OXFORD

The Methodology and Practice of Econometrics

A Festschrift in Honour of David F. Hendry

Edited by Jennifer L. Castle
and Neil Shephard

The Methodology and Practice of Econometrics

This page intentionally left blank

The Methodology and Practice of Econometrics

A Festschrift in Honour of David F. Hendry

Edited by
Jennifer L. Castle
and
Neil Shephard

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street , Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2009

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging-in-Publication Data

The methodology and practice of econometrics : a festschrift in honour of
David F. Hendry / Jennifer Castle and Neil Shephard.
p. cm.
Includes bibliographical references.
ISBN 978–0–19–923719–7 (acid-free paper)
1. Econometrics. 2. Econometric models. I. Hendry, David F. II. Castle,
Jennifer, 1979– III. Shephard, Neil.
HB139.M474 2009
330.01'5195—dc22 2008055210

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
MPG Biddles Ltd., King's Lynn, Norfolk

ISBN 978–0–19–923719–7

1 3 5 7 9 10 8 6 4 2

Preface

This book collects a series of essays to celebrate the work of David Hendry: one of the most influential of all modern econometricians.

David's writing has covered many areas of modern econometrics, which brings together insights from economic theory, past empirical evidence, the power of modern computing, and rigorous statistical theory to try to build useful empirically appealing models. His work led to the blossoming of the use of error correction models in applied and theoretical work. The questions he asked about multivariate nonstationary time-series were the basis of Clive Granger's formalization of cointegration. His sustained research programme has led to a massive increase in the rigour with which many economists carry out applied work on economic time-series. He pioneered the development of strong econometric software (e.g. PcGive and PcGets) to allow applied researchers to use their time effectively and developed the general-to-specific approach to model selection. Throughout the period we have known him he has been the most intellectually generous colleague we have ever had. A brave searcher for truth and clarity, with broad and deep knowledge, he is, to us, what an academic should be.

The volume is a collection of original research in time-series econometrics, both theoretical and applied, and reflects David's interests in econometric methodology. The volume is broadly divided into five sections, including model selection, correlations, forecasting, methodology, and empirical applications, although the boundaries are certainly opaque. The first four chapters focus on issues of model selection, a topic that has received a revival of interest in the last decade, partly due to David's own writings on the subject. Johansen and Nielsen provide rigorous model selection theory by deriving an innovative estimator that is robust to outliers and structural breaks. Three further chapters consider more applied aspects of model selection, including Hoover, Demiralp, and Perez who use automatic causal search and selection search algorithms to identify structural VARs, White and Kennedy who develop methods for defining, identifying, and estimating causal effects, and Doornik who develops a new general-to-specific search algorithm, Autometrics.

The next two chapters focus on techniques for modelling financial data, with applications to equity and commodity prices. Engle proposes a new estimation method for Factor DCC models called the 'McGyver' method, and Trivedi and Zimmer consider the use of copula mixture models to test for dependence.

The third section focuses on economic forecasting, an area that David has been prolific in. Stock and Watson consider the performance of factor based macroeconomic forecasts under structural instability, Banerjee and Marcellino extend the dynamic factor model to incorporate error correction models, and Clements considers whether forecasters are consistent in making forecasts, enabling a test of forecaster rationality. The fourth section considers econometric methodology in the broad sense, commencing with Granger's appeal for pragmatic econometrics, embodying much of David's econometric philosophy. The question of how to undertake simulations in dynamic models is addressed by Abadir and Paruolo, both Dolado, Gonzalo, and Mayoral, and Davidson examine the order of integration of time-series data, either for fractionally integrated processes or stationary versus nonstationary processes, and finally Hendry, Lu, and Mizon, complete the section by considering model identification and the implications it has for model evaluation.

The final section of the volume consists of a range of empirical applications that implement much of the Hendry methodology. Beyer and Juselius consider the question of how to aggregate data, with an application to a small monetary model of the Euro-area, Bårdsen and Nymoen consider the US natural rate of unemployment, selecting between a Phillips curve and a wage equilibrium correction mechanism, and Ericsson and Kamin revisit a model of Argentine broad money demand using general-to-specific model selection algorithms. The contributions cover the full breadth of time-series econometrics but all with the overarching theme of congruent econometric modelling using the coherent and comprehensive methodology that David has pioneered.

The volume assimilates original scholarly work at the frontier of academic research, encapsulating the current thinking in modern day econometrics and reflecting the intellectual impact that David has had, and will continue to have, on the profession.

We are indebted to a great many referees for their hard work in ensuring a consistently high standard of essays. Furthermore, it is a great pleasure to acknowledge all who helped to organize the conference in honour of David, held in Oxford in August 2007. In particular, our thanks go to Maureen Baker, Ann Gibson, Carlos Santos, Bent Nielsen, Jurgen Doornik, and the staff at the Economics Department. The conference was generously supported by our sponsors including the Bank of England, ESRC, Journal of Applied

Econometrics, Oxford-Man Institute of Quantitative Finance, Oxford Economic Papers, Oxford University Press, Royal Economic Society, Timberlake Consultants, and Wiley-Blackwell Publishers, and we are delighted to record our gratitude to them. Finally, our thanks extend to David, whose interactions and discussions with all the authors in the volume have been the inspiration for the research in this collection.

Jennifer L. Castle
Neil Shephard

This page intentionally left blank

Contents

List of Contributors

Karim M. Abadir *Imperial College, London*

Anindya Banerjee *University of Birmingham*

Gunnar Bårdsen *The Norwegian University of Science and Technology*

Andreas Beyer *European Central Bank*

Michael P. Clements *University of Warwick*

James Davidson *University of Exeter*

Selva Damiralp *Kọc University*

Juan J. Dolado *Universidad Carlos III de Madrid*

Jurgen A. Doornik *University of Oxford*

Robert F. Engle *New York University*

Neil R. Ericsson *Board of Governors of the Federal Reserve System*

Jesus Gonzalo *Universidad Carlos III de Madrid*

Clive W. J. Granger *University of California, San Diego*

David F. Hendry *University of Oxford*

Kevin D. Hoover *Duke University*

Søren Johansen *University of Copenhagen, and CREATES, University of Aarhus*

Katarina Juselius *University of Copenhagen*

Steven B. Kamin *Board of Governors of the Federal Reserve System*

Pauline Kennedy *Bates White, LLC*

Maozu Lu *University of Southampton*

Massimiliano Marcellino *Bocconi University and European University Institute, Florence*

Laura Mayoral *Institute for Economic Analysis, Spain (CSIC)*

Grayham E. Mizon *University of Southampton*

Bent Nielsen *University of Oxford*

Ragnar Nymoen *University of Oslo*

Paolo Paruolo *University of Insubria*

Stephen J. Perez *California State University*

James H. Stock *Harvard University and the National Bureau of Economic Research*

Pravin K. Trivedi *Indiana University*

Mark W. Watson *Princeton University, and the National Bureau of Economic Research*

Halbert White *University of California, San Diego*

David H. Zimmer *Western Kentucky University*

1

An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator

*Søren Johansen and Bent Nielsen**

1.1 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 observations. 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 analysed 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 chapter 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

* 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 chapter.

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 1.1 is an asymptotic expansion of the least squares estimator for (β, σ^2) based upon those observations for which $\underline{c} \leq v_t \leq \overline{c}$.

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 behaviour of the estimators in the presence of outliers, but refer to further work which we intend to do in the future.

1.1.1 *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

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

supplemented with an estimator of the 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

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

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 of ε_t.

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 chapter. 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 nonsymmetric error distribution.

We apply the theory of empirical processes using tightness arguments similar to Bickel (1975). The representation in our main result Theorem 1.1 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 1.1 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

$$T^{1/2}(\breve{\mu} - \mu) = (1-\alpha)^{-1} \left\{ T^{-1/2} \sum_{t=1}^{T} \varepsilon_t 1_{(\underline{c}\sigma \leq \varepsilon_t \leq \sigma\overline{c})} + 2c\mathsf{f}(c) T^{1/2}(\hat{\mu} - \mu) \right\} + \mathrm{o_P}(1).$$

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

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

so that

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

$$\overset{\mathsf{D}}{\to} \mathsf{N}\left[0, \sigma^2 \frac{\tau_2^c}{\{1 - \alpha - 2c\mathsf{f}(c)\}^2}\right],$$

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.

1.1.2 *The Structure of the Chapter*

The one-step estimators are described in detail in section 1.2, and in section 1.3 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 section 1.4.1. The situation where the initial estimator is an indicator saturated estimator is then considered for stationary process in section 1.4.2 and for trend stationary autoregressive processes and unit root processes in section 1.5. Section 1.6 contains the proof of the main theorem, which involves techniques for empirical processes, whereas proofs for special cases are given in section 1.7. Finally, section 1.8 concludes.

1.2 The One-step *M*-estimators

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

1.2.1 *The Regression Model*

As a statistical model consider the regression model

$$y_t = \beta' x_t + \sum_{i=1}^{T} \gamma_i 1_{(i=t)} + \varepsilon_t \qquad t = 1, \ldots, T, \tag{1.3}$$

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 $(x_1, \ldots x_t, \varepsilon_1, \ldots, \varepsilon_{t-1})$ by E_{t-1} and Var_{t-1}.

The parameter space of the model is given by $\beta, (\gamma_1, \ldots, \gamma_T), \sigma^2 \in \mathbb{R}^m \times \mathbb{R}^T \times \mathbb{R}_+$. 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

$$\tau_0^c = \int_{\underline{c}}^{\overline{c}} \mathsf{f}(v)dv = 1 - \alpha \quad \text{and} \quad \tau_1^c = \int_{\underline{c}}^{\overline{c}} v\mathsf{f}(v)dv = 0. \tag{1.4}$$

It is convenient to introduce as a general notation

$$\tau_n = \int_{\mathbb{R}} u^n \mathsf{f}(u)du, \qquad \tau_n^c = \int_{\underline{c}}^{\overline{c}} u^n \mathsf{f}(u)du, \tag{1.5}$$

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

1.2.2 *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.

1.2.2.1 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 $\gamma_t = 0$ as

$$v_t = (y_t - \hat{\beta}' x_t)/\hat{\omega}_t, \tag{1.6}$$

where $\hat{\omega}_t^2$ could simply be chosen as $\hat{\sigma}^2$ or as $\hat{\sigma}^2\{1 - x_t'(\sum_{s=1}^T x_s x_s')^{-1} x_t\}$ 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

$$\breve{\beta}_{LS} = \left\{ \sum_{t=1}^{T} x_t x_t' 1_{(\underline{c} \le v_t \le \overline{c})} \right\}^{-1} \sum_{t=1}^{T} x_t y_t 1_{(\underline{c} \le v_t \le \overline{c})}, \tag{1.7}$$

$$\breve{\sigma}^2_{LS} = \left(\frac{\tau_2^c}{1-\alpha} \right)^{-1} \left\{ \sum_{t=1}^{T} 1_{(\underline{c} \le v_t \le \overline{c})} \right\}^{-1} \sum_{t=1}^{T} (y_t - \breve{\beta}_{LS}' x_t)^2 1_{(\underline{c} \le v_t \le \overline{c})}. \tag{1.8}$$

It will be shown that $\{\sum_{t=1}^{T} 1_{(\underline{c} \le v_t \le \overline{c})}\}^{-1} \sum_{t=1}^{T} (y_t - \breve{\beta}_{LS}' x_t)^2 1_{(\underline{c} \le v_t \le \overline{c})} \overset{\mathsf{P}}{\to} \sigma^2 \tau_2^c/(1-\alpha)$, 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 $\breve{\sigma}^2_{LS} = 0$.

1.2.2.2 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 $\mathcal{I}_j$

$$\hat{\beta}_j = \left(\sum_{t \in \mathcal{I}_j} x_t x_t' \right)^{-1} \sum_{t \in \mathcal{I}_j} x_t y_t, \qquad \hat{\sigma}_j^2 = \frac{1}{T_j} \sum_{t \in \mathcal{I}_j} (y_t - \hat{\beta}_j' x_t)^2, \tag{1.9}$$

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

$$v_t = 1_{(t \in \mathcal{I}_2)} (y_t - \hat{\beta}_1' x_t)/\hat{\omega}_{t,1} + 1_{(t \in \mathcal{I}_1)} (y_t - \hat{\beta}_2' x_t)/\hat{\omega}_{t,2}, \tag{1.10}$$

   where $\hat{\omega}^2_{t,j}$ could be chosen as $\hat{\sigma}_j^2$ or $\hat{\sigma}_j^2 \{1 + x_t' (\sum_{s \notin \mathcal{I}_j} x_s x_s')^{-1} x_t\}$ as for fixed regressors.
3. We then compute robustified least squares estimators $\tilde{\beta}$ and $\tilde{\sigma}^2$ by (1.7) and (1.8) based on v_t given by (1.10).
4. Based on the estimators $\tilde{\beta}$ and $\tilde{\sigma}^2$ define the t-ratios for testing $\gamma_t = 0$:

$$\tilde{v}_t = (y_t - \tilde{\beta}' x_t)/\tilde{\omega}_t, \tag{1.11}$$

   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 (1.7) and (1.8) based on $\tilde{v}_t$ given by (1.11).

1.3 The Main Asymptotic Result

Asymptotic distributions will be derived under the assumption that in (1.3) 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 (1.7) and (1.8) become

$$\breve{\beta} = S_{xx}^{-1} S_{xy}, \tag{1.12}$$

$$\breve{\sigma}^2 = (1-\alpha)(\tau_2^c S_{11})^{-1}(S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}). \tag{1.13}$$

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* $\max_{1 \leq t \leq T} T^{1/2} |\hat{\omega}_t^2 - \hat{\sigma}^2| = o_{\mathsf{P}}(1)$.

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

Theorem 1.1. *Consider model (1.3), where $\gamma_i = 0$ for all i, and there exists some estimators $(\hat{\beta}, \hat{\sigma}^2)$ and nonstochastic normalization matrices $N_T \to 0$, so that*

(*i*) *The initial estimators satisfy*
- (*a*) $T^{1/2}(\hat{\sigma}^2 - \sigma^2)$, $(N_T^{-1})'(\hat{\beta} - \beta) = O_{\mathsf{P}}(1)$,
- (*b*) $\hat{\omega}_t^2$ *satisfies Assumption B.*

(*ii*) *The regressors satisfy, jointly,*
- (*a*) $N_T \sum_{t=1}^{T} x_t x_t' N_T' \overset{\mathsf{D}}{\to} \Sigma \overset{a.s.}{>} 0$,
- (*b*) $T^{-1/2} N_T \sum_{t=1}^{T} x_t \overset{\mathsf{D}}{\to} \mu$,
- (*c*) $\max_{t \leq T} \mathsf{E}|T^{1/2} N_T x_t|^4 = O(1)$.

(*iii*) *The density* f *satisfies Assumption A, and* $\underline{c}$ *and* $\overline{c}$ *are chosen so that* $\tau_1^c = 0$.

Then it holds

$$T^{-1}S_{11} \overset{\mathsf{P}}{\to} 1-\alpha, \tag{1.14}$$

$$N_T S_{xx} N_T' \overset{\mathsf{D}}{\to} (1-\alpha)\Sigma, \tag{1.15}$$

$$T^{-1/2} N_T S_{x1} \overset{\mathsf{D}}{\to} (1-\alpha)\mu. \tag{1.16}$$

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*

$$N_T S_{x\varepsilon} = N_T \sum_{t=1}^{T} \left\{ x_t \varepsilon_t 1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)} + \xi_1^c x_t x_t'(\hat{\beta}-\beta) + \xi_2^c(\hat{\sigma}-\sigma)x_t \right\} + \mathrm{o}_{\mathsf{P}}(1), \tag{1.17}$$

$$S_{\varepsilon\varepsilon} = \sum_{t=1}^{T} \left\{ \varepsilon_t^2 1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)} + \sigma\xi_2^c(\hat{\beta}-\beta)'x_t + \sigma\xi_3^c(\hat{\sigma}-\sigma) \right\} + \mathrm{o}_{\mathsf{P}}\left(T^{1/2}\right), \tag{1.18}$$

$$S_{11} = \sum_{t=1}^{T} \left\{ 1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)} + \xi_0^c(\hat{\beta}-\beta)'x_t/\sigma + \xi_1^c(\hat{\sigma}/\sigma - 1) \right\} + \mathrm{o}_{\mathsf{P}}\left(T^{1/2}\right). \tag{1.19}$$

Combining the expressions for the product moments gives expressions for the one-step *M*-estimators of the form (1.12), (1.13). 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 1.2. *Suppose the assumptions of Theorem 1.1 are satisfied. Then*

$$(1-\alpha)\Sigma\left(N_T^{-1}\right)'(\breve{\beta}-\beta) = N_T \sum_{t=1}^{T} x_t \varepsilon_t 1_{(\underline{c}\sigma \le \varepsilon_t \le \sigma\overline{c})}$$
$$+ \xi_1^c \Sigma\left(N_T^{-1}\right)'(\hat{\beta}-\beta) + \xi_2^c T^{1/2}(\hat{\sigma}-\sigma)\mu + \mathrm{o}_{\mathsf{P}}(1), \tag{1.20}$$

$$\tau_2^c T^{1/2}(\breve{\sigma}^2 - \sigma^2) = T^{-1/2} \sum_{t=1}^{T} \left(\varepsilon_t^2 - \sigma^2 \frac{\tau_2^c}{1-\alpha} \right) 1_{(\underline{c}\sigma \le \varepsilon_t \le \sigma\overline{c})}$$
$$+ \sigma\zeta_2^c \mu'\left(N_T^{-1}\right)'(\hat{\beta}-\beta) + \sigma\zeta_3^c T^{1/2}(\hat{\sigma}-\sigma) + \mathrm{o}_{\mathsf{P}}(1), \tag{1.21}$$

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

$$\left\{ (N_T^{-1})'(\breve{\beta}-\beta), T^{1/2}(\breve{\sigma}^2 - \sigma^2) \right\} = \mathrm{O}_{\mathsf{P}}(1), \tag{1.22}$$

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

The proofs of Theorem 1.1 and Corollary 1.2 are given in section 1.6. It involves a series of steps. In section 1.6.1 a number of inequalities are given for the indicator functions appearing in S_{xx} and $S_{x\varepsilon}$, and in section 1.6.2 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 section 1.6.3.

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 $\varepsilon_t 1_{(\underline{c}\sigma\leq\varepsilon\leq\sigma\overline{c})}$ would not have zero mean and the conclusion (1.22) would in general fail because $N_T S_{x\varepsilon}$ would diverge.

1.4 Asymptotic Distributions in the Stationary Case

We now apply Theorem 1.1 and Corollary 1.2 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 1.1 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.

1.4.1 *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 $(\breve{\beta}_{LS}, \breve{\sigma}^2_{LS})$ be the robustified least squares estimators based on these, as given by (1.6), (1.12), and (1.13). We find the asymptotic distribution of these estimators with a proof in section 1.7.

Theorem 1.3. *Consider model (1.3) 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*

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

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

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

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 (1.22),* $N_T S_{xx} N_T/(1-\alpha)$*, see (1.15), and* $T^{-1/2} N_T S_{x1}/(1-\alpha)$*, see (1.16), respectively, so that, for instance*

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

The case where f is symmetric is of special interest. The critical value is then $c = -\underline{c} = \bar{c}$ and $\tau_3 = \tau^c_3 = 0$ and $\xi^c_0 = \xi^c_2 = 0$ so $\zeta^c_2 = 0$, whereas $\xi^c_1 = 2c\mathsf{f}(c)$ and $\xi^c_3 = 2c^3\mathsf{f}(c)$ so $\zeta^c_3 = \{c^2 - \tau^c_2/(1-\alpha)\}2c\mathsf{f}(c)$. It follows that $\kappa_\beta = \kappa_\sigma = \kappa_c = \eta_c = 0$. Theorem 1.3 then has the following Corollary.

Corollary 1.4. *If* f *is symmetric and the assumptions of Theorem 1.3 hold, then*

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

where, with $\xi^c_1 = 2c\mathsf{f}(c)$ *and* $\zeta^c_3 = \{c^2 - \tau^c_2/(1-\alpha)\}2c\mathsf{f}(c)$*, it holds*

$$(1-\alpha)^2 \eta_\beta = \tau^c_2 (1 + 2\xi^c_1) + (\xi^c_1)^2,$$

$$2(\tau^c_2)^2 \eta_\sigma = \{\tau^c_4 - \frac{(\tau^c_2)^2}{1-\alpha}\}(1 + \zeta^c_3) + \frac{(\zeta^c_3)^2}{4}(\tau_4 - 1).$$

Corollary 1.4 shows that the efficiency of the indicator saturated estimator $\breve{\beta}_{LS}$ with respect to the least squares estimator $\hat{\beta}$ is

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

Likewise the efficiency of $\breve{\sigma}_{LS}$ is efficiency$(\hat{\sigma}^2, \breve{\sigma}^2_{LS}) = \eta^{-1}_\sigma$. 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 1.1.

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 1.1. We first prove two representations corresponding to (1.20) and (1.21) 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 1.5. *Suppose* $\gamma_i = 0$ *for all i in model (1.3), 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*

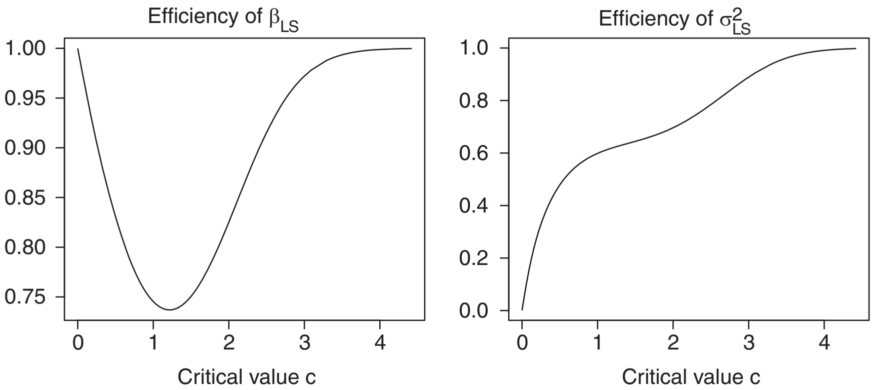

FIG. 1.1. The efficiency of the estimators $\breve{\beta}_{LS}$ and $\breve{\sigma}^2_{LS}$ with respect to the least squares estimators $\hat{\beta}$ and $\hat{\sigma}^2$, respectively, for f equal to the Gaussian density.

that $\tau_1^c = 0$. *Then, for* $j = 1,2$ *it holds, with* $\lambda_1 + \lambda_2 = 1$ *and* $\lambda_j > 0$, *that*

$$T^{-1}\sum_{t\in\mathcal{I}_j} x_t \overset{\mathsf{P}}{\to} \lambda_j \mu, \qquad T^{-1}\sum_{t\in\mathcal{I}_j} x_t x_t' \overset{\mathsf{P}}{\to} \lambda_j \Sigma. \tag{1.23}$$

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*

$$(1-\alpha)\Sigma T^{1/2}(\tilde{\beta}-\beta) = T^{-1/2}\sum_{t=1}^{T}\left[x_t\left\{\varepsilon_t 1_{(\underline{c}\sigma\le\varepsilon_t\le\sigma\bar{c})} + h_t\xi_1^c\varepsilon_t\right\} + \frac{\xi_2^c}{2}\mu h_t\left(\varepsilon_t^2/\sigma - \sigma\right)\right] + \mathrm{o}_{\mathsf{P}}(1), \tag{1.24}$$

$$\tau_2^c T^{1/2}(\tilde{\sigma}^2 - \sigma^2) = T^{-1/2}\sum_{t=1}^{T}\left\{\left(\varepsilon_t^2 - \sigma^2\frac{\tau_2^c}{1-\alpha}\right)1_{(\underline{c}\sigma\le\varepsilon_t\le\sigma\bar{c})} + \sigma\zeta_2^c\mu'\Sigma^{-1}x_t\varepsilon_t h_t + \sigma\frac{\zeta_3^c}{2}\left(\varepsilon_t^2/\sigma - \sigma\right)h_t\right\} + \mathrm{o}_{\mathsf{P}}(1). \tag{1.25}$$

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 1.6. *Suppose* $\gamma_i = 0$ *for all* i *in model (1.3), 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*

$$T^{1/2}(\tilde{\beta}-\beta) \overset{\mathsf{D}}{\to} \mathsf{N}_m\left\{0, \sigma^2(\eta\Sigma^{-1} + \kappa\Sigma^{-1}\mu\mu'\Sigma^{-1})\right\}, \tag{1.26}$$

where

$$(1-\alpha)^2\eta = \tau_2^c\left(1+2\xi_1^c\right) + \left(\xi_1^c\right)^2\left(\frac{\lambda_2^2}{\lambda_1}+\frac{\lambda_1^2}{\lambda_2}\right),$$

$$(1-\alpha)^2\kappa = \xi_2^c\left[\left\{\frac{1}{4}\xi_2^c(\tau_4-1)+\xi_1^c\tau_3\right\}\left(\frac{\lambda_2^2}{\lambda_1}+\frac{\lambda_1^2}{\lambda_2}\right)+\tau_3^c\right].$$

We note that the result of Hendry, Johansen, and Santos (2008) is a special case of Theorem 1.6. 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 1.7. *Suppose $\gamma_t = 0$, $t = 1, \dots, T$ in model (1.3), and that x_t is stationary with mean μ, variance Σ, and finite fourth moment, and that $\hat{\omega}_{t,j}^2$ and $\tilde{\omega}_t^2$ satisfy Assumption B. The symmetric density* f *satisfies Assumption A, and $\underline{c}$ and $\bar{c}$ are chosen so that $\tau_1^c = 0$. Then*

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

where

$$(1-\alpha)^4\eta_{Sat} = \left(1-\alpha+\xi_1^c\right)\tau_2^c\left\{\left(1-\alpha+\xi_1^c\right)+2\left(\xi_1^c\right)^2\right\}+\left(\xi_1^c\right)^4\left(\frac{\lambda_1^2}{\lambda_2}+\frac{\lambda_2^2}{\lambda_1}\right). \quad (1.27)$$

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

1.5 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.

1.5.1 *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

$$\begin{aligned} y_t &= \sum_{i=1}^{k} \delta_i y_{t-i} + \varphi d_{t-1} + \varepsilon_t, \\ d_t &= D d_{t-1}, \end{aligned} \tag{1.28}$$

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 (1.28) is of the form (1.3) with $x_t' = (y_{t-1}, \ldots, y_{t-k}, d_t')$ and $\beta' = (\delta_1, \ldots, \delta_k, \phi)$, 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 formalized 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

$$Y_{t-1} = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-k} \end{pmatrix}, \quad A = \left\{ \begin{matrix} (\delta_1, \ldots, \delta_{k-1}) & \delta_k \\ I_{k-1} & 0 \end{matrix} \right\}, \quad \Phi = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}, \quad e_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix},$$

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, section 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

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

so that $M_T \sum_{t=1}^{T} d_{t-1} d'_{t-1} M'_T = I_\ell$. Equivalently, a block diagonal normalization, N_D, could be chosen if D, without loss of generality, were assumed to have a Jordan structure as in Nielsen (2005, section 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

$$N_T = \begin{pmatrix} T^{-1/2} & 0 \\ 0 & M_T \end{pmatrix} \begin{pmatrix} I_k & -\Psi \\ 0 & I_\ell \end{pmatrix}. \tag{1.29}$$

Theorem 1.8. *Let y_t be the trend stationary process given by (1.28) 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*

$$N_T \sum_{t=1}^{T} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix}' N'_T \overset{\mathsf{P}}{\to} \Sigma \overset{\text{def}}{=} \begin{pmatrix} \Sigma_Y & 0 \\ 0 & \Sigma_D \end{pmatrix}, \tag{1.30}$$

$$T^{-1/2} N_T \sum_{t=1}^{T} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix} \overset{\mathsf{P}}{\to} \mu \overset{\text{def}}{=} \begin{pmatrix} 0 \\ \mu_D \end{pmatrix}, \tag{1.31}$$

$$\max_{1\le t\le T} |M_T d_t| = \mathrm{O}(T^{-1/2}), \tag{1.32}$$

$$N_T \sum_{t=1}^{T} \begin{pmatrix} Y_{t-1} \\ d_{t-1} \end{pmatrix} \varepsilon'_t \overset{\mathsf{D}}{\to} N_m(0, \sigma^2 \Sigma). \tag{1.33}$$

In particular, it holds

$$(N_T^{-1})'(\hat{\beta} - \beta) \overset{\mathsf{D}}{\to} \mathsf{N}_m(0, \sigma^2 \Sigma^{-1}), \tag{1.34}$$

$$T^{1/2}(\hat{\sigma}^2 - \sigma^2) = T^{-1/2} \sum_{t=1}^{T} (\varepsilon_t^2 - \sigma^2) + \mathrm{o}_{\mathsf{P}}(1) = \mathrm{O}_{\mathsf{P}}(1). \tag{1.35}$$

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.

1.5.2 *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 1.1 can then be applied to each of these components.

Theorem 1.9. *Let y_t be the trend stationary process given by (1.28) so $|\mathrm{eigen}(A)| < 1$, with finite fourth moment, deterministic component satisfying Assumption C, and $\hat{\omega}^2_{t,j}$ satisfies Assumption B. Suppose the density* f *satisfies Assumption A, and the truncation points are chosen so that $\tau^c_1 = 0$. Finally, assume that*

$$\lim_{T\to\infty} M_T \sum_{t\in\mathcal{I}_j} d_t d_t' M_T = \Sigma_{D,j} > 0, \tag{1.36}$$

$$\lim_{T\to\infty} T^{-1/2} M_T \sum_{t\in\mathcal{I}_j} d_t = \mu_{D,j}, \tag{1.37}$$

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

$$(N_T')^{-1}(\tilde{\beta} - \beta) \overset{\mathsf{D}}{\to} \mathsf{N}_m(0, \sigma^2 \Sigma^{-1} \Phi \Sigma^{-1}), \tag{1.38}$$

where $\Sigma = \Sigma_1 + \Sigma_2$ and

$$\begin{aligned}(1-\alpha)^2\Phi &= \tau^c_2\left(1 + 2\xi^c_1\right)\Sigma + \left(\xi^c_1\right)^2\left(\Sigma_2\Sigma_1^{-1}\Sigma_2 + \Sigma_1\Sigma_2^{-1}\Sigma_1\right) \\ &\quad + \tau^c_3 \frac{\xi^c_2}{2}(\mu_2\mu_1' + \mu_1\mu_2')\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) + (\tau_4 - 1)\left(\frac{\xi^c_2}{2}\right)^2\left(\frac{\mu_2\mu_2'}{\lambda_1} + \frac{\mu_1\mu_1'}{\lambda_2}\right) \\ &\quad + \tau_3 \frac{\xi^c_1\xi^c_2}{2}\left(\frac{\mu_2\mu_1'\Sigma_1^{-1}\Sigma_2 + \Sigma_2\Sigma_1^{-1}\mu_1\mu_2'}{\lambda_1} + \frac{\mu_1\mu_2'\Sigma_2^{-1}\Sigma_1 + \Sigma_1\Sigma_2^{-1}\mu_2\mu_1'}{\lambda_2}\right).\end{aligned}$$

A closer look at the expression for Φ shows that it is block diagonal. The variance for the autoregressive components is $(1-\alpha)^2\Phi_Y = \Sigma_Y\{\tau^c_2(1 + 2\xi^c_1) + (\xi^c_1)^2(\lambda_2^2\lambda_1^{-1} + \lambda_1^2\lambda_2^{-1})\}$. 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 $\xi^c_2 = 0$ 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 section 1.5.3 it is shown how this proportionality can be achieved by choosing the index sets in a particular way.

Corollary 1.10. *If* f *is symmetric then $\xi^c_2 = 0$ so*

$$(1-\alpha)^2\Phi = \tau^c_2(1 + 2\xi^c_1)\Sigma + (\xi^c_1)^2(\Sigma_2\Sigma_1^{-1}\Sigma_2 + \Sigma_1\Sigma_2^{-1}\Sigma_1).$$

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 1.3.

1.5.3 *Choice of Index Sets in the Nonstationary Case*

Corollary 1.10 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 $0 \leq u \leq v \leq 1$

$$\frac{1}{T} \sum_{t \in \mathrm{int}(Tu)+1}^{\mathrm{int}(Tv)} 1_{(t \in \mathcal{I}_j)} \to \lambda_j (v - u), \tag{1.39}$$

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.

1.5.3.1 RANDOM INDEX SETS

We will consider one particular index set which is alternating in a random way. Generate a series of independent Bernoulli variables, $s_1, \ldots, s_T$ taking the values 1 and 2 so that

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

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

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

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 (1.28). Since the index sets are constructed by independent sampling then

$$\mathsf{E}\left(N_T \sum_{t \in \mathcal{I}_j} x_t x_t' N_T'\right) = \mathsf{E}\left\{N_T \sum_{t=1}^{T} (x_t x_t') N_T'\right\} \mathsf{E}1_{(s_t = j)} = \mathsf{E}\left\{N_T \sum_{t=1}^{T} x_t x_t' N_T'\right\} \lambda_j \to \lambda_j \Sigma,$$

$$\mathsf{E}\left(T^{-1/2} N_T \sum_{t \in \mathcal{I}_j} x_t\right) = \mathsf{E}\left(T^{-1/2} N_T \sum_{t=1}^{T} x_t\right) \mathsf{E}1_{(s_t = j)} = \mathsf{E}\left(T^{-1/2} N_T \sum_{t=1}^{T} x_t\right) \lambda_j \to \lambda_j \mu.$$

1.5.3.2 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 (1.39) with $\lambda_1 = \lambda_2 = 1/2$.

Consider the trend stationary model in (1.28) 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

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

$$T^{-1} Q_T \sum_{t \in \mathcal{I}_j} d_t = T^{-1} Q_T \sum_{t=0}^{T/2-1} d_{2t+j} \to \frac{1}{2} \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}.$$

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

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

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.

1.5.4 *A Few Results for Unit Root Processes*

Consider the first order autoregression

$$X_t = \beta X_{t-1} + \varepsilon_t, \tag{1.40}$$

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

$$T^{-1/2}\sum_{t=1}^{\mathrm{int}(Tu)}\begin{Bmatrix}\varepsilon_t 1_{(t\in\mathcal{I}_1)}\\ \varepsilon_t 1_{(t\in\mathcal{I}_2)}\\ \varepsilon_t 1_{(|\varepsilon_t|<c\sigma)}\end{Bmatrix}\overset{\mathsf{D}}{\to}\begin{pmatrix}w_{1u}\\ w_{2u}\\ w_u^c\end{pmatrix}=W_u,$$

where W_u is a Brownian motion with variance matrix

$$\widetilde{\Omega}\overset{\mathrm{def}}{=}\sigma^2\begin{pmatrix}\lambda_1 & 0 & \lambda_1\tau_2^c\\ 0 & \lambda_2 & \lambda_2\tau_2^c\\ \lambda_1\tau_2^c & \lambda_2\tau_2^c & \tau_2^c\end{pmatrix}.$$

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 $\lambda_1\lambda_2\mathsf{E}(\sum_{t=1}^{T}X_{t-1}^4)$; it is therefore of order $T^{3/2}$. It follows that

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

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 $\int_0^1 w_u^2du$. It follows from Corollary 1.2, that the first round indicator saturated estimator satisfies

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

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

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

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 nonstationary case.

For the case of alternating index sets, nearly the same arguments apply as with random index sets. In this case the definition of the Brownian motions becomes

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

1.6 Proof of Main Result

The results of Theorem 1.1 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 1.1 this replacement is justified using techniques for empirical processes and in particular Koul (2002, Theorem 7.2.1, p. 298).

We define the normalized 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 $T^{1/2}(\hat{a}_T, \hat{b}_T) = \mathrm{O_P}(1)$ and $T^{1/2} \max_{1 \le t \le T} |\hat{a}_{Tt} - \hat{a}_T| = T^{1/2} \max_{1 \le t \le T} |\hat{\omega}_t - \hat{\sigma}| = \mathrm{o_P}(1)$ by assumption (*i*) of Theorem 1.1. Note that

$$\hat{\omega}_t v_t = \varepsilon_t - (\hat{\beta} - \beta)' x_t = \varepsilon_t - \{T^{-1/2}(N_T^{-1})'(\hat{\beta} - \beta)\}'(T^{1/2} N_T x_t) = \varepsilon_t - \hat{b}_T' x_{Tt}, \quad (1.41)$$

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

$$I_t(u) = I_t(a, b) = 1_{\{\underline{c}(\sigma + a) \le \varepsilon_t - b' x_{Tt} \le \overline{c}(\sigma + a)\}} - 1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)}, \quad (1.42)$$

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

$$N_T S_{xx} N_T' = T^{-1} \sum_{t=1}^{T} x_{Tt} x_{Tt}' 1_{(\underline{c} \le v_t \le \overline{c})} = T^{-1} \sum_{t=1}^{T} x_{Tt} x_{Tt}' 1_{(\underline{c}\sigma \le \varepsilon_t \le \overline{c}\sigma)} \quad (1.43)$$
$$+ T^{-1} \sum_{t=1}^{T} x_{Tt} x_{Tt}' \{I_t(\hat{a}_{Tt}, \hat{b}_T) - I_t(\hat{a}_T, \hat{b}_T)\} + T^{-1} \sum_{t=1}^{T} x_{Tt} x_{Tt}' I_t(\hat{a}_T, \hat{b}_T).$$

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 $(\hat{a}_T, \hat{b}_T) = \mathrm{O_P}(T^{1/2})$, 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 $N_T S_{x\varepsilon}$

$$N_T S_{x\varepsilon} = T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t 1_{(\underline{c}\leq \upsilon_t\leq \overline{c})} = T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t 1_{(\underline{c}\sigma\leq \varepsilon_t\leq \overline{c}\sigma)}$$
$$+ T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t\{I_t(\hat{a}_{tT},\hat{b}_T) - I_t(\hat{a}_T,\hat{b}_T)\} + T^{-1/2}\sum_{t=1}^{T} x_{Tt}\varepsilon_t I_t(\hat{a}_T,\hat{b}_T).$$

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

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

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 $\ell,m = 0,1,2$,

$$M_T^{\ell,m} = T^{-1/2}\sum_{t=1}^{T} g_m(x_{Tt})\,\varepsilon_t^{\ell}\{I_t(\hat{a}_{Tt},\hat{b}_T) - I_t(\hat{a}_T,\hat{b}_T)\}, \tag{1.44}$$

$$W_T^{\ell,m}(a,b) = \frac{1}{T}\sum_{t=1}^{T} g_m(x_{Tt})\,\varepsilon_t^{\ell} I_t(a,b), \tag{1.45}$$

$$V_T^{\ell,m}(a,b) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_m(x_{Tt})\left\{\varepsilon_t^{\ell} I_t(a,b) - \sigma^{\ell-1}\left(a\xi_{\ell+1}^c + b' x_{Tt}\xi_{\ell}^c\right)\right\}, \tag{1.46}$$

where the function g_m is given as

$$g_0(x_{Tt}) = 1, \qquad g_1(x_{Tt}) = x_{Tt}, \qquad g_2(x_{Tt}) = x_{Tt}x'_{Tt}. \tag{1.47}$$

Lemma 1.14 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 1.15, 1.16, and 1.17 below show that as $T\to\infty$ and if $T^{1/2}(\hat{a}_T,\hat{b}_T)$ is tight, then

$$M_T^{\ell,m} \overset{\mathsf{P}}{\to} 0, \qquad W_T^{\ell,m}(\hat{a}_T,\hat{b}_T) \overset{\mathsf{P}}{\to} 0 \qquad \text{and} \qquad V_T^{\ell,m}(\hat{a}_T,\hat{b}_T) \overset{\mathsf{P}}{\to} 0. \tag{1.48}$$

Some equalities and expansions are established initially in section 1.6.1. The remainder terms are analysed in section 1.6.2. Finally, the threads are pulled together in a proof of Theorem 1.1 in section 1.6.3.

1.6.1 *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 1.11. *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 1.11. 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 (1.42). 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 1.12. *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 1.12. 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 1.11 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 $(x_s,\varepsilon_s,s \leq t-1,x_t)$.