(u_k)\}\leq\mathsf{E}_{t-1}\left\{\varepsilon_t^\ell I_t(u_k)\right\}^2$ it then holds

$$\mathsf{Var}\{\tilde{V}_T(u_k)\}\leq\frac{8U\sigma^{2\ell-1}}{T^{3/2}}\sup_{v\in\mathbb{R}}\{v^{2\ell}\mathsf{f}(v)\}\sum_{t=1}^{T}\mathsf{E}\{|x_{Tt}|^{2m}(c+|x_{Tt}|)\}\leq\frac{c_0}{T^{1/2}},$$

because $\max_{t\leq T}(\mathsf{E}\,|x_{Tt}|^{2m}, \mathsf{E}\,|x_{Tt}|^{2m+1})$ is bounded. Using first Boole's inequality and then Chebychev's inequality it then holds for a $\zeta>0$ to be chosen later

$$\begin{aligned}\mathsf{P}\{\max_k|\tilde{V}_T(u_k)|\geq\zeta\} &= \mathsf{P}\bigcup_{k=1}^{K}\{|\tilde{V}_T(u_k)|\geq\zeta\}\leq\sum_{k=1}^{K}\mathsf{P}\{|\tilde{V}_T(u_k)|\geq\zeta\} \\ &\leq \frac{1}{\zeta^2}\sum_{k=1}^{K}\mathsf{Var}\{\tilde{V}_T(u_k)\}\leq\frac{c_0K}{T^{1/2}\zeta^2}\to 0, \quad (6.12)\end{aligned}$$

for fixed K (and χ) and $T \to \infty$.

5. *The term* $\max_k \sup_{u \in B_k} |\tilde{V}_T(u) - \tilde{V}_T(u_k)|$. The inequality in Lemma 6.2 shows

$$\sup_{u \in B_k} |\tilde{V}_T(u) - \tilde{V}_T(u_k)| \leq Z_T(k)$$

where

$$Z_T(k) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} |x_{Tt}|^m \left[|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi) + \mathsf{E}_{t-1}\{|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi)\} \right],$$

because $|u - u_k| \leq T^{-1/2}\chi$. Again, write Z_T as a sum of martingale differences $\tilde{Z}_T$ and a correction term $\overline{Z}_T$ so $Z_T(k) = \tilde{Z}_T(k) + \overline{Z}_T(k)$ where

$$\begin{aligned} \tilde{Z}_T(k) &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} |x_{Tt}|^m \left[|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi) - \mathsf{E}_{t-1}\{|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi)\} \right], \\ \overline{Z}_T(k) &= \frac{2}{T^{1/2}} \sum_{t=1}^{T} |x_{Tt}|^m \, \mathsf{E}_{t-1}\{|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi)\}. \end{aligned}$$

6. *The term* $\max_k \overline{Z}_T(k)$. Lemma 6.3 shows

$$\max_k \overline{Z}_T(k) \leq 8\chi \sup_{v \in \mathbb{R}} \{|v|^\ell \mathsf{f}(v)\} \frac{1}{T} \sum_{t=1}^{T} |x_{Tt}|^m (c + |x_{Tt}|) = \mathrm{O}_{\mathsf{P}}(\chi), \tag{6.13}$$

due to Assumptions (ii), (iii).

7. *The term* $\max_k \tilde{Z}_T(k)$. Since $\tilde{Z}_T(k)$ is a sum of martingale differences then

$$\mathsf{Var}\{\tilde{Z}_T(k)\} = \frac{1}{T} \sum_{t=1}^{T} \mathsf{E}[g_m(x_{Tt}) g_m(x_{Tt})' \mathsf{Var}_{t-1}\{|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi)\}].$$

Since $\mathsf{Var}_{t-1}\{|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi)\} \leq \mathsf{E}_{t-1}\{|\varepsilon_t|^\ell J_t(u_k, T^{-1/2}\chi)\}^2$ then (6.11) shows

$$\mathsf{Var}\{\tilde{Z}_T(k)\} \leq 4\chi\sigma^{2\ell-1} \sup_{v \in \mathbb{R}} \{v^{2\ell} \mathsf{f}(v)\} \frac{1}{T^{3/2}} \sum_{t=1}^{T} \mathsf{E}\{|x_{Tt}|^{2m} (c + |x_{Tt}|)\} = \mathrm{O}(T^{-1/2}),$$

because $\max_{t \leq T}(\mathsf{E}\,|x_{Tt}|^{2m}, \mathsf{E}\,|x_{Tt}|^{2m+1})$ is bounded, using Assumptions (ii) and (iii). Then, like the evaluation (6.12), we find

$$\mathsf{P}\{\max_k |\widetilde{Z}_T(k)| \geq \zeta\} \leq \frac{c_0 M}{T^{1/2}\zeta^2} \to 0.$$

8. The proof is now complete by noticing that for given $\zeta > 0$ and $\xi > 0$ we can first choose U so large that

$$\mathsf{P}\{T^{1/2}|(\hat{a}_T, \hat{b}_T)| \geq U\} \leq \xi,$$

using condition (i). Next choose χ so small that (6.13) is small. Finally, choose T so large that the remaining terms are small. ∎

6.3 Proof of main result

Proof of Theorem 3.1. We analyse the properties of the product moments:

$$\begin{aligned} S_{11} &= \sum_{t=1}^{T} 1_{(\underline{c}\leq v_t\leq \overline{c})}, \quad S_{xx} = \sum_{t=1}^{T} x_t x_t' 1_{(\underline{c}\leq v_t\leq \overline{c})}, \\ S_{x\varepsilon} &= \sum_{t=1}^{T} x_t \varepsilon_t 1_{(\underline{c}\leq v_t\leq \overline{c})}, \quad S_{x1} = \sum_{t=1}^{T} x_t 1_{(\underline{c}\leq v_t\leq \overline{c})}. \end{aligned}$$

We define $(\hat{a}_{tT}, \hat{b}_T) = \{\hat{\omega}_t - \sigma, T^{-1/2}(N_T^{-1})'(\hat{\beta} - \beta)\}$, and note, see (6.3) that the definition of $W_T^{\ell,m}(a,b)$ and $M_T^{\ell,m}$ implies that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} g_m(x_{Tt})\,\varepsilon_t^{\ell} 1_{(\underline{c}\leq v_t\leq \overline{c})} &= T^{-1}\sum_{t=1}^{T} g_m(x_{Tt})\,\varepsilon_t^{\ell} 1_{(\underline{c}\sigma\leq \varepsilon_t\leq \overline{c}\sigma)} + \\ &\quad + T^{-1/2} M_T^{\ell,m} + W_T^{\ell,m}(\hat{a}_T, \hat{b}_T), \end{aligned}$$

and that for $x_{Tt} = T^{1/2} N_T x_t$, Theorem 6.6 implies that $W_T^{\ell,m}(\hat{a}_T, \hat{b}_T) = \mathsf{o_P}(1)$ and Theorem 6.5 shows that $M_T^{\ell,m} = \mathsf{o_P}(1)$.

The limits (3.3), (3.4), and (3.5). For $m = 2,\ \ell = 0$ we find

$$N_T S_{xx} N_T' = N_T \sum_{t=1}^{T} x_t x_t' 1_{(\underline{c}\sigma\leq \varepsilon_t\leq \overline{c}\sigma)} N_T' + \mathsf{o_P}(1).$$

Note that $\mathsf{E}_{t-1}\{1_{(\underline{c}\sigma\leq \varepsilon_t\leq \overline{c}\sigma)}\} = 1-\alpha$, so a martingale decomposition of the main term on the right hand side is

$$N_T \sum_{t=1}^{T} x_t x_t' \{1_{(\underline{c}\sigma\leq \varepsilon_t\leq \overline{c}\sigma)} - (1-\alpha)\} N_T' + N_T \sum_{t=1}^{T} x_t x_t' N_T' (1-\alpha).$$

The first term vanishes due to Chebychev's inequality and Assumption (ii,c). The second term converges in probability to $(1-\alpha)\Sigma$ due to Assumption (ii,a).

The limit of S_{x1} is found by a similar argument for $m = 1, \ell = 0$, which gives

$$T^{-1/2} N_T \sum_{t=1}^{T} x_t 1_{(\underline{c}\leq v_t\leq \overline{c})} = T^{-1/2} N_T \sum_{t=1}^{T} x_t 1_{(\underline{c}\sigma\leq \varepsilon_t\leq \overline{c}\sigma)} + \mathsf{o_P}(1).$$

A martingale decomposition of the main term on the right hand side is

$$T^{-1/2} N_T \sum_{t=1}^{T} x_t \{1_{(\underline{c}\sigma\leq \varepsilon_t\leq \overline{c}\sigma)} - (1-\alpha)\} + T^{-1/2} N_T \sum_{t=1}^{T} x_t (1-\alpha).$$

The first term vanishes due to Chebychev's inequality and Assumption (ii,a). The second term converges to $(1-\alpha)\mu$ due to Assumption (ii,b).

Finally for $m = \ell = 0$, we find

$$T^{-1}\sum_{t=1}^{T} 1_{(\underline{c}\le v_t\le \bar{c})} = T^{-1}\sum_{t=1}^{T} 1_{(\underline{c}\sigma\le \varepsilon_t\le \bar{c}\sigma)} + \mathrm{o}_{\mathsf{P}}(1) \xrightarrow{\mathsf{P}} 1-\alpha.$$

The representations (3.6), (3.7), and (3.8): The definition of $V_T^{\ell,m}(\hat{a}_T, \hat{b}_T)$ implies that for $m = 0, 1, \ell = 0, 1, 2$ we have the representation

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{T} g_m(x_{Tt})\varepsilon_t^{\ell} 1_{\{\underline{c}\le v_t\le \bar{c}\}} &= M_T^{\ell,m} + V_T^{\ell,m}(\hat{a}_T, \hat{b}_T) \\ +T^{-1/2}\sum_{t=1}^{T} g_m(x_{Tt})[\varepsilon_t^{\ell} 1_{(\underline{c}\sigma\le \varepsilon_t\le \bar{c}\sigma)} &+ \sigma^{\ell-1}\{x_{Tt}'(\hat{\beta}-\beta)\xi_{\ell}^{c} + (\hat{\sigma}-\sigma)\xi_{\ell+1}^{c}\}], \end{aligned}$$

and that for $x_{Tt} = T^{1/2}N_T x_t$, Theorem 6.7 implies that $V_T^{\ell,m}(\hat{a}_T, \hat{b}_T) = \mathrm{o}_{\mathsf{P}}(1)$ and Theorem 6.5 shows that $M_T^{\ell,m} = \mathrm{o}_{\mathsf{P}}(1)$.

The representation of S_{11} follows for $\ell = m = 0$, and by noting that

$$T^{-1}\sum_{t=1}^{T} 1_{(\underline{c}\sigma\le \varepsilon_t< \bar{c}\sigma)} \xrightarrow{\mathsf{P}} 1-\alpha,$$

we have proved (3.4). The representation of $N_T S_{x\varepsilon}$ follows for $\ell = m = 1$. Finally the representation of term $S_{\varepsilon\varepsilon}$ follows for $m = 0, \ell = 2$. ■

Proof of Corollary 3.2. *Representation of* $(N_T^{-1})'(\breve{\beta}-\beta)$: From (3.1) we have

$$(N_T^{-1})'(\breve{\beta}-\beta) = (N_T S_{xx} N_T')^{-1} N_T S_{x\varepsilon}.$$

Because $N_T S_{xx} N_T' \xrightarrow{\mathsf{P}} (1-\alpha)\Sigma > 0$ by (3.4), we see that $\breve{\beta}$ is defined with probability tending to one, and the representation (3.9) follows from (3.6).

The representation of $T^{1/2}(\breve{\sigma}^2-\sigma^2)$: We use the expression, see (3.2), to show that

$$S_{11}^{-1}(S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}) = S_{11}^{-1}\{S_{\varepsilon\varepsilon} + \mathrm{O}_{\mathsf{P}}(1)\} \xrightarrow{P} \sigma^2\frac{\tau_2^c}{1-\alpha}.$$

This shows that we need to bias correct the empirical variance and therefore we consider

$$\breve{\sigma}^2 - \sigma^2 = (1-\alpha)(\tau_2^c)^{-1}S_{11}^{-1}(S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}) = (1-\alpha)(\tau_2^c)^{-1}S_{11}^{-1}\{S_{\varepsilon\varepsilon} + \mathrm{O}_{\mathsf{P}}(1)\},$$

and hence

$$\tau_2^c T^{1/2}(\breve{\sigma}^2-\sigma^2) = T^{1/2}(S_{\varepsilon\varepsilon} - \sigma^2\frac{\tau_2^c}{1-\alpha}S_{11}) + \mathrm{O}_{\mathsf{P}}(T^{-1/2}).$$

From (3.7) and (3.8) we find the representation

$$\begin{aligned} \tau_2^c T^{1/2}(\breve{\sigma}^2-\sigma^2) \;=\; &\{T^{-1/2}\sum_{t=1}^{T}(\varepsilon_t^2 - \sigma^2\frac{\tau_2^c}{1-\alpha})1_{(\sigma\underline{c}\le\varepsilon_t\le\sigma\bar{c})} \\ &+T^{1/2}(\hat{\sigma}-\sigma)\zeta_3^c + (\hat{\beta}-\beta)'N_T^{-1}T^{-1/2}N_T x_t\zeta_2^c)\} + \mathrm{o}_{\mathsf{P}}(1), \end{aligned}$$

which proves (3.10), because $T^{-1/2}N_T x_t \xrightarrow{\mathsf{P}} \mu$.

Consistency of the estimators: Finally it follows from the Assumption (i), (ii,a), (3.9), and (3.10) that $\{(N_T^{-1})'(\breve{\beta}-\beta), T^{1/2}(\breve{\sigma}^2-\sigma^2)\} = \mathrm{O}_{\mathsf{P}}(1)$, and $N_T \to 0$ and $T \to \infty$ then imply that $(\breve{\beta},\breve{\sigma}^2) \xrightarrow{\mathsf{P}} (\beta,\sigma^2)$. ■

7 Proofs for stationary and trend stationary cases

The proofs relating to §4 and §5 follow.

Proof of Theorem 4.1. We apply Corollary 3.2, using $N_T = T^{-1/2}I_m$. The least squares estimator based on the full sample satisfies condition (i,a): $T^{1/2}(\hat{\sigma}-\sigma, \hat{\beta}-\beta) = \mathrm{O}_{\mathsf{P}}(1)$, and the stationarity of x_t shows that conditions (ii,a,b,c) hold.

For the numerator of the estimator $\breve{\beta}_{LS}$ we therefore consider

$$T^{-1/2}\sum_{t=1}^{T} x_t\varepsilon_t 1_{(\sigma\underline{c}\le\varepsilon_t\le\bar{c}\sigma)} + \xi_1^c\Sigma T^{1/2}(\hat{\beta}-\beta) + \xi_2^c T^{1/2}(\hat{\sigma}-\sigma)\mu,$$

and insert

$$\begin{aligned} T^{1/2}(\hat{\beta}-\beta) &= \Sigma^{-1}T^{-1/2}\sum_{t=1}^{T} x_t\varepsilon_t + \mathrm{o}_{\mathsf{P}}(1), \\ T^{1/2}(\hat{\sigma}-\sigma) &= \frac{1}{2}T^{-1/2}\sum_{t=1}^{T}(\varepsilon_t^2/\sigma-\sigma) + \mathrm{o}_{\mathsf{P}}(1). \end{aligned}$$

This shows that $(1-\alpha)\Sigma T^{1/2}(\breve{\beta}_{LS}-\beta)$ has the same limit distribution as

$$T^{-1/2}\sum_{t=1}^{T}\{x_t(\varepsilon_t 1_{(\sigma\underline{c}\le\varepsilon_t\le\bar{c}\sigma)} + \xi_1^c\varepsilon_t) + \frac{\xi_2^c}{2}(\varepsilon_t^2/\sigma-\sigma)\mu\}, \tag{7.1}$$

where the summand is a martingale difference sequence. The Central Limit Theorem for martingales shows that this expression is asymptotically $\mathsf{N}_m\left(0,\sigma^2\Phi_\beta\right)$. To find Φ_β we calculate the sum of the conditional variances

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T} x_t x_t'\{\tau_2^c + (\xi_1^c)^2 + 2\xi_1^c\tau_c^2\} + \mu\mu'(\frac{\xi_2^c}{2})^2 T^{-1}\sum_{t=1}^{T}(\tau_4-1) \\ & + T^{-1}\sum_{t=1}^{T}(x_t\mu' + \mu x_t')\frac{\xi_2^c}{2}(\tau_3^c + \xi_1^c\tau_3) \\ \xrightarrow{\mathsf{P}} \;& \Sigma\{\tau_2^c + (\xi_1^c)^2 + 2\xi_1^2\tau_2^c\} + \mu\mu'\{(\frac{\xi_2^c}{2})^2(\tau_4-1) + \xi_2^c(\tau_3^c + \xi_1^c\tau_3)\}. \end{aligned}$$

Divide by $(1-\alpha)\Sigma$ from right and left to get the limiting variance for $T^{-1/2}(\breve{\beta}_{LS}-\beta)$.

For the estimator $\breve{\sigma}_{LS}$ the limiting distribution of $\tau_2^c T^{1/2}(\breve{\sigma}_{LS} - \sigma)$ is, in the same way, that of

$$T^{-1/2}\sum_{t=1}^{T}\{(\varepsilon_t^2 - \frac{\sigma^2\tau_2^c}{1-\alpha})1_{(\sigma\underline{c}\leq\varepsilon_t\leq\bar{c}\sigma)} + \sigma\zeta_2^c\mu'\Sigma^{-1}x_t\varepsilon_t + \sigma\frac{\zeta_3^c}{2}(\varepsilon_t^2/\sigma - \sigma)\}. \tag{7.2}$$

This variable is asymptotically normal with variance gives by $\sigma^4\Phi_\sigma$ where

$$\begin{aligned}\Phi_\sigma &= \tau_4^c - \frac{(\tau_2^c)^2}{1-\alpha} + (\zeta_2^c)^2\mu'\Sigma^{-1}\mu + \frac{(\zeta_3^c)^2}{4}(\tau_4 - 1)\\ &\quad + 2\zeta_2^c\mu'\Sigma^{-1}\mu\tau_3^c + 2\frac{\zeta_3^c}{2}(\tau_4^c - \frac{(\tau_2^c)^2}{1-\alpha}) + 2\zeta_2^c\frac{\zeta_3^c}{2}\mu'\Sigma^{-1}\mu\tau_3.\end{aligned}$$

Finally, the asymptotic covariance is of the expressions (7.1), (7.2) is $\sigma^3\Phi_c$ where

$$\begin{aligned}\Phi_c &= \mu(1+\xi_1^c)\tau_3^c + \mu\zeta_2^c(\tau_2^c + \xi_1^c) + \mu\frac{\zeta_3^c}{2}(\tau_3^c + \xi_1^c\tau_3)\\ &\quad + \mu\frac{\xi_2^c}{2}\{\tau_4^c - \frac{(\tau_2^c)^2}{1-\alpha}\} + \mu\mu'\Sigma^{-1}\mu\frac{\xi_2^c\zeta_2^c}{2}\tau_3 + \mu\frac{\xi_2^c\zeta_3^c}{4}(\tau_4 - 1).\end{aligned}$$

■

Proof of Theorem 4.3. We want to apply Theorem 3.1 to the contributions for the two subsets $\mathcal{I}_1$ and $\mathcal{I}_2$. The least squares estimator based on the full sample satisfies condition (i,a): $T^{1/2}(\hat{\sigma}_j - \sigma, \hat{\beta}_j - \beta) = \mathrm{O_P}(1)$, and the stationarity of x_t shows that conditions (ii,a,b,c) hold. Thus, define the product moments

$$S_{uv} = \sum_{t=1}^{T} u_t w_t' 1_{\{\underline{c}\leq v_t\leq\bar{c}\}} = \sum_{t\in\mathcal{I}_1} u_t w_t' 1_{\{\underline{c}\leq v_t\leq\bar{c}\}} + \sum_{t\in\mathcal{I}_2} u_t w_t' 1_{\{\underline{c}\leq v_t\leq\bar{c}\}} = S_{uw}^1 + S_{uw}^2.$$

The stationarity of x_t implies that (4.1) holds. Considering the term S_{xx} apply (3.4) of Theorem (3.1) to get

$$T_1^{-1}S_{xx}^j \xrightarrow{\mathsf{P}} \lambda_j(1-\alpha)\Sigma \qquad \text{so} \qquad T^{-1}S_{xx} \xrightarrow{\mathsf{P}} (1-\alpha)\Sigma, \tag{7.3}$$

since $T_j/T \to \lambda_j$.

Representation of $T^{1/2}(\tilde{\beta} - \beta)$: The estimator $\tilde{\beta}$ satisfies

$$\tilde{\beta} - \beta = S_{xx}^{-1}S_{x\varepsilon} = S_{xx}^{-1}\{S_{xx}^1(\tilde{\beta}^1 - \beta) + S_{xx}^2(\tilde{\beta}^2 - \beta)\},$$

where $\tilde{\beta}^j - \beta = (S_{xx}^j)^{-1}S_{x\varepsilon}^j$ is the contribution from $\mathcal{I}_j$. Due to (7.3) we then find

$$\tilde{\beta} - \beta = \lambda_1(\tilde{\beta}^1 - \beta) + \lambda_2(\tilde{\beta}^2 - \beta) + \mathrm{o_P}(1). \tag{7.4}$$

Turning to the individual contributions $\tilde{\beta}^j$, Theorem 3.1 shows

$$\{(1-\alpha)\Sigma\}(\tilde{\beta}^1 - \beta) = T_1^{-1}\sum_{t\in\mathcal{I}_1} x_t\varepsilon_t 1_{(\underline{c}\sigma\leq\varepsilon_t\leq\sigma\bar{c})} + \xi_1^c\Sigma(\hat{\beta}_2 - \beta) + \xi_2^c(\hat{\sigma}_2 - \sigma)\mu + \mathrm{o_P}(T_1^{-1/2}),$$

where $\hat{\beta}_2, \hat{\sigma}_2$ are the initial least squares estimators satisfying

$$\begin{aligned}\hat{\beta}_2 - \beta &= \Sigma^{-1} T_2^{-1} \sum_{t \in \mathcal{I}_2} x_t \varepsilon_t + \mathsf{o}_{\mathsf{P}}(1), \\ \hat{\sigma}_2 - \sigma &= \frac{1}{2} T_2^{-1} \sum_{t \in \mathcal{I}_2} (\varepsilon_t^2/\sigma - \sigma) + \mathsf{o}_{\mathsf{P}}(1).\end{aligned}$$

Inserting these in the expression for $\tilde{\beta}^1$ then gives

$$\begin{aligned}&\{(1-\alpha)\Sigma\}(\tilde{\beta}^1 - \beta) \\ = \; & T_1^{-1} \sum_{t \in \mathcal{I}_1} x_t \varepsilon_t 1_{(\underline{c}\sigma \le \varepsilon_t \le \sigma \bar{c})} + \xi_1^c T_2^{-1} \sum_{t \in \mathcal{I}_2} x_t \varepsilon_t + \frac{1}{2} \xi_2^c \mu T_2^{-1} \sum_{t \in \mathcal{I}_2} (\varepsilon_t^2/\sigma - \sigma) + \mathsf{o}_{\mathsf{P}}\left(T^{-1/2}\right).\end{aligned}$$

Interchanging the role of the indices $1, 2$ gives a similar expression for $\tilde{\beta}^2 - \beta$. Combining these expressions according to (7.4) then proves (4.2).

Representation of $T^{1/2}(\tilde{\sigma}^2 - \sigma^2)$: We use the expression (3.2) showing

$$\begin{aligned}\tilde{\sigma}^2 &= (1-\alpha)(\tau_2^c S_{11})^{-1}(S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}) \\ &= (1-\alpha)(\tau_2^c S_{11})^{-1}\{S_{\varepsilon\varepsilon} - (\tilde{\beta} - \beta)' S_{xx} (\tilde{\beta} - \beta)\}.\end{aligned}$$

Since (4.2) implies $\tilde{\beta} - \beta = \mathrm{O}_{\mathsf{P}}\left(T^{-1/2}\right)$ while $S_{xx} = \mathrm{O}_{\mathsf{P}}(T)$ by (7.3) then

$$\tilde{\sigma}^2 = (1-\alpha)(\tau_2^c S_{11})^{-1}\{S_{\varepsilon\varepsilon} + \mathrm{O}_{\mathsf{P}}(1)\},$$

so that, using $T^{-1} S_{11} \xrightarrow{\mathsf{P}} 1 - \alpha$,

$$\tau_2^c T^{1/2}(\tilde{\sigma}^2 - \sigma^2) = T^{-1/2}(S_{\varepsilon\varepsilon} - \sigma^2 \frac{\tau_2^c}{1-\alpha} S_{11}) + \mathsf{o}_{\mathsf{P}}(1).$$

We apply (3.7) and (3.8) and find that the contribution from $\mathcal{I}_1$ is

$$\begin{aligned}&T^{-1/2} \sum_{t \in \mathcal{I}_1} [(\varepsilon_t^2 - \sigma^2 \frac{\tau_2^c}{1-\alpha}) 1_{(\underline{c}\sigma \le \varepsilon_t \le \bar{c}\sigma)} \\ &+ \sigma \frac{\lambda_1}{\lambda_2} \{\zeta_2^c \mu' \Sigma^{-1} T^{-1/2} \sum_{t \in \mathcal{I}_2} x_t \varepsilon_t + \frac{\zeta_3^c}{2} T^{-1/2} \sum_{\mathcal{I}_2} (\varepsilon_t^2/\sigma - \sigma)\}] + \mathrm{O}_{\mathsf{P}}\left(T^{1/2}\right),\end{aligned}$$

which together with an expression for the contribution from $\mathcal{I}_2$ shows (4.3). ∎

Proof of Theorem 4.4. We apply (4.2) where the summands on the right hand side is a martingale difference sequence and we apply the Central Limit Theorem for martingales to shows that $T^{1/2}(\tilde{\beta} - \beta) \xrightarrow{\mathsf{D}} \mathsf{N}_m(0, \Phi)$. In order to find Φ we calculate the sum of the conditional variances and find

$$\begin{aligned}&T^{-1} \sum_{t=1}^{T} x_t x_t' \{\tau_c^2 + h_t^2 (\xi_1^c)^2 + 2 h_t \xi_1^c \tau_2^c\} + \mu\mu' (\frac{\xi_2^c}{2})^2 h_t^2 (\tau_4 - 1) \\ &+ T^{-1} \sum_{t=1}^{T} (x_t \mu' + \mu x_t') \frac{\xi_2^c}{2} (h_t \tau_3^c + h_t^2 \xi_1^c \tau_3).\end{aligned}$$

Using the relations

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} h_t &= 1 & T^{-1}\sum_{t=1}^{T} h_t^2 &= \frac{\lambda_1^2}{\lambda_2}+\frac{\lambda_2^2}{\lambda_1}, \\ T^{-1}\sum_{t=1}^{T} x_t h_t &\xrightarrow{\mathsf{P}} \mu & T^{-1}\sum_{t=1}^{T} x_t h_t^2 &\xrightarrow{\mathsf{P}} \mu(\frac{\lambda_1^2}{\lambda_2}+\frac{\lambda_2^2}{\lambda_1}), \\ T^{-1}\sum_{t=1}^{T} x_t x_t' h_t &\xrightarrow{\mathsf{P}} \Sigma, & T^{-1}\sum_{t=1}^{T} x_t x_t' h_t^2 &\xrightarrow{\mathsf{P}} \Sigma(\frac{\lambda_1^2}{\lambda_2}+\frac{\lambda_2^2}{\lambda_1}), \end{aligned} \tag{7.5}$$

we find the limit (4.4). ■

Proof of Theorem 4.5. We apply Theorem 3.1 and Corollary 3.2. The initial estimators $\tilde{\beta}$ and $\tilde{\sigma}^2$ satisfy condition (i,a), and the stationarity implies conditions (ii,a,b,c). We therefore get that $T^{-1}S_{xx} \xrightarrow{\mathsf{P}} (1-\alpha)\Sigma$, and $T^{-1}S_{x1} \xrightarrow{\mathsf{P}} (1-\alpha)\mu$. Finally we find from (3.9), that, when the density is symmetric so that $\xi_2^c = 0$,

$$(1-\alpha)\Sigma T^{1/2}(\breve{\beta}_{Sat}-\beta) = T^{-1/2}\{\sum_{t=1}^{T} x_t\varepsilon_t 1_{(\underline{c}\sigma\le\varepsilon_t\le\sigma\bar{c})} + \xi_1^c\Sigma T^{1/2}(\tilde{\beta}-\beta)\} + \mathsf{o}_{\mathsf{P}}(1),$$

Now insert the expression for $\tilde{\beta}$ in (4.2) with $\xi_2^c = 0$, which is

$$(1-\alpha)\Sigma T^{1/2}(\tilde{\beta}-\beta) = T^{-1/2}\sum_{t=1}^{T} x_t\{\varepsilon_t 1_{(\underline{c}\sigma\le\varepsilon_t\le\sigma\bar{c})} + \xi_1^c\varepsilon_t h_t\} + \mathsf{o}_{\mathsf{P}}(1)$$

and we find

$$(1-\alpha)\Sigma T^{1/2}(\breve{\beta}_{Sat}-\beta) = \{T^{-1/2}\sum_{t=1}^{T} x_t\{(1+\frac{\xi_1^c}{1-\alpha})\varepsilon_t 1_{(\underline{c}\sigma\le\varepsilon_t\le\sigma\bar{c})} + \varepsilon_t\frac{(\xi_1^c)^2}{1-\alpha}h_t\} + \mathsf{o}_{\mathsf{P}}(1).$$

Again the summands form a martingale difference sequence and the Central Limit Theorem for martingales shows that $T^{1/2}(\breve{\beta}-\beta) \xrightarrow{\mathsf{D}} N(0,\Phi)$. We calculate the sum of the conditional variances

$$\sigma^2 T^{-1}\sum_{t=1}^{T} x_t x_t'\{(1+\frac{\xi_1^c}{1-\alpha})^2\tau_2^c + h_t^2\frac{(\xi_1^c)^4}{(1-\alpha)^2} + 2(1+\frac{\xi_1^c}{1-\alpha})\frac{(\xi_1^c)^2}{1-\alpha}h_t\tau_2^c\},$$

which converges in probability towards

$$\frac{\sigma^2}{(1-\alpha)^2}\Sigma\{(1-\alpha+\xi_1^c)^2\tau_2^c + 2(1-\alpha+\xi_1^c)(\xi_1^c)^2\tau_2^c + (\frac{\lambda_1^2}{\lambda_2}+\frac{\lambda_2^2}{\lambda_1})(\xi_1^c)^4\},$$

which gives the expression (4.5) after dividing by $(1-\alpha)^2\Sigma^2$. ■

Proof of Theorem 5.1. The results (5.3), (5.4), (5.5): Note that it can be assumed without loss of generality that D has the Jordan form of Nielsen (2005, §4). Using the normalisation N_D suggested in that paper it follows that $T^{-1}N_D'\sum_{t=1}^{T} d_t d_t' N_D$ converges. The results then follow from Nielsen (2005, Theorem 4.1, 6.2, 6.4).

The result (5.6) follows from the Central Limit Theorem for martingales noting that the Lindeberg conditions hold:

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\mathsf{E}|(Y_{t-1}-\Psi d_{t-1})\varepsilon_t'|^2 1_{(|Y_{t-1}-\Psi d_{t-1})\varepsilon_t'|\geq T^{1/2}a)} &\to 0, \\ \sum_{t=1}^{T}\mathsf{E}|M_T d_t\varepsilon_t'|^2 1_{(|M_T d_{t-1}\varepsilon_t'|\geq T^{1/2}a)} &\leq c\frac{1}{T}\sum_{t=1}^{T}|M_T d_t|^4 \to 0. \end{aligned}$$

Finally, (5.7), (5.8) follow by combining (5.3) and (5.6). ■

Proof of Theorem 5.2. We can mimic the steps of the proof of Theorem 5.1 for the sums over the subsets $t \in \mathcal{I}_j$ rather than $t \in \mathcal{I}_1 \cup \mathcal{I}_2$. Thus, the assumptions of Theorem 3.1 are satisfied for each subset. In particular, it holds

$$(N_T^{-1})'(\hat{\beta}_j-\beta) = \Sigma_j^{-1}N_T\sum_{t\in\mathcal{I}_j}x_t\varepsilon_t + \mathsf{o_P}(1), \tag{7.6}$$

$$T^{1/2}(\hat{\sigma}_j-\sigma) = (2\sigma\lambda_j)^{-1}T^{-1/2}\sum_{t\in\mathcal{I}_j}\left(\varepsilon^2-\sigma^2\right) + \mathsf{o_P}(1). \tag{7.7}$$

We can now apply Theorem 3.1 to the estimator

$$(N_T^{-1})'(\tilde{\beta}-\beta) = (N_T S_{xx} N_T')^{-1}N_T S_{x\varepsilon}.$$

Apply (3.4), (3.5), (3.6) of Theorem 3.1 to each component to get

$$\begin{aligned} N_T S_{xx} N_T' &\xrightarrow{\mathsf{P}} (1-\alpha)(\Sigma_1+\Sigma_2) = (1-\alpha)\Sigma, \\ T^{-1/2}N_T S_{x1} &\xrightarrow{\mathsf{P}} (1-\alpha)(\mu_1+\mu_2) = (1-\alpha)\mu. \\ N_T S_{x\varepsilon} &= N_T\sum_{j=1}^{2}\sum_{t\notin\mathcal{I}_j}\{x_t\varepsilon_t 1_{(\underline{c}\sigma<\varepsilon_t<\bar{c}\sigma)} + \xi_1^c x_t x_t'(\hat{\beta}_j-\beta) + \xi_2^c x_t(\hat{\sigma}_j-\sigma)\} + \mathsf{o_P}(1). \end{aligned}$$

For $S_{x\varepsilon}$ insert the expressions (7.6), (7.7) for the estimators to get the expression

$$\sum_{t=1}^{T}N_T x_t\varepsilon_t 1_{(\underline{c}\sigma<\varepsilon_t<\bar{c}\sigma)} + \xi_1^c\sum_{t=1}^{T}H_t N_T x_t\varepsilon_t + \frac{\xi_2^c}{2\sigma}T^{-1/2}\sum_{t=1}^{T}K_t\left(\varepsilon^2-\sigma^2\right) + \mathsf{o_P}(1),$$

where

$$\begin{aligned} H_t &= \Sigma_2\Sigma_1^{-1}1_{(t\in\mathcal{I}_1)} + \Sigma_1\Sigma_2^{-1}1_{(t\in\mathcal{I}_2)}, \\ K_t &= \mu_2\lambda_1^{-1}1_{(t\in\mathcal{I}_1)} + \mu_1\lambda_2^{-1}1_{(t\in\mathcal{I}_2)}. \end{aligned}$$

This expression is a sum of a martingale difference sequence and we therefore apply the Central Limit Theorem for martingales. We calculate the sum of the conditional variances to be

$$\sum_{t=1}^{T} N_T x_t x_t' N_T' \sigma^2 \tau_2^c + \sigma^2 (\xi_1^c)^2 \sum_{t=1}^{T} H_t N_T x_t x_t' N_T' H_t' + (\frac{\xi_2^c}{2})^2 T^{-1} \sum_{t=1}^{T} K_t K_t (\tau_4 - 1)$$
$$+\sigma^2 \xi_1^c \tau_2^c \sum_{t=1}^{T} (N_T x_t x_t' N_T' H_t' + H_t N_T x_t x_t' N_T')$$
$$+\sigma^2 \tau_3 \frac{\xi_2^c}{2} \xi_1^c T^{-1/2} \sum_{t=1}^{T} (H_t N_T x_t K_t' + K_t x_t' N_T' H_t') + \sigma^2 \frac{\xi_2^c}{2} \tau_3^c T^{-1/2} \sum_{t=1}^{T} (N_T x_t K_t' + K_t x_t' N_T')$$

Now we apply the results that

$$\sum_{t=1}^{T} N_T x_t x_t' N_T' \xrightarrow{\mathsf{P}} \Sigma, \qquad T^{-1} \sum_{t=1}^{T} K_t K_t' \to \tfrac{\mu_2 \mu_2'}{\lambda_1} + \tfrac{\mu_1 \mu_1'}{\lambda_2}$$
$$\sum_{t=1}^{T} H_t N_T x_t x_t' N_T' H_t' \xrightarrow{\mathsf{P}} \Sigma_2 \Sigma_1^{-1} \Sigma_2 + \Sigma_1 \Sigma_2^{-1} \Sigma_1, \qquad T^{-1/2} \sum_{t=1}^{T} N_T x_t K_t' \xrightarrow{\mathsf{P}} \mu_1 \mu_2' + \mu_2 \mu_1'$$
$$\sum_{t=1}^{T} N_T x_t x_t' N_T' H_t' \xrightarrow{\mathsf{P}} \Sigma, \qquad T^{-1/2} \sum_{t=1}^{T} H_t N_T x_t K_t' \xrightarrow{\mathsf{P}} \Sigma_2 \Sigma_1^{-1} \mu_1 \mu_2' \lambda_1^{-1} + \Sigma_1 \Sigma_2^{-1} \mu_2 \mu_1' \lambda_2^{-1}$$

which gives the result. ■

8 References

Bickel, P. J. (1975). One step Huber estimates in the linear model. *Journal of the American Statistical Association* 70, 428–434.

Hendry, D.F. (1999). An econometric analysis of US food expenditure, 1931-1989. In Magnus, J.R. and Morgan, M.S. (eds) *Methodology and tacit knowledge: Two experiments in econometrics.* p.341-361. New York: Wiley.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Selecting a Regression Saturated by Indicators. Forthcoming *Statistical Computing*

Huber, P. (1981). *Robust Statistics*, New York: Wiley.

Koenker, R. (2005) *Quantile Regression,* Cambridge: Cambridge University Press.

Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edition. New York: Springer.

Nielsen, B. (2005). Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. *Econometric Theory* 21, 534-561.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Rousseeuw, P.J. (1984) Least median of squares regression. *Journal of the American Statistical Association* 79, 871-880.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust regression and outlier detection.* New Jersey, John Wiley and Sons.

Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimatyion mi the linear model. *Journal of the American Statistical Association* 75, 828–838.

Yohai V. and Maronna, R. (1976). Location estimators based on linear combinations of modified order statistics. *Communications in Statistics Theory and Methods* 5, 481–486.

Research Papers 2008

2007-41 Søren Johansen and Anders Rygh Swensen: Exact rational expectations, cointegration, and reduced rank regression

2007-42 Ole E. Barndorff-Nielsen, José Manuel Corcuera and Mark Podolskij: Power variation for Gaussian processes with stationary increments

2007-43 Jean Jacod, Yingying Li, Per A. Mykland, Mark Podolskij and Mathias Vetter: Microstructure Noise in the Continuous Case: The Pre-Averaging Approach - JLMPV-9

2007-44 Charles S. Bos, Siem Jan Koopman and Marius Ooms: Long memory modelling of inflation with stochastic variance and structural breaks

2007-45 James Davidson and Nigar Hashimzade: Representation and Weak Convergence of Stochastic Integrals with Fractional Integrator Processes

2008-01 John A. Carlson, Christian M. Dahl, and Carol L. Osler: Short-run Exchange-Rate Dynamics: Theory and Evidence

2008-02 Peter Reinhard Hansen: Reduced-Rank Regression: A Useful Determinant Identity

2008-03 Søren Johansen, Katarina Juselius, Roman Frydman and Michael Goldberg: Testing hypotheses in an I(2) model with applications to the persistent long swings in the Dmk/$ rate

2008-04 Olaf Posch: Explaining output volatility: The case of taxation

2008-05 Annastiina Silvennoinen and Timo Teräsvirta: Modelling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model

2008-06 Annastiina Silvennoinen and Timo Teräsvirta: Multivariate GARCH models. To appear in T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch, eds. Handbook of Financial Time Series. New York: Springer.

2008-07 Changli He, Annastiina Silvennoinen and Timo Teräsvirta: Parameterizing unconditional skewness in models for financial time series

2008-08 Cristina Amado and Timo Teräsvirta: Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure

2008-09 Søren Johansen and Bent Nielsen: An analysis of the indicator saturation estimator as a robust regression estimator