Asymptotic Properties of the Gauge of Step-Indicator Saturation

Bent Nielsen[1]
Nuffield College, University of Oxford

Matthias Qian
Department of Economics, University of Oxford

April 9, 2018

Abstract

We investigate the asymptotic properties of Step-indicator Saturation which is an algorithm to handle unmodelled location shifts in time series. We consider a stylized version of the algorithm that uses the split-half approach. We present asymptotic convergence and distribution results on the gauge of the algorithm which is the frequency of falsely retained step-indicators when the data generating process has no shifts. The proofs rely on empirical process results of temporal differences of residuals. Our results offer an asymptotic justification to use the gauge in choosing the tuning parameter of this statistical procedure.

1 Introduction

Step-indicator saturation detects location shifts which are changes to the previously unconditional mean of time-series data. As a member of the class of saturation algorithms, the idea of Step-indicator Saturation is simple: Extend the feature space, which is the space spanned by the regressors, with step-indicators that capture location shifts in the periods of interest. Then proceed to eliminate unnecessary step-indicators from the generalised feature space. The elimination procedure hinges on a cut-off value c. We present results in this paper to choose the appropriate cut-off value c for selection.

As location shifts we consider non-zero coefficients δ_ℓ in the data generating process,

$$y_i = \sum_{\ell=1}^{L} \delta_\ell 1_{(\tau_{\ell-1} \le i \le \tau_\ell)} + \beta' x_i + \varepsilon_i, \quad i = 1, ..., n$$

where x_i contains an intercept and ε_i have common reference density f and are independent from $\mathcal{F}_{i-1} = \sigma(x_1, ..., x_i, \varepsilon_1, ..., \varepsilon_{i-1})$. In fact, f can be very general and does not need to be symmetric. The number of location shifts L in the data generating process is typically unknown to the researcher. The breakpoints τ_ℓ determine the timing of the location shift.

Macroeconomic data are packed with location shifts. Bai and Perron (2003) find empirical evidence for location shifts in the US ex-post real interest rate (the three-month treasury bill rate deflated by the CPI inflation rate taken from the Citibase data) in 1966, 1972 and 1980. Hansen (2001) finds empirical evidence of location shifts in US labour productivity in 1982 and 1995. Perron (1989) finds that most US macroeconomic variables can be modelled as trend stationary with two location shifts in 1929 and 1973 which represents the financial crash of the Great Depression and an oil price shock following the OPEC oil embargo.

Unmodelled location shifts give rise to three distinct statistical problems. First, model selection procedures are affected that start from a general unrestricted model and target a parsimonious and encompassing empirical model. Castle and Hendry (2014) document the effect of unmodelled location shifts on selection probabilities of the variables, their lags and non-linear functions. Second, parameter estimation is distorted. Hendry and Mizon (2011) show that an unmodelled location shift caused by the food stamp program of the Great Depression leads to implausible regression coefficients in their model for the US food expenditure. Third, as

[1]B. Nielsen is supported by the Institute for Economic Modeling at the Oxford Martin School.

described in Clements and Hendry (1998), unanticipated location shifts at or near the forecast origin lead to forecast failure.

A number of variations of Step-indicator Saturation algorithms have been proposed, see Castle et al. (2015). Empirical researchers use the Autometrics implementation of Step-indicator Saturation as discussed in Doornik (2009) or the R package *gets* discussed in Pretis, Reade and Sucarrat (2016). In this paper we consider a stylised split-half version of the algorithm, which we present in section 2.

For split-half Step-indicator Saturation, the decision rule to categorise period i as a breakpoints hinges on some preliminary estimator $(\hat{\beta}, \hat{\sigma}^2)$ and the test-statistic $v_i = (\nabla y_i - \hat{\beta}' \nabla x_i)/\sqrt{2}\hat{\sigma}$, where $\nabla y_i = y_i - y_{i+1}$. A period i is classified as a breakpoint if $|v_i| \geq c$, where c is a suitable cut-off value. The cut-off value c is the only tuning parameter of the algorithm.

In order to choose the cut-off value c, we use the concept of the gauge, which considers the frequency of wrongly detected shifts under the null hypothesis of no shifts in the data generating process. The (empirical) gauge is formally defined as

$$\hat{\gamma}_n = n^{-1} \sum_{i=1}^{n} r_i,$$

where $r_i = 1$ denotes the existence of a breakpoint in period i and $r_i = 0$ the opposite case. The empirical gauge is similar, but different to the concept of the size used in hypothesis testing, as it is contaminated by estimation error within the decision rule for breakpoints.

The gauge was first used in the simulation studies of Hoover and Perez (1999). It was formally defined and named by Hendry and Santos (2010) and Castle et al. (2011). The gauge of a number of outlier detection algorithms was studied in Johansen and Nielsen (2016b).

1.1 Purpose of paper and results

The purpose of this paper is to provide an asymptotic theory for setting the cut-off value c which should be chosen to control the gauge. We find that under the null of no location shift in the data generating process, the empirical gauge converges in probability to

$$\hat{\gamma}_n(c) \overset{p}{\to} \mathsf{P}(|\nabla \varepsilon_i| \geq \sqrt{2}\sigma c) = \gamma(c), \tag{1}$$

the size of the underlying test.

We also derive the asymptotic distribution of the empirical gauge of Step-indicator Saturation. Its asymptotic expansion is

$$\begin{aligned} n^{1/2}\{\hat{\gamma}_n(c) - \gamma(c)\} &= -n^{-1/2}(\gamma(c) - 1_{(|\nabla \varepsilon_i| \leq \sqrt{2}\sigma c)} - 1) \\ &+ \frac{c\mathsf{h}(c)}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{\varepsilon_i^2}{\sigma^2} - 1 \right) h_i + \xi_c \Sigma^{-1} \frac{\mathsf{h}(c)}{n^{1/2}} \sum_{i=1}^{n} x_i \varepsilon_i h_i + \mathrm{o}_{\mathsf{P}}(1). \end{aligned}$$

which consists of the density h of $\nabla \varepsilon_i/(\sqrt{2}\sigma)$, a function h_i that depends on the samples sizes of the subsamples we form in the split-half implementation of Step-indicator Saturation and the function ξ_c which depends on average of the conditional expectations of ∇x_i, conditional on $\nabla \varepsilon_i = \sqrt{2}\sigma c$ and past information.

The asymptotic results are based on a number of empirical process results of temporal differences of residuals. While there is a growing literature on empirical processes for residuals, see Koul (2002), Engler and Nielsen (2009) Johansen and Nielsen (2009, 2016a, 2016b) and Jiao and Nielsen (2016), to the best of our knowledge, the class of empirical processes of temporal differences of residuals has not yet been studied.

Our paper contributes to the understanding of the asymptotic properties of automatic model selection techniques using the general-to-specific framework (see Doornik, 2009 and Hendry and

Krolzig, 2005). We are the first paper that considers asymptotic properties of this class of algorithms where we select over dependent regressors.

In section 2, we outline the model and the Step-indicator Saturation algorithm. In section 3,, we present the asymptotic results on the empirical gauge of Step-indicator Saturation and provide simulation results. In section 4, we state our results on the empirical processes of temporal differences of residuals. In section 5, we highlight two further useful lemmas.

2 Model and Step-indicator Saturation

2.1 Model

The data generating process in this paper is a linear time-series regression model

$$y_i = \beta' x_i + \varepsilon_i \quad i = 1, ..., n+1 \tag{2}$$

where y_i is a one-dimensional dependent variable, x_i is a k-dimensional vector of contemporaneous regressors which contains the intercept and β is the k-dimensional vector of regression coefficients. The error terms ε_i are independent of the filtration $\mathcal{F}_{i-1} = \sigma(x_1, \ldots, x_i, \varepsilon_1, \ldots, \varepsilon_{i-1})$ and identically distributed so that ε_i/σ has density f. In applications, we often assume f to be the normal density. We allow the regressors to be from a wide range of different classes, namely stationary regressors, trend stationary regressors and unit-root regressors. We do not allow for explosive regressors. Our sample size extents to $n+1$ to simplify the expression of the gauge.

The scaled forward differences $\nabla\varepsilon_i/(\sqrt{2}\sigma)$ have density

$$\mathsf{h}(x) = \sqrt{2}\int_{-\infty}^{\infty} \mathsf{f}(y)\mathsf{f}(x+y)\,\mathrm{d}y. \tag{3}$$

We summarize three properties of the density h in the following Lemma.

Lemma 2.1. *The density* h *satisfies the following properties.*
(*a*) *Symmetry:* $\mathsf{h}(x) = \mathsf{h}(-x)$*;*
(*b*) *Suppose* f *is continuous with second moments. Then* $\mathsf{f} = \mathsf{h}$ *if and only if* f *is normal;*
(*c*) *If* $\sup_{v\in\mathbb{R}} |\mathsf{f}(x)| < \infty$*, then* $\sup_{v\in\mathbb{R}} |\mathsf{h}(x)| < \infty$*;*
(*d*) *If* $\sup_{v\in\mathbb{R}} |\dot{\mathsf{f}}(v)| < \infty$*, then* $\sup_{v\in\mathbb{R}} |\dot{\mathsf{h}}(v)| < \infty$*;*
(*e*) *If* $\sup_{v\in\mathbb{R}} |v|\mathsf{f}(v) < \infty$*, then* $\sup_{v\in\mathbb{R}} |v|\mathsf{h}(v) < \infty$*.*

Define $\mathsf{H}(x) = \mathsf{P}(\nabla\varepsilon_i \leq \sqrt{2}\sigma c)$ as the corresponding distribution function. The absolute value of the scaled forward differences then has distribution function

$$\psi = 1 - \gamma = \mathsf{G}(c) = \mathsf{P}\left(|\varepsilon_i - \varepsilon_{i+1}| < \sqrt{2}\sigma c\right) = \int_{-c}^{c} \mathsf{h}(x)dx. \tag{4}$$

The conditional joint density $\mathsf{m}_i(y, x)$ given $\mathcal{F}_{i-1}$ is the density of the two random variables

$$\frac{\nabla\varepsilon_i}{\sqrt{2}\sigma} \quad \text{and} \quad \frac{n^{1/2}N'\nabla x_i}{\sqrt{2}\sigma}.$$

Let $\mathsf{k}_i(x)$ be the marginal density of the regressors x_i. Then the conditional density $\mathsf{m}_i(x|y)$ given $\mathcal{F}_{i-1}$ and the derivative of the conditional density $\mathsf{m}_i(y|x)$ given $\mathcal{F}_{i-1}$ with respect to the differenced error term is

$$\mathsf{m}_i(c|y) = \frac{\mathsf{m}_i(y,c)}{\mathsf{h}(y)} \quad \text{and} \quad \dot{\mathsf{m}}_i(c|x) = \frac{1}{\mathsf{k}_i(x)}\left.\frac{\partial \mathsf{m}_i(y,x)}{\partial y}\right|_{y=c}. \tag{5}$$

Finally, define the third and fourth moment of the scaled innovations ε_i/σ as

$$\varkappa_3 = \int_{-\infty}^{\infty} u^3\mathsf{f}(u)du, \quad \varkappa_4 = \int_{-\infty}^{\infty} u^4\mathsf{f}(u)du. \tag{6}$$

Table 1: Numerical computations of ϑ_1 and ϑ_2 given a standard normal density f.

	c									
	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75
ϑ_1	0.17	0.33	0.50	0.65	0.77	0.86	0.92	0.95	0.98	0.99
ϑ_2	-0.18	-0.23	-0.24	-0.23	-0.19	-0.15	-0.11	-0.07	-0.04	-0.03

If $\mathsf{f} = \varphi$ is standard normal then $\mathsf{h} = \varphi$ while $\mathsf{G} = 2\Psi - 1$. Further $\varkappa_3 = 0$ and $\varkappa_4 = 3$.

For later reference, define the following covariance terms, where $\mathsf{k}_i(x, y)$ is the joint density of x_{i+1} and ε_i/σ

$$\vartheta_1 = \mathsf{E}\{1_{(|\nabla\varepsilon_i|\le\sqrt{2}c\sigma)}1_{(|\nabla\varepsilon_{i+1}|\le\sqrt{2}c\sigma)}\}$$
$$= \int_{-\sqrt{2}c}^{\sqrt{2}c}\int_{-\sqrt{2}c}^{\sqrt{2}c}\int_{-\infty}^{\infty}\mathsf{f}(x+y+z)\mathsf{f}(x+z)\mathsf{f}(z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x, \tag{7}$$
$$\vartheta_2 = \mathsf{E}\{1_{(|\nabla\varepsilon_i|\le\sqrt{2}c\sigma)}(\varepsilon_{i+1}^2/\sigma^2 - 1)\} = \mathsf{E}\{1_{(|\nabla\varepsilon_i|\le\sqrt{2}c\sigma)}(\varepsilon_i^2/\sigma^2 - 1)\}$$
$$= \int_{-\infty}^{\infty}\int_{-\sqrt{2}c}^{\sqrt{2}c}(z^2-1)\mathsf{f}(z)\mathsf{f}(y+z)\,\mathrm{d}y\,\mathrm{d}z, \tag{8}$$
$$\vartheta_3 = \mathsf{E}\{1_{(|\nabla\varepsilon_i|\le\sqrt{2}c\sigma)}x_{i+1}\varepsilon_{i+1}/\sigma\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\sqrt{2}c}^{\sqrt{2}c}xy\mathsf{f}(y)\mathsf{k}_i(x, z+y)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x. \tag{9}$$

We tabulate the values of ϑ_1 and ϑ_2 for a standard normal density f in Table 1. The values for ϑ_3 depend on the regressors x_i. In the special case of x_i being the first-order autoregressive regressor we find $\vartheta_3 = -\vartheta_2$.

2.2 Algorithm

We are concerned about detecting location shifts, with an unknown location, duration, magnitude, sign and number. If there are L locations shifts, say, with breakpoints $0 = \tau_0 < \tau_1 < \cdots \tau_{L-1} < \tau_L = n$, then model (2) is modified as

$$y_i = \beta' x_i + \sum_{\ell=1}^{L}\mu_\ell 1_{(\tau_{\ell-1}<i\le\tau_\ell)} + \varepsilon_i \quad i = 1, ..., n+1. \tag{10}$$

As a member of an algorithm in the general-to-specific framework, Step-indicator Saturation extends the feature space, which is the space spanned by regressors, with step-indicators which model location shifts. We saturate the feature space with one step-indicator per period. The generalised feature space, with $1 < n_1 < n+1$, is then

$$y_i = \beta' x_i + \sum_{\ell=1}^{n_1}\gamma_\ell 1_{(i<\ell)} + \sum_{\ell=n_1+1}^{n+1}\gamma_\ell 1_{(n_1<i<\ell)} + \varepsilon_i \quad i = 1, ..., n+1. \tag{11}$$

The set of step-indicators is separated into two groups to avoid multicollinearity of the final step-indicators with the intercept in x_i.

We proceed by eliminating unnecessary step-indicators from the generalised feature space. In practise the Autometrics algorithm (see Doornik (2009)) or the related R package *gets* (see Pretis, Reade and Sucarrat (2016)) is used. In this paper, we will look at the highly stylised split-half algorithm, which mimics some of the basic structures in the Autometrics algorithm.

Algorithm 2.2. *The split-half step-indicator saturation algorithm.*

1. *Choose a cut-off value $c > 0$ to select breakpoints.*

2. *Split the sample of* n *observations into two sets* I_1 *and* I_2*, with* n_1 *and* n_2 *observations.*

3. *Calculate the least squares estimators for* (β, σ^2) *based on sample* I_j *as*

$$\hat{\beta}_j = (\sum_{n \in I_j} x_i x_i')^{-1} \sum_{n \in I_j} x_i y_i, \quad \hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{n \in I_j} (y_i - \hat{\beta}_j x_i)^2.$$

4. *Calculate the t-ratio's for* $i = \{1, ..., n-1\}$ *as*

$$v_i = 1_{(i \in I_1)} \frac{(\nabla y_i - \hat{\beta}_2' \nabla x_i)}{\hat{\omega}_{2,i}} + 1_{(i \in I_2)} \frac{(\nabla y_i - \hat{\beta}_1' \nabla x_i)}{\hat{\omega}_{1,i}} \quad (12)$$

where $\hat{\omega}_{j,i}^2$ *could be chosen as* $\hat{\sigma}_j^2 \{2 + \nabla x_i' (\sum_{k \in I_j} x_k x_k')^{-1} \nabla x_i\}$.

5. *Pick breakpoints* τ_ℓ *when* $|v_i| \geq c$.

6. *Let* $I_\ell = 1_{(\tau_{\ell-1} < i \leq \tau_\ell)}$*. Calculate the means* $\bar{x}_\ell = n_\ell^{-1} \sum_{n \in I_\ell} x_i$*. Demean the regressors* $\tilde{x}_n = x_i - \sum_{\ell=1}^{L+1} \bar{x}_\ell 1_{(i \in I_\ell)}$.

7. *Calculate the step-indicator saturation estimator as*

$$\hat{\beta}_2 = \left(\sum_{i=1}^{n} \tilde{x}_n \tilde{x}_n' \right)^{-1} \left(\sum_{i=1}^{n} \tilde{x}_n y_i \right) \quad \text{and} \quad \hat{\sigma}_2^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_2' x_i)^2.$$

Because we study the empirical gauge of Step-indicator Saturation, which is the frequency of detected location shifts under the null hypothesis of no location shift in the data generating process, we only consider step 1 to 5 of Algorithm 2.1.

2.3 Variations of the Algorithm

The split-half Step-indicator Saturation estimator is a special case of a general class of split-half algorithms. In this section we present a general framework for the split-half algorithms and proceed to discuss a number of variations of the algorithm within the framework.

In order to set up the split-half algorithm we need some notation. We divide the sample in two parts: the first $n_1 = \text{integer}(n/2)$ observations and the last $n_2 = n - n_1$ observations. We write the regression model (2) conformably in matrix notation as

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \quad (13)$$

We choose a set of dummy variables of the form

$$\begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} \quad (14)$$

where D_1 and D_2 are invertible matrices. With this general set-up we can cover a number of algorithms that are in use. For instance if we choose these matrices as identity matrices we get a dummy variable for each observation. This is known as impulse indicator saturation and is related to outlier detection methods in robust statistics, see Hendry, Johansen and Santos (2008), Johansen and Nielsen (2009). If instead D_1 and D_2 are suitable triangular matrices we get Step-indicator Saturation. This can be done in various ways and we return to a discussion of these choices later.

We start the split-half algorithm by estimating β from the second sample giving the estimators $\hat{\beta}_2 = (X_2' X_2)^{-1} X_2' Y_2$ and $\hat{\sigma}_2^2 = (Y_2 - X_2 \hat{\beta}_2)'(Y_2 - X_2 \hat{\beta}_2)/n_2$. From this we compute residuals for the first sample, $\hat\varepsilon_1 = Y_1 - X_1\hat\beta_2$ and regress these on D_1 giving $\hat\delta_1 = (D_1'D_1)^{-1}D_1'\hat\varepsilon_1$, which reduced to $\hat\delta_1 = D_1^{-1}\hat\varepsilon_1$, since D_1 is invertible. We will now form t-statistics for each of the coefficients in $\hat\delta_1$. To do this we compute the variance of $\hat\delta_1$ as if the regressors are fixed. This gives

$$\begin{aligned} \mathsf{V}(\hat\delta_1|X) &= \mathsf{V}\{D_1^{-1}(Y_1 - X_1\hat\beta_2)|X\} \\ &= \mathsf{V}[D_1^{-1}\{\varepsilon_1 - X_1(X_2'X_2)^{-1}X_2'\varepsilon_2\}|X] \\ &= \sigma^2 D_1^{-1}\{I + X_1(X_2'X_2)^{-1}X_1'\}(D_1^{-1})'. \end{aligned} \tag{15}$$

Here σ^2 is estimated by $\hat\sigma_2^2$. For each component of $\hat\delta_1$ we get the t-statistic

$$v_{1,\ell} = \frac{\hat\delta_{1,\ell}}{\sqrt{\hat{\mathsf{V}}(\hat\delta_{1,\ell}|X)}}.$$

We keep those dummy variables where $|v_{1,\ell}|$ is larger than a cut-off value c. We discuss later the rational to choose the correct c.

We swap the role of the two sub-samples, compute the same quantities $\hat\beta_1$, $\hat\sigma_1^2$ to get $\mathsf{V}(\hat\delta_2|X)$ and $v_{2,\ell}$, which are defined as before. We select those dummy variables where $|v_{2,\ell}| > c$. This concludes the split-half procedure.

For Step-indicator Saturation, we can choose D_1 and D_2 to be upper of lower triangular matrices of the form

$$D_\nabla = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & & 1 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad \text{or} \quad D_\Delta = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \tag{16}$$

Their inverses, used for the computation of $\hat\delta_i$, are

$$D_\nabla^{-1} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad \text{and} \quad D_\Delta^{-1} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix} \tag{17}$$

Both set of step-indicators D_∇ and D_Δ span the same space. However, they represent re-parametrisations. Either choice results in a single element of $\hat\delta_2$ that is computed differently from the remaining ones. For the choice of $D_2 = D_\nabla$, say, the estimate of the last dummy in half sample is $\hat\delta_{2,n} = \hat\varepsilon_n$ while it is $\hat\delta_{2,\ell} = \hat\varepsilon_\ell - \hat\varepsilon_{\ell+1}$ for the previous dummies. The variance of $\hat\delta_{2,\ell}$ changes accordingly from $\mathsf{V}(\hat\delta_{2,\ell}|X) = \sigma^2\{2 + \nabla x_{2,\ell}'(X_1'X_1)^{-1}\nabla x_{2,\ell}\}$ for $n_1 + 1 \le \ell \le n - 1$ to $\mathsf{V}(\hat\delta_{2,n}|X) = \sigma^2\{1 + x_{2,n}'(X_1'X_1)^{-1}x_{2,n}\}$ for the final dummy. In the case of $D_2 = D_\Delta$, the first dummy is estimated differently as $\hat\delta_{2,1} = \hat\varepsilon_{n_1+1}$ while the latter ones are estimated as $\hat\delta_{2,\ell} = \hat\varepsilon_{2,\ell} - \hat\varepsilon_{2,\ell-1}$. For the variance of $\hat\delta_{2,\ell}$ we get $\mathsf{V}(\hat\delta_{2,1}|X) = \sigma^2\{1 + x_{2,1}'(X_1'X_1)^{-1}x_{2,1}\}$ for the first dummy and $\mathsf{V}(\hat\delta_{2,\ell}|X) = \sigma^2\{2 + \nabla x_{2,\ell}'(X_1'X_1)^{-1}\nabla x_{2,\ell}\}$ for the subsequent ones. Both D_∇ and D_Δ are valid choices for D_i. The choice on the D_i depends on the application.

We illustrate the choice of the triangular matrices for the case of intercept correction in forecasting. Intercept correction, also known as setting the forecast back on track, becomes necessary if a structural break has occurred prior to the forecasting period, see Hendry and Clements (1998), p.197–198. Assume an AR(1) model with intercept is used for forecasting, such that we simplify model (2) to $y_i = \mu + \beta y_{i-1} + \varepsilon_i$. Then, the one period ahead forecast at period n is $\hat y_{n+1|n} = \hat\mu + \hat\beta y_n$. The forecast error is then $e_{n+1} = y_{n+1} - \hat y_{n+1} = (\mu - \hat\mu) + (\beta - \hat\beta)y_n + \varepsilon_{n+1}$. The idea of intercept correction is to add the forecast error of period n to the forecast for period

$n+1$: $\hat{y}_{n+1|n} = \hat{\beta} y_n + e_n$. Then, the corrected forecast error becomes $e^*_{n+1} = e_{n+1} - e_n$. In this model, if we use $\hat{\delta}_{2,\ell}$ when $D_2 = D_\Delta$, then we have

$$\hat{\delta}_{2,\ell} = \hat{\varepsilon}_{2,\ell} - \hat{\varepsilon}_{2,\ell-1} = e_\ell - e_{\ell-1} = e^*_\ell.$$

The challenge in forecasting exercises is to determine, whether a structural break has occurred in the forecast origin. If no structural break has occurred, then no intercept correction is necessary. In contrast, if a structural break has occurred, intercept correction can be useful in order to reduce the forecast error.

Step-indicator saturation can provide guidance on whether or not to intercept correct. The significance of the final t-statistic, $|v_n| > c$, can be used as a decision rule for intercept correction. The choices D_Δ and D_∇ offer different decision rules. In the case of D_Δ, for a forecast origin in n, the decision rule is based on the corrected forecast error $e^*_n = e_n - e_{n-1}$. In contrast, with D_∇, the decision rule is based on the forecast error e_n.

In Castle et al. (2016) the set of dummy variables is of the form

$$\begin{pmatrix} D_\nabla & F \\ 0 & D_\nabla \end{pmatrix} \tag{18}$$

where F is a matrix of ones. They arrive at matrix (18) by saturating the regression equation with n step-indicators which form the set $\mathcal{S} = \{1_{(i \leq j)}, j = 1, ..., n\}$. The split-half algorithm with this variation of dummy variables faces multicollinearity of the last step-indicator with the intercept, as the last column of matrix (18) is a column of ones. There are two ways to address the multicollinearity. In the first, the final step-indicator is dropped. The second is to modify the final step-indicator to be $1_{(n_1 < i \leq n)}$ instead. However, this unnecessarily complicates the asymptotic theory.

For our asymptotic analysis, we consider only those dummies with have symmetric decision rules. We select the symmetric dummies either using $M_\nabla = (I_{n_i-1}, 0)$ or $M_\Delta = (0, I_{n_i-1})$. It holds that

$$M_\nabla D_\nabla^{-1} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & -1 \end{pmatrix} = -M_\Delta D_\Delta^{-1}.$$

Thus, we have for, say, the first sample dummies, with $\hat{\delta}_\Delta = D_\Delta^{-1} \hat{\varepsilon}_i$

$$-M_\Delta \hat{\delta}_\Delta = -M_\Delta D_\Delta^{-1} \hat{\varepsilon}_1 = M_\nabla D_\nabla^{-1} \hat{\varepsilon}_1 = M_\nabla \hat{\delta}_\nabla.$$

and

$$\begin{aligned} \mathsf{V}(M_\Delta \delta_\Delta | X) &= \sigma^2 M_\Delta D_\Delta^{-1} \{I - X_1 (X_2' X_2)^{-1} X_1\} D_\Delta^{-1} M_\Delta \\ &= \sigma^2 M_\nabla D_\nabla^{-1} \{I - X_1 (X_2' X_2)^{-1} X_1\} D_\nabla^{-1} M_\nabla = \mathsf{V}(M_\Delta \delta_\Delta | X). \end{aligned}$$

The choice of D_Δ and D_∇ results in $n_1 - 1$ t-statistics for the symmetric dummies which are identical up the the sign, whereas t-statistics for end points are different.

2.4 The empirical gauge

We introduce the empirical gauge to help us choose the right cut-off value c. Following Hendry and Santos (2010), the empirical gauge of Step-indicator Saturation based on model (2) is defined as

$$\hat{\gamma}_n(c) = \frac{1}{n} \sum_{i=1}^{n} 1_{(|v_i| > c)}, \tag{19}$$

where the v_i is defined in equation (12) of Algorithm 2.2. The empirical gauge is the frequency of wrongly detected location shifts and can be interpreted as the rate of false positives.

The empirical gauge relates to the size of a hypothesis test. The size is the probability that we reject the null hypothesis of no location shifts in a specific period i given model (2). Our theory shows that the gauge converges asymptotically to the average of the size for all n periods. The asymptotic result is non-trivial, because of the presence of an estimation error for the regression parameter β and the scale parameter σ in the definition of v_i.

By controlling the gauge, we control the average frequency of false positives. This helps to benchmark algorithms, where we compare the power of the different algorithms when each algorithm is calibrated to satisfy the same gauge.

3 The main results

We proceed to present the theorems containing the asymptotic results on the empirical gauge of split-half Step-indicator Saturation. We start with the assumptions, which are sufficiently general to capture a large number of possible models.

3.1 Assumptions

In this section, we present a set of sufficient assumptions for the asymptotic theory of the empirical gauge of Step-indicator Saturation. When we use the Step-indicator Saturation estimator, we assume the density f is known. In most applications, our assumed density will be the normal density. But our assumptions allow for a range of alternative densities such as the t-density. The density f does not need to be symmetric.

Assumption 3.1. *Let $\mathcal{F}_i$ be a filtration so that ε_{i-1} and x_i are $\mathcal{F}_{i-1}$-measurable and ϵ_i/σ is independent of $\mathcal{F}_{i-1}$ with distribution function F and positive density f on $\mathbb{R}$ with derivative $\dot{\mathsf{f}}$ which is continuously differentiable and non-zero on the real line.*

Assumption 3.2. *Let $\mathcal{F}_i$ be an increasing sequence of σ-fields so ε_{i-1} and x_i are $\mathcal{F}_{i-1}$ measurable and ε_i is independent of $\mathcal{F}_{i-1}$. Let ε_i/σ have a continuously differentiable density f. For some values of $0 \leq \kappa < \eta \leq 1/4$, suppose*

(i) the density f satisfies

for $c > 0$, $c\mathsf{f}(c)$, $|\dot{\mathsf{f}}(c)|$ are decreasing large c and increasing for small c;

(ii) the conditional density $\mathsf{m}_i(y|x)$ given $\mathcal{F}_{i-1}$ satisfies

$\max_{1\leq i\leq n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}|\dot{\mathsf{m}}_i(y|x)| < \infty$;

(iii) the regressors x_i and the normalisation matrix N satisfy

(a) $\Sigma_n = \sum_{i=1}^n N'x_ix_i'N \overset{\mathsf{P}}{\to} \Sigma \overset{a.s.}{>} 0$;

(b) $\max_{1\leq i\leq n}|n^{1/2-\kappa}N'x_i| = \mathrm{O}_\mathsf{P}(1)$;

(c) $n^{-1}\sum_{i=1}^n \mathsf{E}_{i-1}|n^{1/2}N'\nabla x_i|^2 = \mathrm{O}_\mathsf{P}(1)$;

(d) $n^{-1}\mathsf{E}\sum_{i=1}^n|n^{1/2}N'\nabla x_i| = \mathrm{O}(1)$;

(iv) the estimators $(\hat\beta, \hat\sigma^2, \hat\omega_i)$ satisfy

(a) $N^{-1}(\hat\beta - \beta) = \mathrm{O}_\mathsf{P}(n^{1/4-\eta})$;

(b) $n^{1/2}(\hat\sigma^2 - \sigma^2) = \mathrm{O}_\mathsf{P}(n^{1/4-\eta})$;

(c) $\max_{1\leq i\leq n} n^{1/2}|\hat\omega_i - \hat\sigma| = \mathrm{o}_\mathsf{P}(1)$.

The conditions (i) and (iii) are satisfied in a wide range of models, as discussed in Johansen and Nielsen (2013, 2016a, 2016b). The normal and t-distribution both satisfy condition (i) on the density. Stationary, random walk and deterministically trending regressors satisfy condition (iii). Condition (iv) allows the standardised estimation errors to diverge at the rate of $n^{1/4-\eta}$ instead of being bounded in probability. By choosing $\eta = 1/4$, we would consider estimators with the standard rate of convergence.

3.2 The asymptotic results

The main results of this paper relate to the empirical gauge of split-half Step-indicator Saturation. Because we analyse the empirical gauge, we only need to consider the steps 1 to 4 in Algorithm 2.2. The first theorem states that the empirical gauge converges in probability to $\gamma = \mathsf{P}(|\nabla\varepsilon_i| \geq \sqrt{2}c\sigma)$. The second theorem provides an asymptotic distribution of the empirical gauge. The proofs are given in the Appendix E.

Theorem 3.3. *Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2. Suppose Assumptions 3.1(i), (iii)(a), (iii)(d) and (iv)(a)-(c) hold. Then, for fixed $c \in \mathbb{R}$, we have $\hat{\gamma}_n(c) \overset{p}{\to} \gamma$.*

As by definition the empirical gauge is bounded above and in particular for all c we have $\hat{\gamma}_n(c) < 1$. We can therefore strengthen the mode of convergence to mean square convergence with an appeal to Billingsley (1968, p. 32) and uniform integrability.

This theorem is used for the calibration of the Step-indicator Saturation algorithm. The unique tuning parameter of the algorithm is c. The theorem provides the rational to choose c to match the desired gauge. The empirical gauge $\hat{\gamma}_n$ converges to γ which in turn depends on c by (4). Thus, we first specify our tolerance for false positives expressed by γ, which given our assumed distribution of the error term, results in a selection quantile c.

For example, we assume a standard normal reference distribution f and a sample size $n = 100$. When we tolerate on average 1 falsely detected location shift, then we want the gauge to be $\hat{\gamma}_{100} = 1\%$. Thus we choose $\psi = 0.99$ and $c_{0.99} = 2.58$.

The next result gives us the asymptotic distribution of the empirical gauge. It relies on some new definitions. Let $x_{in} = N'x_i$ refer to normalised regressors. The normalisation depends on the type of the regressors. If a regressor is stationary $N = n^{-1/2}$. If it is random walk, the normalisation is $N = n^{-1}$. Define a row vector related to the covariance of the differenced regressors and the differenced error terms

$$\xi(c) = n^{-1}\sum_{i=1}^{n} \mathsf{E}_{i-1}\left(\frac{n^{1/2}\nabla x'_{in}}{\sqrt{2}\sigma} \middle| \frac{\nabla\varepsilon_i}{\sqrt{2}\sigma} = c \right). \tag{20}$$

The term

$$\xi_c = \xi(c) - \xi(-c) \tag{21}$$

vanishes in two distinct cases. If the regressors are strictly exogenous, then $\mathsf{E}_{i-1}\left(\nabla x'_{in} \middle| \nabla\varepsilon_i = c\right) = \mathsf{E}_{i-1}\left(\nabla x'_{in}\right)$, which does not depend on c and we have $\xi_c = 0$ for all c. Alternatively, if the regressors are of the random walk type, then as $\mathsf{E}_{i-1}\left(\nabla x'_{i} \middle| \nabla\varepsilon_i = c\right)$ is of order $\mathrm{O}(1)$, the normalisation in $x_{in} = N'x_i$, which is $N = n^{-1}$, dominates the expression and $\xi(c)$ vanishes for all c.

If the regressor x_i and the error term ε_i are stationary with geometrically declining autocorrelation (as an AR-process), then ∇x_{in}, $\nabla\varepsilon_i$ and consequently $\xi(c)$ are all stationary. In this case, use an appropriate Law of Large Numbers and the normalisation $N = n^{-1/2}$ to get that $\xi(c) \overset{\mathsf{P}}{\to} \mathsf{E}(\nabla x'_i|\nabla\varepsilon_i = c)$.

If ∇x_i and $\nabla\varepsilon_i$ are also jointly normally distributed and $\nabla x_{in} = v_i + w_i$ where v_i is $\mathcal{F}_{i-1}$ measurable and w_i is independent of $\mathcal{F}_{i-1}$ with $\mathsf{E}w_i = 0$ and

$$\begin{pmatrix} w_i \\ \nabla\varepsilon_i \end{pmatrix} = \mathsf{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{ww} & \sigma_{w\varepsilon} \\ * & 1 \end{pmatrix} \right\},$$

then $\xi(c) = n^{-1}\sum_{i=1}^{n}(v_i + c\sigma_{w\varepsilon}) \overset{\mathsf{P}}{\to} \mathsf{E}v_i + c\sigma_{w\varepsilon}$. We then get for the term $\xi(c) - \xi(-c) = 2c\sigma_{w\varepsilon}$. For example, if we run SIS on $y_i = \beta y_{i-1} + \varepsilon_i$ and $-1 < \beta < 1$, we get $\sigma_{\omega\varepsilon} = -\beta$ and thus $\xi(c) - \xi(-c) = -2c\beta$.

Theorem 3.4. *Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2. Suppose the Assumptions 3.2(i,ii,iiia,iiib,iiic,iv) hold. Then, for fixed $c \in \mathbb{R}$ and with*

$$h_i = \sqrt{\frac{n_2}{n_1}} 1_{(i \in I_1)} + \sqrt{\frac{n_1}{n_2}} 1_{(i \in I_2)}, \tag{22}$$

we get

$$\sqrt{n}\{\hat{\gamma}_n(c) - \gamma(c)\} = -\sqrt{n}\left\{\widehat{G}_n(0,0,c) - \psi\right\} - \frac{\mathsf{h}(c)}{\sqrt{n}} \sum_{i=1}^{n} h_i \left\{c\left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right) + \xi_c \Sigma^{-1} x_i \varepsilon_i\right\} + \mathsf{o}_{\mathsf{P}}(1).$$

Depending on the characteristics of the regressors, we get different asymptotic distributions. We distinguish three different cases: stationary regressors, unit-root and trend stationary regressors as well as the combination of both forms of regressors.

Corollary 3.5. *Let the regressors x_i be stationary. Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2. Suppose the Assumptions 3.2 (i,ii,iiia-iiic, iva-ivc) hold. Then, for fixed $c \in \mathbb{R}$, and with $\mu = \mathsf{E} x_i$ and*

$$\tilde{n} = 2\sqrt{n_1 n_2}/n \tag{23}$$

we have $n^{1/2}\{\hat{\gamma}_n(c) - \gamma\} \stackrel{\mathsf{D}}{\to} \mathsf{N}\{0, \gamma(1-\gamma) + \omega_* + \omega_{**}\}$ where

$$\omega_* = c\mathsf{h}(c)\{c\mathsf{h}(c)(\varkappa_4 - 1) + 4\vartheta_2\tilde{n}\} + 2(\vartheta_1 - \psi^2),$$
$$\omega_{**} = \sigma^2\xi_c^2\mathsf{h}^2(c) + 2\xi_c\sigma\Sigma^{-1}\mathsf{h}(c)\{c\mathsf{h}(c)\mu\varkappa_3 + \tilde{n}\vartheta_3\}.$$

The two components ω_* and ω_{**} of the asymptotic variance are zero in special cases. The term $\omega_* = 0$, if the variance of the error term σ is known and does not need to be estimated. The second term $\omega_{**} = 0$, when the regressors are all strictly exogenous, as in that case $\xi_c = 0$ for all c. Similarly, the term $\omega_{**} = 0$ vanishes asymptotically, if the regressors are of random walk type or trend stationary.

Corollary 3.6. *Let the regressors x_i be random walk or trend stationary regressors. Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2. Suppose the Assumptions 3.2(i,ii,iiia-iiic,iv) hold. Then, for fixed $c \in \mathbb{R}$, and with $\tilde{n} = 2\sqrt{n_1 n_2}/n$ we have $n^{1/2}\{\hat{\gamma}_n(c) - \gamma\} \stackrel{\mathsf{D}}{\to} \mathsf{N}\{0, \gamma(1-\gamma) + \omega\}$ where $\omega = c\mathsf{h}(c)\{c\mathsf{h}(c)(\varkappa_4 - 1) + 4\vartheta_2\tilde{n}\} + 2(\vartheta_1 - \psi^2)$.*

The expressions for ϑ_1, ϑ_2 and ϑ_3 are given in equation (7), (8) and (9). We also tabulated the resulting values for ϑ_1 and ϑ_2 for a standard normal density f in Table 1.

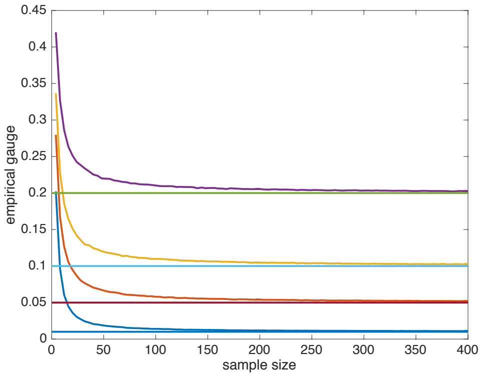
Example. If the error terms ε_i are standard normal distributed, the samples have equal size, and with stationary regressors x_i, the asymptotic distribution of the empirical gauge simplifies to

$$n^{1/2}\{\hat{\gamma}_n^{sn}(c) - \gamma\} \stackrel{\mathsf{D}}{\to} \mathsf{N}(0, \gamma(1-\gamma) + \omega_*^{sn} + \omega_{**}^{sn}) \tag{24}$$

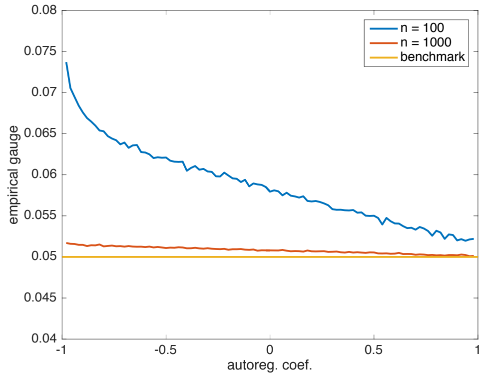
with $\omega_*^{sn} = 2c\mathsf{h}(c)\{c\mathsf{h}(c) + 2\vartheta_2\} + 2(\vartheta_1 - \psi^2)$ and $\omega_{**}^{sn} = \xi_c^2\mathsf{h}^2(c) + 2\xi_c\Sigma^{-1}\mathsf{h}(c)\vartheta_4$.

3.3 The simulation results

We illustrate the asymptotic results using the simulation results on convergence and asymptotic distribution of the Step-indicator Saturation estimator. All simulation exercises have been coded in MATLAB and I used the following simulation design. Our data generating process is a first-order auto-regression with an intercept equal to zero, where the coefficient varies by simulation. The first observation is equal to the realisation of the first error term. We replicate each simulation $M = 10000$ times.



(a) Bias at different sample sizes and selection quantiles



(b) Bias with different AR (1) coefficients

Figure 1: Simulation results compare the small sample bias for different sample sizes, selection probabilities and autoregressive coefficients.

3.3.1 Simulations on the convergence

In this section, we complement our analysis of the convergence of the empirical gauge with two simulations that illustrate finite sample properties of Step-indicator Saturation.

In Figure 1a we plot the empirical gauge for an increasing sample size and different selection probabilities γ, in order to analyse the small sample properties. The DGP we use is $y_i = \varepsilon_i$, where ε_i is standard normal. The sample size varies from 10 to 300 and we increase the sample size one-by-one. When we increase the sample size, we redraw all n observations. We run Algorithm 2.2 on the regression equation $y_i = \beta x_i + \varepsilon_i$, for a strictly exogenous one-dimensional regressor x_i which are i.i.d. draws from a standard normal distribution. We observe that the small sample bias is positive and vanishes quickly with growing sample size. Moreover, we observe that the small sample bias depends on the selection quantile c. In absolute value, the small sample bias is larger for bigger selection probabilities γ.

Figure 1b shows the relationship between the first-order autoregressive coefficient and the small sample bias. The DGP is $y_i = \beta y_{i-1} + \varepsilon_i$, where $-1 < \beta < 1$. We increase the coefficients by 0.1 in every step. The errors are standard normal. We run Algorithm 2.2 on the correctly specified first-order autoregressive model for two sample sizes $n = 100$ and $n = 1000$. In each iteration of the simulation, we redraw all observations. The figure illustrates that for constant n, the small sample bias increases for negative β and is the largest for β close to -1. In contrast, a positive β reduces the small sample bias. However, the small sample bias vanishes asymptotically, as predicted by the consistency of the gauge.

3.3.2 Simulations on the distribution

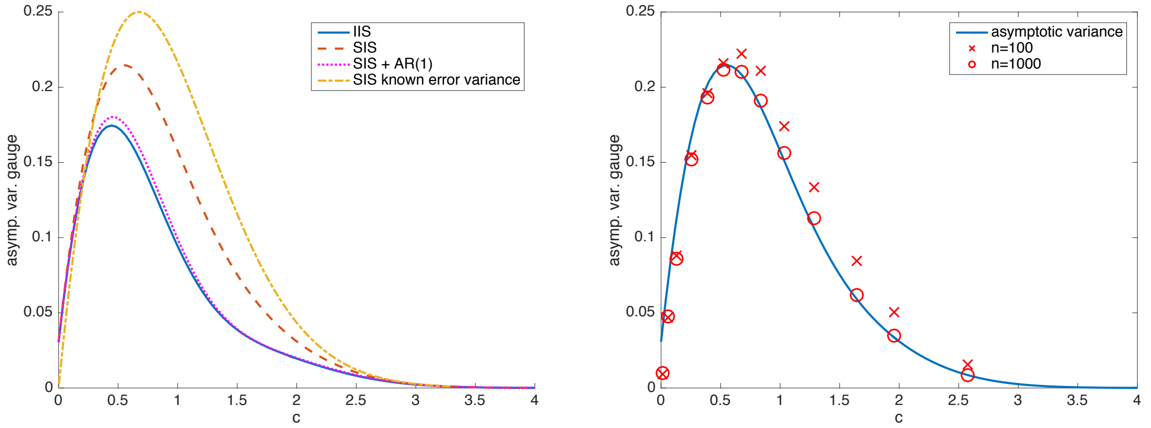
In this section we provide simulation results on the asymptotic distribution of the empirical gauge.

We compare the asymptotic variance of the gauge of the Impulse-indicator Saturation (IIS), as discussed in section 2.3, to Step-indicator Saturation (SIS). Johansen and Nielsen (2016b) present in their Corollary 5 the asymptotic distribution of the empirical gauge of Impulse-indicator Saturation as

$$n^{1/2}\{\hat{\gamma}_n^{IIS}(c) - \gamma\} \overset{\mathsf{D}}{\to} \mathsf{N}\{0, \gamma(1-\gamma) + 2c\mathsf{h}(c)\varkappa_2 + 2c^2\mathsf{h}^2(c)\} \tag{25}$$

with the truncated moment $\varkappa_2 = \int_{-c}^{c}(\varepsilon^2 - 1)\mathsf{f}(\varepsilon)d\varepsilon$.

In Figure 2a, we display the different theoretical asymptotic variance curves of the gauge depending on the selection quantile c for IIS and different models of SIS. The DGP is $y_i = \varepsilon_i$



(a) Asymptotic variance IIS and variations of SIS

(b) Small sample bias of the asymptotic variance

Figure 2: Simulation results plot the asymptotic variance curve as a function of the selection quantile c for different models and investigate the small sample bias.

Table 2: Table of the asymptotic variance of the empirical gauge of Step-indicator Saturation (in %) for different selection probabilities. The expression'-AR' refers to strictly exogenous regressors which are drawn from an i.i.d. standard normal distribution. The expression '+AR' refers to an autoregressive regressor. The expression '-σ^2' refers to an unknown error variance. The expression '+σ^2' refers to a known error variance.

	Algorithm			
ψ	IIS -AR -σ^2	SIS -AR -σ^2	SIS +AR -σ^2	SIS -AR +σ^2
5%	2.13	3.48	2.16	4.75
1%	0.71	0.89	0.75	0.99
0.5%	0.40	0.47	0.43	0.50
0.1%	0.01	0.01	0.01	0.01

with standard normal ε_i. We consider three distinct models for SIS. First, we run Algorithm 2.2 on the model $y_i = \beta x_i + \varepsilon_i$, where x_i are i.i.d. draws from a standard normal distribution and where $\hat{\omega}^2_{j,i} = \hat{\sigma}^2_j$. We therefore have from Theorem 3.4 that $\omega_{**} = 0$. Second, we run Algorithm 2.2, where we have an autoregressive regressor $x_i = y_{i-1}$ and where $\hat{\omega}^2_{j,i} = \hat{\sigma}^2_j$. Third, we run Algorithm 2.2 with x_i as in the first case, but with known error variance $\hat{\omega}^2_{j,i} = \sigma^2 = 1$. We therefore have from Theorem 3.4 that $\omega_* = 0$. We run the corresponding IIS algorithm, as outlined in section 2.3 on the model $y_i = \beta x_i + \varepsilon_i$, where x_i are i.i.d. draws from a standard normal distribution and where $\hat{\omega}^2_{j,i} = \hat{\sigma}^2_j$.

The following facts stand out. First, for all c the asymptotic variance of the gauge in IIS is lower than all three competing SIS models. Second, running SIS knowing the the variance of the error term σ results in a higher asymptotic variance of the gauge. Third, adding an autoregressive regressor to the estimated model, although unnecessary, as the DGP has no relevant regressor, results in a lower asymptotic variance of the gauge for SIS. This does not hold for IIS, as the theoretical asymptotic variance does not include terms that depend on the regressors. Finally, we observe that the asymptotic variance of the gauge falls rapidly for growing c. This gives a rational to choose a large c corresponding to selection probabilities of 1% or lower, as recommended by Castle et al. (2015), as the large c reduces the uncertainty on the empirical gauge.

How precise is the approximation of the asymptotic variance in small samples? Figure 2b shows the small sample bias of the estimated variance of the gauge of Step-indicator Saturation.

The simulation is based the DGP $y_i = \epsilon_i$, where the error terms are standard normal. We run Algorithm 2.2 on the model $y_i = \beta x_i + \varepsilon_i$ for a strictly exogenous one-dimensional regressor x_i which are i.i.d. draws from a standard normal distribution. We plot the results for two different sample sizes $n = 100$ and $n = 1000$. In each iteration of the simulation, we redraw all n observations. We observe that large selection quantiles c, the small sample bias is positive. Thus, on these large selection quantiles c and with small sample sizes, the empirical gauge will have a variance larger than the asymptotic variance.

In Table 2 we tabulate the asymptotic variance of the empirical gauge of Step-indicator Saturation for different selection probabilities $\psi = \{5\%, 1\%, 0.5\%, 0.1\%\}$ and compare to the asymptotic variance of the empirical gauge of Impulse-indicator Saturation as stated in (25). The data generating process is $y_i = \varepsilon_i$ with standard normal ε_i. We consider four models, which are the same models as in Figure 2a. We find that at a five percent selection probability, the variance of Step-indicator Saturation with an strictly exogenous regressor and with unknown variance of the error term is over 60 percent higher than for a similar setting with Impulse-indicator Saturation. This difference vanishes for smaller selection probabilities.

4 A class of auxiliary empirical process results of temporal differences of residuals

It is useful to consider an auxiliary class of empirical processes of temporal differences of residuals. Define the two-sided empirical distribution function as

$$\widehat{\mathsf{G}}_n(a, b, c) = \frac{1}{n} \sum_{i=1}^{n} 1_{(|\nabla \varepsilon_i - \nabla x'_{in} b| \leq \sqrt{2}\sigma c + \sqrt{2} n^{-1/2} a c)}. \tag{26}$$

We have $\sigma + n^{-1/2} a > 0$ by our choice of a as $\hat{a} = n^{1/2}(\hat{\sigma} - \sigma)$. In this case $\sigma + n^{-1/2} a = \hat{\sigma} > 0$ and we only need to consider $c > 0$. The (pseudo-)compensator (as $1_{(\nabla \varepsilon_i \leq c)}$ is not $\mathcal{F}_i$-measurable) is

$$\overline{\mathsf{G}}_n(a, b, c) = \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}_{i-1} 1_{(|\nabla \varepsilon_i - \nabla x'_{in} b| \leq \sqrt{2}\sigma c + \sqrt{2} n^{-1/2} a c)}, \tag{27}$$

where $\mathsf{E}_{i-1}(\cdot) = \mathsf{E}(\cdot|\mathcal{F}_{i-1})$. Note that $\overline{\mathsf{G}}_n(0, 0, c) = \mathsf{G}(c)$ as defined in (4). We standardize the empirical distribution function and get our empirical process of interest

$$\mathbb{G}_n(a, b, c) = n^{1/2}\{\widehat{\mathsf{G}}_n(a, b, c) - \overline{\mathsf{G}}_n(a, b, c)\}. \tag{28}$$

4.1 Explicit Assumptions

We keep track of the assumptions in a more explicit way than above.

Assumption 4.1. *Set $0 \leq \kappa < \eta \leq 1/4$ and suppose*
(i) the marginal density f *satisfies*
(a) moments: $\int_{-\infty}^{\infty} |u| \mathsf{f}(u) du < \infty$*;*
(b) boundedness: $\sup_{v \in \mathbb{R}} \{(1 + |v|) \mathsf{f}(v) + (1 + v^2) |\dot{\mathsf{f}}(v)|\} < \infty$*;*
(ii) the marginal density h *is smooth in that some* $C_\mathsf{H} > 0$ *exist such that for all* $a > 0$

$$\frac{\sup_{c \geq a} \mathsf{h}(c)}{\inf_{0 \leq c \leq a} \mathsf{h}(c)} \leq C_\mathsf{H} \quad \textit{and} \quad \frac{\sup_{c \leq -a} \mathsf{h}(c)}{\inf_{-a \leq c \leq 0} \mathsf{h}(c)} \leq C_\mathsf{H};$$

(iii) the conditional density $\mathsf{m}_i(y|x)$ *given* $\mathcal{F}_{i-1}$ *satisfies* $\max_{1 \leq i \leq n} \sup_{y \in \mathbb{R}, x \in \mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)| < \infty$*;*
(iv) the regressors x_i *satisfy*
(a) $\max_{1 \leq i \leq n} |n^{1/2-\kappa} N' \nabla x_i| = \mathrm{O}_\mathsf{P}(1)$*;*
(b) $n^{-1} \sum_{i=1}^{n} \mathsf{E}_{i-1} |n^{1/2} N' \nabla x_i|^2 = \mathrm{O}_\mathsf{P}(1)$*.*

Remark 1. If the conditional joint density m_i given $\mathcal{F}_{i-1}$ follows the multivariate normal distribution, 4.1(ib) would imply 4.1(ii). To see this, write the conditional density as,

$$\mathsf{f}_{Y|X}(y|x) = \phi(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu - \beta' x)^2\right\},$$

where $\omega = y - \mu - \beta' x$. Its derivative is then,

$$\frac{\partial \mathsf{f}_{Y|X}(y|x)}{\partial y} = \mathsf{f}_{Y|X}(y|x)\left\{-\sigma^{-2}(y - \mu - \beta' x)\right\} = -\frac{\omega\phi(\omega)}{\sigma^2}.$$

As $\sup_{v\in\mathbb{R}} |v\phi(v)| < \infty$, we have $\sup_{y\in\mathbb{R},x\in\mathbb{R}^p} \left|\dot{\mathsf{f}}_{Y|X}(y|x)\right| < \infty$.

Remark 2. The assumption of normally distributed error terms ϵ_i would imply Assumption 4.1(iia), because of the monotonicity of its tails is preserved in the convolution of the density f.

4.2 The empirical process results

The first empirical process result controls perturbations of the empirical process that are due to the estimation error of the location and the scale parameters from a regression. We prove uniform convergence in all three parameters a, b and c of the empirical process.

Theorem 4.2. *Suppose Assumptions 3.1 and 4.1(i,ii,iva) hold. Then, for any $B > 0$,*

$$\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\le n^{1/4-\eta}B} |\mathbb{G}_n(a,b,c) - \mathbb{G}_n(0,0,c)| = \mathsf{o}_\mathsf{P}(1).$$

The second empirical process result provides us with a bias correction term resulting from the quasi-compensators.

Theorem 4.3. *Suppose Assumptions 3.1 and 4.1(ib,iii,ivb) hold. Then, for all $B > 0$ and bias term $\mathcal{G}_n(a,b,c) = \mathsf{h}(c)(2ca\sigma^{-1} + \xi_c b)$ with ξ_c defined as (21), we have*

$$\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\le n^{1/4-\eta}B} \left| n^{1/2}\{\overline{G}(a,b,c_\psi) - \overline{G}(0,0,c_\psi)\} - \mathcal{G}_n(a,b,c_\psi)\right| = \mathsf{o}_\mathsf{P}(n^{-2\eta}).$$

The proof of these results relates to those of Koul and Ossiander (1994), Johansen and Nielsen (2016a), Jiao and Nielsen (2016) and in particular Berenguer-Rico and Nielsen (2018). An important difference is that those papers have indicators for residuals whereas we consider indicators for differences of residuals. This gives a more complicated dependence structure.

We also consider a result for an alternative empirical distribution function $\widehat{\mathsf{H}}_n$. It is used to handle the estimates of the standard error $\hat{\omega}^2_{j,i}$ in Algorithm 2.2. Define

$$\widehat{\mathsf{H}}_n(a,b,c) = \frac{1}{n}\sum_{i=1}^n 1_{(|\nabla\varepsilon_i - \sqrt{2}c(1+n^{-1/2}a) - \nabla x'_{in}b| \le \sqrt{2}|c|n^{-1/2}\delta_n)}, \tag{29}$$

where δ_n is some sequence of real numbers. Its pseudo-compensator which is constructed conditioning on filtration $\mathcal{F}_i$,

$$\overline{\mathsf{H}}_n(a,b,c) = \frac{1}{n}\sum_{i=1}^n \mathsf{E}_i 1_{(|\nabla\varepsilon_i - \sqrt{2}c(1+n^{-1/2}a) - \nabla x'_{in}b| \le \sqrt{2}|c|n^{-1/2}\delta_n)}. \tag{30}$$

The empirical process $\mathbb{H}_n(a,b,c)$ is defined conformably. The following result holds uniformly in a, b and c.

Theorem 4.4. *Suppose Assumptions 3.1 and 4.1(i,ii,iva) hold. Then, for any $B > 0$, it follows for some $\delta_n \to 0$ as $n \to \infty$ that*

$$\sup_{|a|,|b| \le n^{1/4-\eta}B} \sup_{c \in \mathbb{R}} n^{1/2} \widehat{\mathsf{H}}_n(a,b,c) = \mathrm{o}_{\mathsf{P}}(1).$$

To derive the asymptotic distribution of the gauge, we use the McLeish (1975) mixingale central limit theorem. A mixingale is defined as a sequence of $\mathcal{F}_i$-adapted random variables $\{X_i\}_{i=1}^n$ for which there exists a positive sequence $\psi_k \to 0$ as $k \to \infty$, such that for all $i \ge 1$, $k \ge 0$, $\mathsf{E}^{1/2}|\mathsf{E}_{i-k} X_i|^2 \le \psi_k$ and $\mathsf{E}^{1/2}|X_i - \mathsf{E}_{i+k} X_{i+1}|^2 \le \psi_{k+1}$.

The sequence ψ_k is considered to be of size $-p$ if there exists a positive sequence $\{L(k)\}$ which satisfies (i) $\sum_{k=1}^n \{kL(k)\}^{-1} < \infty$; (ii) $L(k) - L(k-1) = O\{L(k)/k\}$; (iii) $L(k)$ is eventually non-decreasing; such that $\psi_k = o\{(k^{1/2}L(k))^{-2p}\}$.

Theorem 4.5. *(McLeish, 1975, Theorem 2.6) Let $\mathcal{F}_i$-adapted sequence of random variables $\{X_i\}_{i=1}^n$ be a mixingale with ψ_k of size $-1/2$ satisfying $\mathsf{E}X_i = 0$ for all i and $\mathsf{E}S_n^2/n \to \sigma^2$, where $S_n = \sum_{i=1}^n X_i$. If X_i is uniformly square integrable and $\mathsf{E}\left\{n^{-1}(S_{k+n} - S_k)^2 | \mathcal{F}_{k-m}\right\} \to \sigma^2$ in $L_1(\Omega)$-norm as $\min(m,k,n) \to \infty$, then $n^{-1/2}S_n \overset{d}{\to} \mathsf{N}(0,\sigma^2)$.*

5 Two lemmas of general interest

We highlight two lemmas that have general applicability beyond the asymptotic results presented in this paper. The first one is the bias correction for differences in probabilities involving two random variables. The second lemma expands on the iterated martingale inequalities from Johansen and Nielsen (2016a).

5.1 Lemma for bias correction

The next lemma is used to derive the bias term for the compensators. The compensators require a more intricate analysis as compared to a standard Taylor approximation in Johansen and Nielsen (2009), Johansen and Nielsen (2016a) and Jiao and Nielsen (2016), because the conditioning on the filtration $\mathcal{F}_{i-1}$ leaves both $\nabla\varepsilon_i$ and $x_{i+1,n}$ as random variables.

Lemma 5.1. *Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}$ be random with density $\mathsf{m}(y,x)$ with respect to the product of the Lebesgue measure and some measure υ on $\mathbb{R}^p$. Suppose the density decomposes into conditional and marginal densities as $\mathsf{m}(y,x) = \mathsf{m}(y|x)\mathsf{m}_X(x) = \mathsf{m}(x|y)\mathsf{m}_Y(y)$ and that it is differentiable in y. Let $\mathsf{E}|N'X|^2 < \infty$ and let $b \in \mathbb{R}^p$. Then, with a normalisation matrix N we get*

$$\sup_{c \in \mathbb{R}} |\mathsf{P}(Y - X'Nb \le c) - \mathsf{P}(Y \le c) - \mathsf{m}_Y(c)\mathsf{E}(X'Nb|Y=c)| \le \frac{1}{2}|b|^2 \sup_{y \in \mathbb{R}, x \in \mathbb{R}^p} |\dot{\mathsf{m}}(y|x)| \, \mathsf{E}|N'X|^2,$$

where the derivative of the conditional density of X given Y is given as

$$\dot{\mathsf{m}}(c|x) = \frac{1}{\mathsf{m}_X(x)} \left.\frac{\partial \mathsf{m}(y,x)}{\partial y}\right|_{y=c}. \tag{31}$$

5.2 Exponential martingale inequalities

The chaining argument transforms a problem of proving tightness of a class of empirical processes to a problem of bounding tail probabilities for the maximum of a certain family of martingales. We bound the latter tail probabilities using the exponential martingale inequality that we present in this section. We modify the iterated exponential martingale inequality in Theorem 5.1 of Johansen and Nielsen (2016a). Because we work with bounded martingale difference sequences, we use the inequality in Freedman (1975) instead of Bercu and Touati (2008). Moreover, no iteration is necessary, as the Freedman inequality requires fewer bounds on the moment conditions of the martingale difference sequences.

Theorem 5.2. *For $1 \leq l \leq L$, let z_{li} be bounded and $\mathcal{F}_i$-adapted with $|z_{li} - \mathsf{E}_{i-1} z_{li}| \leq b$. Then for all $\kappa_0, \kappa_1 > 0$, we have*

$$\mathsf{P}\left(\max_{1\leq l \leq L} \left| \sum_{i=1}^{n} (z_{li} - \mathsf{E}_{i-1} z_{li}) \right| > \kappa_0 \right) \leq \frac{1}{\kappa_1} \mathsf{E} \max_{1\leq l\leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} z_{li}^2 + 2L \exp\left\{ -\frac{\kappa_0^2}{2(\kappa_1 + b\kappa_0)} \right\}.$$

The next result is a corollary to Theorem 5.2 and it modifies Theorem 5.2 in Johansen and Nielsen (2016a). Compared with their Theorem, we only require one condition, which is $2v > \varsigma$.

Theorem 5.3. *For $1 \leq l \leq L$, let z_{li} be bounded and $\mathcal{F}_i$-adapted such that $|z_{li} - \mathsf{E}_{i-1} z_{li}| \leq b$. Suppose $\exists \varsigma \geq 0$, $\lambda > 0$ so that $\mathsf{E} \max_{1\leq l\leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} z_{li}^2 = \mathrm{O}(n^{\varsigma})$ and $L = \mathrm{O}(n^{\lambda})$. Then, for all $2v > \varsigma$ and $\kappa > 0$ we get*

$$\lim_{n\to\infty} \mathsf{P}\left\{ \max_{1\leq l\leq L} \left| \sum_{i=1}^{n} (z_{li} - \mathsf{E}_{i-1} z_{li}) \right| > \kappa n^{v} \right\} = 0.$$

The next theorem is used to control the terms that arise in the chaining argument of the proof of Theorem 4.2. It is used when we apply the iterated martingale inequality of Theorem 5.2 on the $\mathsf{E} \max_{1\leq l\leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} z_{li}^2$ term in Theorem 5.3, in order to move backwards in the filtration.

Lemma 5.4. *For $1 \leq l \leq L$, let z_{li} be bounded and $\mathcal{F}_i$-adapted such that $|z_{li} - \mathsf{E}_{i-1} z_{li}| \leq b$. Suppose $\exists \varsigma \geq 0$, $\lambda > 0$ so that $\mathsf{E} \max_{1\leq l\leq L} \sum_{l=1}^{L} \mathsf{E}_{i-1} z_{li}^2 = \mathrm{O}(n^{\varsigma})$ and $L = \mathrm{O}(n^{\lambda})$. Then, for all $2v > \varsigma$ we get*

$$\int_0^{\infty} \mathsf{P}\left\{ \max_{1\leq l\leq L} \left| \sum_{i=1}^{n} (z_{li} - \mathsf{E}_{i-1} z_{li}) \right| > \kappa \right\} d\kappa = \mathrm{O}(n^{v}).$$

Appendix A: Proofs of Lemma 2.1, 5.1 and 5.2

Proof of Lemma 2.1. *a) Symmetry.* As $\mathsf{h}(x)$ is the density of the difference of two independent random variables ε_i and ε_j, symmetry is given, as

$$\varepsilon_i - \varepsilon_j \overset{\mathsf{D}}{=} \varepsilon_j - \varepsilon_i = -(\varepsilon_i - \varepsilon_j);$$

the ordering of ε_i and ε_j in the difference does not change the distribution of the difference.
b) Normal distribution. If f is the normal density with mean 0 and variance σ^2, then $\mathsf{h}(x)$, the density of the normalised difference of two i.i.d. normally distributed random variables, is also the normal density with mean 0 and variance σ^2. For the proof of the opposite direction, note first that from Lemma 2.1(a) we know that h is symmetric. Therefore it holds

$$\varepsilon_i - \varepsilon_j \overset{\mathsf{D}}{=} \varepsilon_i + \varepsilon_j.$$

By Pólya (1923), as we assumed a continuous density with bounded second moment, the normal density is the only density satisfying $\mathsf{f} = \mathsf{h}$.
c) Bounded densities. The boundedness of f implies the boundedness of h as

$$\mathsf{h}(v) = \sqrt{2} \int_{-\infty}^{\infty} \mathsf{f}(y)\mathsf{f}(v+y)dy \leq \sqrt{2} \sup_{v\in\mathbb{R}} \mathsf{f}(v) \int_{-\infty}^{\infty} \mathsf{f}(y)dy = \sqrt{2} \sup_{v\in\mathbb{R}} \mathsf{f}(v).$$

d) Bounded derivatives. The condition $\sup_{v\in\mathbb{R}} |\dot{\mathsf{f}}(v)| < \infty$ implies by the Leibniz rule for improper integrals and the triangle inequality that $\sup_{v\in\mathbb{R}} |\dot{\mathsf{h}}(v)| < \infty$, as

$$|\dot{\mathsf{h}}(v)| = \sqrt{2}\left|\frac{\partial}{\partial v}\int_{-\infty}^{\infty} \mathsf{f}(y)\mathsf{f}(v+y)dy\right| \leq \sqrt{2}\int_{-\infty}^{\infty} \mathsf{f}(y)\sup_{v\in\mathbb{R}}|\dot{\mathsf{f}}(v)|dy.$$

e) Bounded product. We have with the triangle inequality that

$$|v\mathsf{h}(v)| \leq \sqrt{2}\left|\int_{-\infty}^{\infty} \mathsf{f}(y)(v+y)\mathsf{f}(v+y)dy\right| + \sqrt{2}\left|\int_{-\infty}^{\infty} y\mathsf{f}(y)\mathsf{f}(v+y)dy\right| \leq 2\sqrt{2}\sup_{v\in\mathbb{R}}|v\mathsf{f}(v)|.$$

Therefore, the boundedness for $\sup_{v\in\mathbb{R}} |v|\mathsf{h}(v) < \infty$ holds, as long as $\sup_{v\in\mathbb{R}} |v|\mathsf{f}(v) < \infty$. □

Proof of Lemma 5.1. Let $\mathcal{P} = \mathsf{P}(Y - X'Nb \leq c) - \mathsf{P}(Y \leq c)$. Write the probability as an expectation and then expand the expectation to an integral

$$\mathcal{P} = \mathsf{E}\{1_{(Y-X'Nb\leq c)} - 1_{(Y\leq c)}\} = \int_{\mathbb{R}^p}\int_c^{c+x'Nb} \mathsf{m}(y,x)dyd\upsilon(x)$$

Differentiate $\mathsf{m}(y,x)$ w.r.t. y, and apply the mean value theorem to the inner integral, to get

$$\int_c^{c+x'Nb} \mathsf{m}(y,x)dy = (x'Nb)\mathsf{m}(c,x) + \frac{1}{2}(x'Nb)^2\dot{\mathsf{m}}(c^*|x)\mathsf{m}_X(x),$$

where $|c^* - c| \leq |x'Nb|$. Then, decompose $\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2$ with

$$\mathcal{P}_1 = \int_{\mathbb{R}^p} x'Nb\mathsf{m}(c,x)d\upsilon(x) \quad \text{and} \quad \mathcal{P}_2 = \frac{1}{2}\int_{\mathbb{R}^p}(x'Nb)^2\dot{\mathsf{m}}(c^*|x)\mathsf{m}_X(x)d\upsilon(x).$$

The first term is

$$\mathcal{P}_1 = \mathsf{m}_Y(c)\int_{\mathbb{R}^p} x'Nb\mathsf{m}_{X|Y}(x|c)d\upsilon(x) = \mathsf{m}_Y(c)\mathsf{E}(X'Nb|Y=c).$$

We have by the triangle inequality

$$|\mathcal{P} - \mathcal{P}_1| = |\mathcal{P}_2| \leq \frac{1}{2}\int_{\mathbb{R}^p}(x'Nb)^2\,|\dot{\mathsf{m}}(c^*|x)|\,\mathsf{m}_X(x)d\upsilon(x).$$

The norm inequality gives $(x'Nb)^2 \leq |N'x|^2|b|^2$ and we get uniformly in c that

$$|\mathcal{P}_2| \leq \frac{1}{2}|b|^2\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}|\dot{\mathsf{m}}(y|x)|\,\mathsf{E}|N'X|^2.$$

□

Proof of Theorem 5.2. Let $m_{li} = |z_{li} - \mathsf{E}_{i-1}z_{li}|$. Let $A_l = \sum_{i=1}^n m_{li}$ and $\mathcal{A} = (\max_{1\leq l\leq L}|A_l| > \kappa_0)$. Define $B_l = \sum_{i=1}^n \mathsf{E}_{i-1}m_{li}^2$ and $\mathcal{B} = (\max_{1\leq l\leq L} B_l \leq \kappa_1)$. Then we bound

$$\mathcal{P}(\mathcal{A}) = \mathsf{P}(\mathcal{A}\cap\mathcal{B}) + \mathsf{P}(\mathcal{A}\cap\mathcal{B}^c) \leq \mathsf{P}(\mathcal{A}\cap\mathcal{B}) + \mathsf{P}(\mathcal{B}^c).$$

Bounding $\mathsf{P}(\mathcal{A}\cap\mathcal{B})$. Let $\mathcal{A}_l = (|A_l| > \kappa_0)$ and $\mathcal{B}_l = (|B_l| \leq \kappa_1)$. Note $\mathcal{A} = \bigcup_{l=1}^L \mathcal{A}_l$ and $\mathcal{B} \subset \mathcal{B}_l$ and apply Boole's Inequality to get

$$\mathsf{P}(\mathcal{A}\cap\mathcal{B}) \leq \sum_{l=1}^L \mathsf{P}(\mathcal{A}_l\cap\mathcal{B}) \leq \sum_{l=1}^L \mathsf{P}(\mathcal{A}_l\cap\mathcal{B}_l).$$

Apply Freedman (1975, Theorem 1.6) with $|m_{li}| \le b$ to get

$$\mathsf{P}(\mathcal{A}\cap\mathcal{B}) \le \sum_{l=1}^{L} \mathsf{P}\left\{(A_l > \kappa_0)\cap\mathcal{B}_l\right\} + \mathsf{P}\left\{(-A_l > \kappa_0)\cap\mathcal{B}_l\right\} \le 2L\exp\left\{-\frac{\kappa_0^2}{2(\kappa_1 + b\kappa_0)}\right\}.$$

Bounding $\mathsf{P}(\mathcal{B}^c)$. The Markov inequality and $\mathsf{E}_{i-1}m_{li}^2 \le \mathsf{E}_{i-1}z_{li}^2$ give

$$\begin{aligned}\mathsf{P}(\mathcal{B}^c) &= \mathsf{P}(\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_{i-1}m_{li}^2 > \kappa_1)\\ &\le \frac{1}{\kappa_1}\mathsf{E}\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_{i-1}m_{li}^2 \le \frac{1}{\kappa_1}\mathsf{E}\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{li}^2.\end{aligned}$$

Finally, combine the bounds on $\mathsf{P}(\mathcal{A}\cap\mathcal{B})$ and $\mathsf{P}(\mathcal{B}^c)$. □

Proof of Theorem 5.3. Apply Theorem 5.2 with $\kappa_0 = \kappa n^v$ and $\kappa_1 = \kappa^2 n^{2v}\{4(1+b)\lambda\log n\}^{-1}$ for some $\kappa > 0$. Let n be fixed and sufficiently large such that $\kappa_0 < \kappa_1$. Use the bound

$$\exp\left\{-\frac{\kappa_0^2}{2(\kappa_1 + b\kappa_0)}\right\} \le \exp\left\{-\frac{\kappa_0^2}{2\kappa_1(1+b)}\right\} = n^{-2\lambda}, \tag{A.1}$$

For our probability of interest, we get that

$$\mathcal{P}_n = \mathrm{O}\left\{\frac{4(1+b)n^{\varsigma}\lambda\log n}{\kappa^2 n^{2v}} + n^{-\lambda}\right\} = \mathrm{o}(1),$$

where we used $2v > \varsigma \ge 0$. □

Proof of Lemma 5.4. Let $W = \max_{1\le l\le L}|\sum_{i=1}^{n}(z_{li} - \mathsf{E}_{i-1}z_{li})|$. Our object of interest is then

$$\mathcal{E}_W = \int_0^\infty \mathsf{P}\left(W > \kappa\right)d\kappa \le n^v + \int_{n^v}^\infty \mathsf{P}\left(W > \kappa\right)d\kappa.$$

Use the substitution formula with $\kappa = n^v\tilde\kappa_0^2$ so $d\kappa = 2n^v\tilde\kappa_0 d\tilde\kappa_0$ to get $\mathcal{E}_W \le n^v + 2n^v\mathcal{J}$, where $\mathcal{J} = \int_1^\infty \tilde\kappa_0\mathsf{P}\left(W > \tilde\kappa_0^2 n^v\right)d\tilde\kappa_0$. Apply Theorem 5.2 to get $\mathcal{J} \le \mathcal{J}_1 + \mathcal{J}_2$, where

$$\mathcal{J}_1 = \int_1^\infty \frac{\tilde\kappa_0}{\kappa_1}\mathsf{E}\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{li}^2 d\tilde\kappa_0, \quad \mathcal{J}_2 = 2L\int_1^\infty \tilde\kappa_0\exp\left\{-\frac{\kappa_0^2}{2(\kappa_1 + b\kappa_0)}\right\}d\tilde\kappa_0,$$

with $\kappa_0 = \tilde\kappa_0^2 n^v$ and where we choose $\kappa_1 = \tilde\kappa_0^4 n^{2v}\{4(1+b)\lambda\log n + 6(1+b)\log\tilde\kappa_0\}^{-1}$. We show that $\mathcal{J}_1$ and $\mathcal{J}_2$ vanish with $n\to\infty$.

For $\mathcal{J}_1$, write out κ_0 and κ_1 to get

$$\mathcal{J}_1 = \frac{\mathsf{E}\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{li}^2}{n^{2v}}\int_1^\infty \frac{4(1+b)\lambda\log n + 6(1+b)\log\tilde\kappa_0}{\tilde\kappa_0^3}d\tilde\kappa_0.$$

Decompose this term, such that $\mathcal{J}_1 \le \mathcal{J}_1^\dagger + \mathcal{J}_1^\ddagger$ and use $\mathsf{E}\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{li}^2 = \mathrm{O}(n^\varsigma)$ with

$$\mathcal{J}_1^\dagger = \mathrm{O}\left(\frac{n^\varsigma\log n}{n^{2v}}\int_1^\infty \frac{1}{\tilde\kappa_0^3}d\tilde\kappa_0\right), \quad \mathcal{J}_1^\ddagger = \mathrm{O}\left(\frac{n^\varsigma}{n^{2v}}\int_1^\infty \frac{\log\tilde\kappa_0}{\tilde\kappa_0^3}d\tilde\kappa_0\right).$$

With $\varsigma < 2v$ and as both integrals are finite, we have $\mathcal{J}_1^\dagger = \mathrm{o}(1)$, $\mathcal{J}_1^\ddagger = \mathrm{o}(1)$ and thus $\mathcal{J}_1 = \mathrm{o}(1)$.

For $\mathcal{J}_2$, use the bound in (A.1) and by our choice of κ_0 and κ_1 we get

$$\mathcal{J}_2 \le 2L\int_1^\infty \tilde\kappa_0\exp\left\{-\frac{\kappa_0^2}{2\kappa_1(1+b)}\right\}d\tilde\kappa_0 = 2L\int_1^\infty n^{-2\lambda}\tilde\kappa_0^{-2}d\tilde\kappa_0.$$

Then, by the assumption $L = \mathrm{O}(n^\lambda)$ we get that $\mathcal{J}_2 = \mathrm{o}(1)$, as the integral is finite. □

Appendix B: A metric on $\mathbb{R}$ and some inequalities

Introduce notation for the difference of two indicator functions

$$J_i(x, y) = 1_{(\nabla\varepsilon_i \le \sqrt{2}\sigma y)} - 1_{(\nabla\varepsilon_i \le \sqrt{2}\sigma x)}, \tag{B.1}$$

where the normalised differenced error terms have density h, see (3).

We partition the unit interval into K intervals of equal size. The corresponding quantiles of the distribution function

$$\mathsf{H}(c) = \int_{-\infty}^{c} \mathsf{h}(u)du \tag{B.2}$$

are $c_k = \mathsf{H}^{-1}(k/K)$ for $0 \le k \le K$, such that

$$|\mathsf{E}J_i(c_{k-1}, c_k)| = \mathsf{E}J_i(c_{k-1}, c_k) = \mathsf{H}(c_k) - \mathsf{H}(c_{k-1}) = K^{-1}. \tag{B.3}$$

Choose the number of intervals K so large so that c_-, c_+ exist which are (weakly) separated from zero by grid points in the sense that $c_{k_--1} \le c_- \le c_{k_-} \le 0$ and $0 \le c_{k_+-1} \le c_+ \le c_{k_+}$ and

$$\mathsf{h}(c_-) = \mathsf{h}(c_+) = \frac{1}{C_\mathsf{H} K^{1/2}}, \tag{B.4}$$

where C_H is defined as in Assumption 4.1(ii). The condition is satisfied for large K since h is assumed to be continuous. Next, we restate Lemma A.6. from Berenguer and Nielsen (2017) with their H^m redefined in terms of our distribution function H in (B.2).

Lemma B.1. *(Berenguer and Nielsen, 2017, Lemma A.6.) Let Assumptions 3.1, 4.1 (ib,ii) hold. Then*

$$\max_{1 \le k \le K} |\dot{\mathsf{H}}(c_k) - \dot{\mathsf{H}}(c_{k-1})| = \mathsf{O}(K^{-1/2}).$$

We state a preliminary inequality from Jiao and Nielsen (2016), which will be used repeatedly in the proofs. The purpose of this lemma is to handle terms involving $c\mathsf{f}(c^*)$, where $c \neq c^*$ but where the assumptions in the Lemma are satisfied.

Lemma B.2. *(Jiao and Nielsen, 2016, Lemma 1.11) If $|c^* - c| \le |Ac + B|$ and $|A| \le 1/2$, then $|c| \le 2(|c^*| + |B|)$ and $(Ac + B)^2 \le 16\{A^2(c^*)^2 + B^2\}$.*

The next lemma is used to control the difference of two indicator functions where a limit on the distance between the boundaries of the indicator functions exists.

Lemma B.3. *(Johansen and Nielsen, 2009, Lemma 1) Let $e < f$ and $e_0 < f_0$, and choose $\zeta \ge \max(|e - e_0|, |f - f_0|)$ to get*

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

The final lemma, stated without proof, lists the terms in the covariance matrix that are used in the derivation of the asymptotic distribution of the empirical gauge.

Lemma B.4. *Suppose the Assumptions 3.2 hold. Introduce the vector*

$$V_i = \left\{1_{(|\nabla\varepsilon_i| \le \sqrt{2}c\sigma)} - \psi, c\mathsf{h}(c)\left(\varepsilon_i^2/\sigma^2 - 1\right)h_i, \xi_c \Sigma^{-1}\mathsf{h}(c) x_i \varepsilon_i h_i\right\}',$$

where the regressors x_i are stationary with first moment $\mu = \mathsf{E}x_i$. Recall the definitions of $\Sigma = \mathsf{E}(x_i x_i')$; γ, ψ from (1), (4); $\varkappa_3$, $\varkappa_4$ from (6); $\vartheta_1, \vartheta_2, \vartheta_3$ from (7), (8), (9); ξ_c from (21) and h_i from (22). The vector V_i has mean zero and variance-covariance matrices

$$\mathsf{E}(V_i^2) = \begin{pmatrix} (1-\gamma)\gamma & c\mathsf{h}(c)\vartheta_2 h_i & 0 \\ * & c^2\mathsf{h}^2(c)(\varkappa_4 - 1)h_i^2 & \mu\xi_c\Sigma^{-1}c\mathsf{h}^2(c)\sigma\varkappa_3 h_i^2 \\ * & * & \sigma^2\xi_c^2\mathsf{h}^2(c)h_i^2 \end{pmatrix}$$

and

$$\mathsf{E}(V_iV_{i+1}') = \begin{pmatrix} \vartheta_1 - \psi^2 & c\mathsf{h}(c)\vartheta_2 h_i & \xi_c\Sigma^{-1}\mathsf{h}(c)\vartheta_3 h_i \\ 0 & c^2\mathsf{h}^2(c)(\varkappa_4-1)h_i^2 & 0 \\ 0 & 0 & \sigma^2\xi_c^2\mathsf{h}^2(c)h_i^2 \end{pmatrix}.$$

For random walk and trend stationary regressors x_i, we have for $n \to \infty$ that $\xi_c \to 0$ and therefore $\mu\xi_c\Sigma^{-1}c\mathsf{h}^2(c)\sigma\varkappa_3 h_i^2 \to 0$, which simplifies the variance-covariance matrices accordingly.

Appendix C: Applications of the martingale inequality

The proof of Theorem 4.2 uses a chaining argument. In this argument, we have to control the tail probabilities of the maximum of a certain family of martingales. The lemmas in this section provide the bounds to these tail probabilities, by repeated use of the exponential martingale inequality in Theorem 5.2.

Lemma C.1. *Suppose Assumptions 3.1, 4.1(i) hold. For $n > 0$, let $c_k = \mathsf{H}^{-1}(k/K)$ for $0 \leq k \leq K$ with $K = \mathrm{O}(n^{1/2})$. For all $c \in \mathbb{R}$ and $\tilde{c} = c(1+\sigma^{-1}n^{-1/2}a)$ with $0 < a \leq n^{1/4-\eta}B$ and $\eta < 1/4$ choose $k, \tilde{k}$ so that $c \leq c_k < c_{\tilde{k}} < \tilde{c}$. Then, with $8\omega < 1 + 4\eta$, we have that*

$$\max_{0<k<\tilde{k}<K}\left|\sum_{i=1}^n\{J_i(c_k, c_{\tilde{k}}) - \mathsf{E}_i J_i(c_k, c_{\tilde{k}})\}\right| = \mathrm{o}_{\mathsf{P}}(n^{1/2-\omega}), \tag{C.1}$$

$$\max_{0<k<\tilde{k}<K}\left|\sum_{i=1}^n\{\mathsf{E}_i J_i(c_k, c_{\tilde{k}}) - \mathsf{E}_{i-1} J_i(c_k, c_{\tilde{k}})\}\right| = \mathrm{o}_{\mathsf{P}}(n^{1/2-\omega}). \tag{C.2}$$

Note that the martingale difference sequence in (C.1) *is $\mathcal{F}_{i+1}$ and in* (C.2) *$\mathcal{F}_i$-adapted.*

Proof of Lemma C.1. Without loss of generality, let $\sigma = 1$. Use Theorem 5.3 with $v = 1/2 - \omega$. To show (C.1), consider the $\mathcal{F}_{i+1}$-adapted term

$$z_{li} = J_i(c_k, c_{\tilde{k}}) = 1_{(\varepsilon_{i+1} \geq \varepsilon_i - \sqrt{2}c_{\tilde{k}})} - 1_{(\varepsilon_{i+1} \geq \varepsilon_i - \sqrt{2}c_k)},$$

where l represents the indices $k, \tilde{k}$. To show (C.2), consider the $\mathcal{F}_i$-adapted term $z_{li}^+ = \mathsf{E}_i z_{li}$. Note the bound $0 \leq z_{li}, z_{li}^+ \leq 1$.

The parameter λ. The set of indices l has the size $L = \mathrm{O}(n)$, since $L \leq K^2$ and $K = \mathrm{O}(n^{1/2})$.

The parameter ς. Use $c \leq c_k < c_{\tilde{k}} < \tilde{c}$ to bound $z_{li} \leq 1_{(\varepsilon_{i+1} \geq \varepsilon_i - \sqrt{2}\tilde{c})} - 1_{(\varepsilon_{i+1} \geq \varepsilon_i - \sqrt{2}c)}$. Then, by the Law of Iterated Expectations in the latter term, we can write

$$\mathsf{E}_i z_{li} \leq \mathsf{F}(\varepsilon_i - \sqrt{2}c) - \mathsf{F}(\varepsilon_i - \sqrt{2}\tilde{c}), \quad \mathsf{E}_{i-1} z_{li}^+ = \mathsf{E}_{i-1} z_{li} \leq \mathsf{H}(\tilde{c}) - \mathsf{H}(c),$$

since F is the cdf of ε_{i+1} where ε_i is known and H is the cdf of $(\varepsilon_i - \varepsilon_{i+1})/\sqrt{2}$. By the Mean Value Theorem with $\tilde{c} = c(1+n^{-1/2}a)$ we get

$$\mathsf{E}_i z_{li} \leq \sqrt{2}n^{-1/2}a|c|\mathsf{f}(c_1^*), \quad \mathsf{E}_{i-1} z_{li}^+ \leq n^{-1/2}a|c|\mathsf{h}(c_2^*),$$

where

$$|c_1^* - \varepsilon_i + \sqrt{2}c| \leq \sqrt{2}n^{-1/2}a|c|, \quad |c_2^* - c| \leq n^{-1/2}a|c|.$$

As $|X - Y| \leq |Z|$ implies $|X| \leq |Z| + |Y|$, we have $|c_1^* + \sqrt{2}c| \leq n^{-1/2}a|\sqrt{2}c| + |\varepsilon_i|$. For a sufficiently large n, we have $n^{-1/2}a \leq 1/2$ and Lemma B.2 yields $|c| \leq \sqrt{2}(|c_1^*| + |\varepsilon_i|)$ and $|c| \leq 2|c_2^*|$. By the bounds on $|c|$ and $a \leq n^{1/4-\eta}B$ we get

$$\mathsf{E}_i z_{li} \leq 2\sqrt{2}Bn^{-1/4-\eta}\left\{\sup_{v\in\mathbb{R}}(|\varepsilon_i| + |v|)\mathsf{f}(v)\right\}, \quad \mathsf{E}_i z_{li}^+ \leq 2Bn^{-1/4-\eta}\sup_{v\in\mathbb{R}}|v|\mathsf{h}(v).$$

As $\mathsf{E}|\varepsilon_i| < \infty$ by Assumption 4.1(ia) and $\sup_{v\in\mathbb{R}}(1+|v|)\mathsf{f}(v) < \infty$ by Assumption 4.1(ib) we have with $\varsigma = 3/4 - \eta$ that

$$\mathsf{E}\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_i z_{li} = 2Bn^{-1/4-\eta}\sum_{i=1}^{n}\left\{\sup_{v\in\mathbb{R}}(\mathsf{E}|\varepsilon_i| + |v|)\mathsf{f}(v)\right\} = \mathrm{O}(n^{\varsigma}),$$

$$\mathsf{E}\max_{1\le l\le L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{li}^{+} = 2Bn^{3/4-\eta}\sup_{v\in\mathbb{R}}|v|\mathsf{h}(v) = \mathrm{O}(n^{\varsigma}).$$

The condition $2v > \varsigma$. To satisfy this condition, we set $v = 1/2 - \omega$, with $8\omega < 1 + 4\eta$. □

Lemma C.2. *Suppose Assumptions 3.1, 4.1 (ib,ii) hold. For $\delta, n > 0$, partition c such that $c_k = \mathsf{H}^{-1}(k/K)$ for $0 \le k \le K$ with $K = int(n^{1/2}/\delta)$. Consider two cases: $d - 0$ and $d = K^{-1/2}$. Then, with $4\omega < 1$, we get for both cases*

$$\max_{1\le k\le K}\left|\sum_{i=1}^{n} J_i(c_{k-1}-d, c_k+d) - \mathsf{E}_i J_i(c_{k-1}-d, c_k+d)\right| = \mathrm{o}_{\mathsf{P}}(n^{1/2-\omega}), \tag{C.3}$$

$$\max_{1\le k\le K}\left|\sum_{i=1}^{n} \mathsf{E}_i J_i(c_{k-1}-d, c_k+d) - \mathsf{E}_{i-1} J_i(c_{k-1}-d, c_k+d)\right| = \mathrm{o}_{\mathsf{P}}(n^{1/2-\omega}). \tag{C.4}$$

Note that the martingale difference sequence in (C.3) *is $\mathcal{F}_{i+1}$ and in* (C.4) *$\mathcal{F}_i$-adapted. If $d = 0$ then Assumptions 4.1 (ib,ii) are unnecessary.*

Proof of Lemma C.2. Use Theorem 5.3 with $v = 1/2 - \omega$. Let $z_{ki,l} = J_i(c_{k-1} - d, c_k + d)$. The martingale difference sequences are bounded by $|z_{ki,d} - \mathsf{E}_i z_{ki,d}|$, $|\mathsf{E}_i z_{ki,d} - \mathsf{E}_{i-1} z_{ki,d}| < 2$.

The parameter λ. The index k has size $K = \mathrm{O}(n^{\lambda})$ with $\lambda = 1/2$.

The parameter ς. Consider $\mathcal{E}_{n2} = \mathsf{E}\max_{1\le k\le K}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{ki,d}$ and $d = K^{-1/2}$. Define the term $\mathcal{H}_k = \mathsf{H}(c_k + K^{-1/2}) - \mathsf{H}(c_{k-1} - K^{-1/2})$, such that (B.3) implies $\mathsf{E}_{i-1}z_{ki,d} = \mathcal{H}_k$. Use the mean value theorem, for c_k^*, c_{k-1}^* so $|c_k^* - c_k|, |c_{k-1}^* - c_{k-1}| \le 2K^{-1/2}$, such that

$$\mathsf{H}(c_k + K^{-1/2}) = \mathsf{H}(c_k) + K^{-1/2}\dot{\mathsf{H}}(c_k) + (K^{-1}/2)\ddot{\mathsf{H}}(c_k^*),$$

$$\mathsf{H}(c_{k-1} + K^{-1/2}) = \mathsf{H}(c_{k-1}) + K^{-1/2}\dot{\mathsf{H}}(c_{k-1}) + (K^{-1}/2)\ddot{\mathsf{H}}(c_{k-1}^*).$$

Take the difference and apply the triangle inequality

$$|\mathcal{H}_k| = |\mathsf{H}(c_k) - \mathsf{H}(c_{k-1})| + K^{-1/2}|\dot{\mathsf{H}}(c_k) - \dot{\mathsf{H}}(c_{k-1})| + (K^{-1}/2)|\ddot{\mathsf{H}}(c_k^*) - \ddot{\mathsf{H}}(c_{k-1}^*)|.$$

By construction, the first term is $1/K$. By Lemma B.1 and Assumption 4.1 (ib,ii) we have $|\dot{\mathsf{H}}(c_k) - \dot{\mathsf{H}}(c_{k-1})| = \mathrm{O}(K^{-1/2})$. The second term is therefore $\mathrm{O}(K^{-1})$. By Assumption 4.1 (ib) $\ddot{\mathsf{H}}(.)$ is uniformly bounded and thus the third term is $\mathrm{O}(K^{-1})$. Consequently, $|\mathcal{H}_k| = \mathrm{O}(K^{-1})$ so that

$$\max_{1\le k\le K}\mathsf{E}_{i-1}z_{ki,d} = \max_{1\le k\le K}\mathsf{E}_{i-1}J_i(c_{k-1} - K^{-1/2}, c_k + K^{-1/2}) = \delta\mathrm{O}(n^{-1/2}). \tag{C.5}$$

Then, it holds that $\mathcal{E}_{n2} = \mathsf{E}\sum_{i=1}^{n}\delta\mathrm{O}(n^{-1/2}) = \delta\mathrm{O}(n^{1/2})$. As $\delta > 0$ can be chosen arbitrarily small, we have $\mathcal{E}_{n2} = \mathrm{o}(n^{1/2})$. If $d = 0$, the previous assumptions are unnecessary. We get by the construction of the partition in equation (B.3) that $\mathsf{E}_{i-1}J_i(c_{k-1}, c_k) = \mathrm{O}(n^{-1/2})$.

Consider $\mathcal{E}_{n1} = \mathsf{E}\max_{1\le k\le K}\sum_{i=1}^{n}\mathsf{E}_i z_{ki,d}$. Add and subtract $\mathsf{E}_{i-1}z_{ki,d}$ and use the triangle inequality to get

$$\mathcal{E}_{n1} \le \mathsf{E}\max_{1\le k\le K}\left|\sum_{i=1}^{n}\mathsf{E}_i z_{ki,d} - \mathsf{E}_{i-1}z_{ki,d}\right| + \mathsf{E}\max_{1\le k\le K}\left|\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{ki,d}\right|. \tag{C.6}$$

The second term in (C.6) is of the same order than $\mathcal{E}_{n2}$. For the first term in (C.6), use $\mathsf{E}[X] = \int_0^\infty \{1 - F(x)\}dx$ on a non-negative random variable X, to get

$$\mathsf{E} \max_{1\le k\le K} \left| \sum_{i=1}^{n} \mathsf{E}_i z_{ki,d} - \mathsf{E}_{i-1} z_{ki,d} \right| \le \int_0^\infty \mathsf{P} \left\{ \max_{1\le k\le K} \left| \sum_{i=1}^{n} \mathsf{E}_i z_{ki,d} - \mathsf{E}_{i-1} z_{ki,d} \right| > \kappa \right\} d\kappa.$$

Apply Lemma 5.4 with the parameters λ', ς' and $1/4 < v'$.

The parameter λ'. Given the same set-up, $\lambda' = \lambda$.

The parameter ς'. If $d = K^{-1/2}$, we argue as in (C.5) and use Assumption 4.1 (ib,ii) to get $\mathsf{E} \max_{1\le l\le L} \sum_{i=1}^{n} \mathsf{E}_{i-1} z_{ki,d} = \mathrm{O}(n^{1/2})$ and $\varsigma' = 1/2$. If $d = 0$, the parameter $\varsigma' = 1/2$ follows from the construction of the partition in equation (B.3).

The condition $2v' > \varsigma'$. This is satisfied for our choice of $1/4 < v'$. Therefore, the first term in decomposition (C.6) is $\mathrm{O}(n^{1/2} - \omega)$ with $\omega < 1/4$.

We now know the order of both terms in (C.6). The larger bound determines the order of $\mathcal{E}_{n2} = \mathrm{O}(n^{1/2})$. Thus, for (C.3), we find that $\varsigma = 1/2$.

Condition $2v > \varsigma$. To satisfy this condition, we set $v = 1/2 - \omega$ with $\omega < 1/4$. □

Appendix D: Proofs of auxiliary Theorems 4.2-4.4

While the proof of Theorem 4.2 uses an intricate chaining argument, Theorem 4.3 relies on Lemma 5.1 which was highlighted in section 5. The proof of both theorems proceed by first establishing a similar result for one-sided empirical processes.

D.1 The empirical process result in Theorem 4.2

We prove Theorem 4.2 by first considering the one-sided indicator functions. Define the one-sided empirical distribution function with differenced error terms

$$\widehat{F}_n(a, b, c) = \frac{1}{n} \sum_{i=1}^{n} 1_{(\nabla\varepsilon_i \le \sqrt{2}\sigma c + \sqrt{2} n^{-1/2} ac + \nabla x_{in}' b)}, \tag{D.1}$$

which has the (pseudo-)compensator

$$\overline{F}_n(a, b, c) = \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}_{i-1} 1_{(\nabla\varepsilon_i \le \sqrt{2}\sigma c + \sqrt{2} n^{-1/2} ac + \nabla x_{in}' b)}, \tag{D.2}$$

and which together form the empirical process

$$\mathbb{F}_n(a, b, c) = n^{1/2}\{\widehat{F}_n(a, b, c) - \overline{F}_n(a, b, c)\}.$$

Theorem D.1. *Suppose Assumptions 3.1, 4.1(i,ii,iva) holds. Then, for any $B > 0$, we get*

$$\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\le n^{1/4-\eta}B} |\mathbb{F}_n(a, b, c) - \mathbb{F}_n(0, 0, c)| = \mathrm{o}_{\mathsf{P}}(1).$$

We prove this theorem in two steps. We first set $b = 0$ and introduce Theorem D.2. In a second step we set $a = 0$ and introduce Theorem D.3. The proof of Theorem D.1 then pulls these strings together.

Theorem D.2. *Suppose Assumptions 3.1, 4.1(i) are satisfied. Then, for any $B > 0$, we get*

$$\sup_{c\in\mathbb{R}} \sup_{|a|\le n^{1/4-\eta}B} |\mathbb{F}_n(a, 0, c) - \mathbb{F}_n(0, 0, c)| = \mathrm{o}_{\mathsf{P}}(1).$$

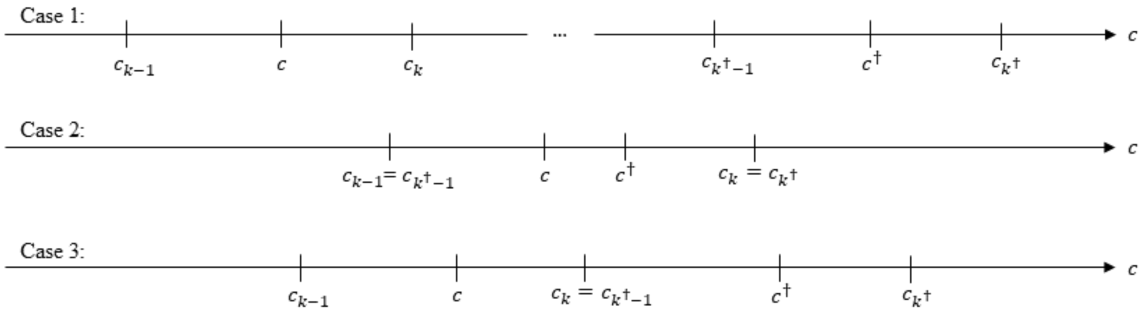


Figure 3: An illustration of the chaining argument

Proof of Theorem D.2. 1. *Notation.* Let $c^\dagger = c(1+\sigma^{-1}n^{-1/2}a)$ and write the term $\mathbb{F}_n(a, 0, c)$ as $\mathbb{F}_n(0, 0, c^\dagger)$. Assume $c < c^\dagger$ and set $\sqrt{2}\sigma = 1$ without loss of generality. We define the term $M_n(c, c^\dagger) = n^{1/2}\{\mathbb{F}_n(0, 0, c^\dagger) - \mathbb{F}_n(0, 0, c)\}$, which can be rewritten in terms of J_i as in (B.1) to be $\sum_{i=1}^n\{J_i(c, c^\dagger) - \mathsf{E}_{i-1}J_i(c, c^\dagger)\}$. Our aim is to prove $M_n = \mathrm{o}_\mathsf{P}(n^{1/2})$ uniformly in c and a.

2. *Decompose $M_n(b, c)$ into martingales.* The term $M_n(c, c^\dagger)$ is not a martingale because ε_{i+1} in $J_i(c, c^\dagger)$ is $\mathcal{F}_{i+1}$ and not $\mathcal{F}_i$-adapted. To transform $M_n(c, c^\dagger)$ to be a martingale, add and subtract $\mathsf{E}_iJ_i(c, c^\dagger)$ to get $M_n(c, c^\dagger) = M_n^1(c, c^\dagger) + M_n^2(c, c^\dagger)$ with

$$M_n^1(c, c^\dagger) = \sum_{i=1}^n\{J_i(c, c^\dagger) - \mathsf{E}_iJ_i(c, c^\dagger)\}, \quad M_n^2(c, c^\dagger) = \sum_{i=1}^n\{\mathsf{E}_iJ_i(c, c^\dagger) - \mathsf{E}_{i-1}J_i(c, c^\dagger)\},$$

and where their respective martingale difference sequences are $\mathcal{F}_{i+1}$ and $\mathcal{F}_i$-adapted. In order to prove $M_n = \mathrm{o}_\mathsf{P}(n^{1/2})$ uniformly in c and a, we then have to show for $q = 1, 2$ that

$$\mathcal{M}_n^q = \sup_{c\in\mathbb{R}}\ \sup_{|a|\le n^{1/4-\eta}B} |M_n^q(c, c^\dagger)| = \mathrm{o}_\mathsf{P}(n^{1/2}).$$

3. *Partition c* such that $c_k = \mathsf{H}^{-1}(k/K)$ for $0 \le k \le K$ with $K = \mathrm{int}(n^{1/2}/\delta)$ and $\delta, n > 0$.

4. *Assign c and $c^\dagger$ to the partitioned support.* For each c and $c^\dagger$ there exist $k \le k^\dagger$ and grid points so that $c_{k-1} < c \le c_k$ and $c_{k^\dagger-1} < c^\dagger \le c_{k^\dagger}$.

5. *Apply chaining to the martingale M_n^1.* Relate c to the nearest right grid point c_k and $c^\dagger$ to the nearest left grid point $c_{k^\dagger-1}$. As illustrated in Figure 3, three distinct cases arise.

- Case 1: We have $J_i(c, c^\dagger) = J_i(c, c_k)+J_i(c_k, c_{k^\dagger-1})+J_i(c_{k^\dagger-1}, c^\dagger)$. By the triangle inequality $|M_n^1(c, c^\dagger)| \le |M_n^1(c, c_k)| + |M_n^1(c_k, c_{k^\dagger-1})| + |M_n^1(c_{k^\dagger-1}, c^\dagger)|$. Define for $j = 1, 2$

$$\mathcal{M}_{an}^j = \max_{0<k<\tilde{k}<K} |M_n^j(c_k, c_{\tilde{k}})|, \tag{D.3}$$

where k and $\tilde{k}$ satisfies $\mathsf{H}(c) \le k/K < \tilde{k}/K \le \mathsf{H}(c^\dagger)$ and define

$$\mathcal{M}_{bn}^j = \max_{1\le k\le K}\ \sup_{c_{k-1}<c\le c_k} |M_n^j(c, c_k)|. \tag{D.4}$$

Then, $\mathcal{M}_{an}^1$ bounds $|M_n^1(c_k, c_{k^\dagger-1})|$ and $\mathcal{M}_{bn}^1$ bounds $|M_n^1(c, c_k)|$ and $|M_n^1(c_{k^\dagger-1}, c^\dagger)|$.

- Case 2: We have $J_i(c, c^\dagger) = J_i(c, c_k) - J_i(c^\dagger, c_{k^\dagger})$. Then, by the triangle inequality $|M_n^1(c, c^\dagger)| \le |M_n^1(c, c_k)| + |M_n^1(c^\dagger, c_{k^\dagger})|$. Both terms are bounded by $\mathcal{M}_{bn}^1$.

- Case 3: We have $J_i(c, c^\dagger) = J_i(c, c_k) + J_i(c_{k^\dagger-1}, c^\dagger)$. Then, by the triangle inequality $|M_n^1(c, c^\dagger)| \le |M_n^1(c, c_k)| + |M_n^1(c_{k^\dagger-1}, c^\dagger)|$. Again, both terms are bounded by $\mathcal{M}_{bn}^1$.

Combine the results from all three cases, and we have $\mathcal{M}^1_n \leq \mathcal{M}^1_{an} + 2\mathcal{M}^1_{bn}$.

6. *The martingale is* $\mathcal{M}^1_{an} = \mathsf{o_P}(n^{1/2})$ by Assumption 4.1(i) and (C.1) in Lemma C.1.

7. *Decompose the martingale* $\mathcal{M}^1_{bn}$. For $c_{k-1} < c \leq c_k$ and $1 \leq k \leq K$, we have the bound $J_i(c, c_k) \leq J_i(c_{k-1}, c_k)$. Therefore $M^1_n(c, c_k) \leq \sum_{i=1}^n \{J_i(c_{k-1}, c_k) + \mathsf{E}_i J_i(c_{k-1}, c_k)\}$. Consequently, $\mathcal{M}^1_{bn} \leq \widetilde{\mathcal{M}}^1_{bn} + 2\overline{\mathcal{M}}^1_{bn}$ where the martingale and the compensator term are

$$\widetilde{\mathcal{M}}^1_{bn} = \max_{1\leq k\leq K} \left|M^1_n(c_{k-1}, c_k)\right|, \quad \overline{\mathcal{M}}^1_{bn} = \max_{1\leq k\leq K} \sum_{i=1}^n \mathsf{E}_i J_i(c_{k-1}, c_k).$$

8. *The martingale is* $\widetilde{\mathcal{M}}^1_{bn} = \mathsf{o_P}(n^{1/2})$ by (C.3) in Lemma C.2 with $b_m = 0\ \forall m$ and $d = 0$.

9. *The compensator is* $\overline{\mathcal{M}}^1_{bn} = \mathsf{o_P}(n^{1/2})$. Add and subtract $\mathsf{E}_i J_i(c_{k-1}, c_k)$ to the sum in $\overline{\mathcal{M}}^1_{bn}$ and apply the triangle inequality to get

$$\overline{\mathcal{M}}^1_{bn} \leq \max_{1\leq k\leq K} \left|M^2_n(c_{k-1}, c_k)\right| + \max_{1\leq k\leq K} \sum_{i=1}^n \mathsf{E}_{i-1} J_i(c_{k-1}, c_k).$$

The first term is $\mathsf{o_P}(n^{1/2})$ by (C.4) in Lemma C.2 with $b_m = 0$ for all m and $d = 0$. For the second term use (B.3) to find $\sum_{i=1}^n \mathsf{E}_{i-1} J_i(c_{k-1}, c_k) = \delta\mathrm{O}(n^{1/2})$. Since $\delta > 0$ can be chosen arbitrarily small, we have $\overline{\mathcal{M}}^1_{bn} = \mathsf{o_P}(n^{1/2})$.

10. *Apply chaining to martingale* M^2_n. Argue as in point 5, to show $\mathcal{M}^2_n \leq \mathcal{M}^2_{an} + 2\mathcal{M}^2_{bn}$.

11. *The martingale is* $\mathcal{M}^2_{an} = \mathsf{o_P}(n^{1/2})$ by Assumption 4.1(ib) and (C.2) in Lemma C.1.

12. *Decompose the martingale* $\mathcal{M}^2_{bn}$. Argue as in point 7 with $\mathcal{M}^2_{bn} \leq \widetilde{\mathcal{M}}^2_{bn} + 2\overline{\mathcal{M}}^2_{bn}$ as

$$\widetilde{\mathcal{M}}^2_{bn} = \max_{1\leq k\leq K} \left|M^2_n(c_{k-1}, c_k)\right|, \quad \overline{\mathcal{M}}^2_{bn} = \max_{1\leq k\leq K} \sum_{i=1}^n \mathsf{E}_{i-1} J_i(c_{k-1}, c_k).$$

13. *The martingale is* $\widetilde{\mathcal{M}}^1_{bn} = \mathsf{o_P}(n^{1/2})$ by (C.4) in Lemma C.2 with $b_m = 0\ \forall m$ and $d = 0$.

14. *The compensator is* $\overline{\mathcal{M}}^1_{bn} = o(n^{1/2})$ by the same argument as in point 9. ☐

Theorem D.3. *Suppose Assumptions 3.1, 4.1(ib,ii,iva) are satisfied. Then, for any $B > 0$, we get*

$$\sup_{c\in\mathbb{R}} \sup_{|b|\leq n^{1/4-\eta}B} |\mathbb{F}_n(0, b, c) - \mathbb{F}_n(0, 0, c)| = \mathsf{o_P}(1).$$

Proof of Theorem D.3. 1. *Notation.* Let $z_i(b, c) = J_i(c, c + \nabla x'_{in} b)$ and let $M_n(b, c) = \sum_{i=1}^n \{z_i(b, c) - \mathsf{E}_{i-1} z_i(b, c)\}$. We want to show $|M_n(b, c)| = \mathsf{o_P}(n^{1/2})$ uniformly in b and c. For $\delta, n > 0$, partition c such that $c_k = \mathsf{H}^{-1}(k/K)$ for $0 \leq k \leq K$ with $K = \mathrm{int}(n^{1/2}/\delta)$. Let c_k be the nearest right grid point to c. We rewrite $z_i(b, c)$ by adding and subtracting $1_{(\nabla\varepsilon_i \leq \sqrt{2}\sigma c_k)}$ to get $z_i(b, c) = z^\dagger_i(b, c, c_k) - z^\dagger_i(0, c, c_k)$, where $z^\dagger_i(b, c, c_k) = J_i(c_k, c + \nabla x'_{in} b)$. Hence, we have $M_n(b, c) = M^\dagger_n(b, c, c_k) - M^\dagger_n(0, c, c_k)$ with $M^\dagger_n(b, c, c_k) = \sum_{i=1}^n \{z^\dagger_i(b, c, c_k) - \mathsf{E}_{i-1} z^\dagger_i(b, c, c_k)\}$. We transform $M^\dagger_n(b, c, c_k)$ into a sum of two martingales by adding and subtracting $\mathsf{E}_i z^\dagger_i(b, c, c_k)$ to get

$$M_n(b, c) = M^{\dagger,1}_n(b, c, c_k) - M^{\dagger,2}_n(b, c, c_k) - M^{\dagger,1}_n(0, c, c_k) + M^{\dagger,2}_n(0, c, c_k),$$

where we first define the term $M^{\dagger,1}_n(b, c, c_k) = \sum_{i=1}^n \{z^\dagger_i(b, c, c_k) - \mathsf{E}_i z^\dagger_i(b, c, c_k)\}$ and then define the second term $M^{\dagger,2}_n(b, c, c_k) = \sum_{i=1}^n \{\mathsf{E}_i z^\dagger_i(b, c, c_k) - \mathsf{E}_{i-1} z^\dagger_i(b, c, c_k)\}$. To show that $|M_n(b, c)| = \mathsf{o_P}(n^{1/2})$ uniformly in b and c, it suffices to show that with $q = 1, 2$,

$$\sup_{|b|\leq n^{1/4-\eta}B} \max_{1\leq k\leq K} \sup_{c_{k-1}\leq c\leq c_k} |M^{\dagger,q}_n(b, c, c_k)|, \quad \max_{1\leq k\leq K} \sup_{c_{k-1}\leq c\leq c_k} |M^{\dagger,q}_n(0, c, c_k)|.$$

are $\mathsf{o_P}(n^{1/2})$. As the first term bounds the second term, we only need to study the first term.

2. *Truncating regressors and martingale decomposition.* By Assumption 4.1(iva), for all $\varepsilon > 0$, $\exists C_x, n_0 > 0$ so that the sets

$$\mathcal{C}_n = \left(\max_{1 \le n \le N} |n^{1/2} x_{in}| \le C_x n^{\kappa} \right), \quad \mathcal{C}_{in} \left(|n^{1/2} x_{in}| \le C_x n^{\kappa} \right),$$

satisfy $\mathsf{P}(\mathcal{C}_n^c < \varepsilon)$ for $n > n_0$, while $\mathcal{C}_n \subseteq \mathcal{C}_{in}$ and $\mathcal{C}_{in}$ is $\mathcal{F}_{i-1}$ adapted. Thus, for $q = 1, 2$, we get that $n^{1/2} |M_n^{\dagger,q}(b, c, c_k)|$ vanishes if $n^{1/2} |M_n^{\dagger,q}(b, c, c_k)| 1_{\mathcal{C}_n}$ vanishes. By the triangle inequality and by $\mathcal{C}_n \subseteq \mathcal{C}_{in}$ we get $|M_n^{\dagger,1}(b, c, c_k)| 1_{\mathcal{C}_n} \le \sum_{i=1}^n |z_i^\dagger(b, c, c_k)| 1_{\mathcal{C}_n} + |\mathsf{E}_i z_i^\dagger(b, c, c_k)| 1_{\mathcal{C}_n}$ and $|M_n^{\dagger,2}(b, c, c_k)| 1_{\mathcal{C}_n} \le \sum_{i=1}^n |\mathsf{E}_i z_i^\dagger(b, c, c_k)| 1_{\mathcal{C}_n} + |\mathsf{E}_{i-1} z_i^\dagger(b, c, c_k)| 1_{\mathcal{C}_n}$.

We bound $|z_i^\dagger(b, c, c_k)| 1_{\mathcal{C}_n}$. First, recall the bound to b, to get on $\mathcal{C}_{in}$,

$$|\nabla x_{in}' b| \le |\nabla x_{in}| |b| \le B n^{1/4 - \eta} C_x n^{\kappa - 1/2} = B C_x n^{\kappa - \eta - 1/4} \le K^{-1/2},$$

where the last inequality holds for large n since $\eta > \kappa$ while $K = \mathrm{int}(n^{1/2}/\delta)$ for fixed δ. Since $c_{k-1} \le c \le c_k$, exploit the truncation on $\mathcal{C}_{in}$ to get that uniformly in b, c we have

$$0 \le |z_i^\dagger(b, c, c_k)| 1_{\mathcal{C}_n} = |J_i(c_k, c + \nabla x_{in}' b)| 1_{\mathcal{C}_n} \le J_i(c_{k-1} - K^{-1/2}, c_k + K^{-1/2}) = z_i^\ddagger(c_{k-1}, c_k).$$

Thus, we can bound $|M_n^{\dagger,1}(b, c, c_k)| 1_{\mathcal{C}_n} \le M_n^{\ddagger,1}(c_k, c_{k-1})$ and $|M_n^{\dagger,2}(b, c, c_k)| 1_{\mathcal{C}_n} \le M_n^{\ddagger,2}(c_k, c_{k-1})$ where we first define the term $M_n^{\ddagger,1}(c_k, c_{k-1}) = \sum_{i=1}^n \{ z_i^\ddagger(c_{k-1}, c_k) + \mathsf{E}_i z_i^\ddagger(c_{k-1}, c_k) \}$ and then define as the second term $M_n^{\ddagger,2}(c_k, c_{k-1}) = \sum_{i=1}^n \{ \mathsf{E}_i z_i^\ddagger(c_{k-1}, c_k) + \mathsf{E}_{i-1} z_i^\ddagger(c_{k-1}, c_k) \}$. Now, for $q = 1, 2$, the term $M_n^{\ddagger,q}$ has martingale decomposition $M_n^{\ddagger,q} = \widetilde{M}_n^{\ddagger,q} + 2\overline{M}_n^{\ddagger,q}$, where

$$\widetilde{M}_n^{\ddagger,1}(c_{k-1}, c_k) = \sum_{i=1}^n \{ z_i^\ddagger(c_{k-1}, c_k) - \mathsf{E}_i z_i^\ddagger(c_{k-1}, c_k) \}, \quad \overline{M}_n^{\ddagger,1}(c_{k-1}, c_k) = \sum_{i=1}^n \mathsf{E}_i z_i^\ddagger(c_{k-1}, c_k),$$

$$\widetilde{M}_n^{\ddagger,2}(c_{k-1}, c_k) = \sum_{i=1}^n \{ \mathsf{E}_i z_i^\ddagger(c_{k-1}, c_k) - \mathsf{E}_{i-1} z_i^\ddagger(c_{k-1}, c_k) \}, \quad \overline{M}_n^{\ddagger,2}(c_{k-1}, c_k) = \sum_{i=1}^n \mathsf{E}_{i-1} z_i^\ddagger(c_{k-1}, c_k).$$

3. *The compensator is* $\max_k \overline{M}_n^{\ddagger,2} = \mathrm{o}(n^{1/2})$. Analyse this term as in (C.5) to get using Assumption 4.1 (ib,ii) that $\max_k \overline{M}_n^{\ddagger,2} = \delta \mathrm{O}(n^{1/2})$. Since $\delta > 0$ can be chosen arbitrarily small, $\max_k \overline{M}_n^{\ddagger,2} = \mathrm{o}(n^{1/2})$.

4. *The martingale is* $\max_k \widetilde{M}_n^{\ddagger,2} = \mathrm{o}_{\mathsf{P}}(n^{1/2})$ by Assumption 4.1(ib,ii), the choice $d = K^{-1/2}$ and equation (C.4) in Lemma C.2.

5. *The compensator is* $\max_k \overline{M}_n^{\ddagger,1} = \mathrm{o}_{\mathsf{P}}(n^{1/2})$. Add and subtract $\mathsf{E}_{i-1} z_i^\ddagger(c, c_k)$ such that

$$\max_{1 \le k \le K} \overline{M}_n^{\ddagger,1}(c_{k-1}, c_k) \le \max_{1 \le k \le K} \widetilde{M}_n^{\ddagger,2}(c_{k-1}, c_k) + \max_{1 \le k \le K} \overline{M}_n^{\ddagger,2}(c_{k-1}, c_k).$$

Both terms vanish as in point 3 and 4 of this argument.

6. *The martingale is* $\max_k \widetilde{M}_n^{\ddagger,1} = \mathrm{o}_{\mathsf{P}}(n^{1/2})$ by Assumption 4.1(ib,ii), the choice $d = K^{-1/2}$ and equation (C.3) in Lemma C.2. □

Proof of Theorem D.1. Let $\mathcal{W} = \mathbb{F}_n(a, b, c) - \mathbb{F}_n(0, 0, c)$. Denote $c^\dagger = c(1 + \sigma^{-1} n^{-1/2} a)$. Use $\mathbb{F}_n(a, b, c) = \mathbb{F}_n(0, b, c^\dagger)$ so it follows that $\mathcal{W} = \mathbb{F}_n(0, b, c^\dagger) - \mathbb{F}_n(0, 0, c)$. Add and subtract $\mathbb{F}_n(a, 0, c) = \mathbb{F}_n(0, 0, c^\dagger)$ and apply the triangle inequality to get

$$|\mathcal{W}| \le |\mathbb{F}_n(0, b, c^\dagger) - \mathbb{F}_n(0, 0, c^\dagger)| + |\mathbb{F}_n(a, 0, c) - \mathbb{F}_n(0, 0, c)|.$$

Thus, the problem reduces to showing

$$\sup_{c \in \mathbb{R}} \sup_{|b| \le n^{1/4 - \eta} B} |\mathbb{F}_n(0, b, c^\dagger) - \mathbb{F}_n(0, 0, c^\dagger)| = \mathrm{o}_{\mathsf{P}}(1),$$

$$\sup_{c \in \mathbb{R}} \sup_{|a| \le n^{1/4 - \eta} B} |\mathbb{F}_n(a, 0, c) - \mathbb{F}_n(0, 0, c)| = \mathrm{o}_{\mathsf{P}}(1).$$

The first equation follows from Theorem D.3 using Assumption 3.1, 4.1 (ib,ii,iva). The second equation follows from Theorem D.2 using Assumption 3.1, 4.1 (i). □

Finally, we prove Theorem 4.2 by decomposing the two-sided indicator functions into two one-sided indicator functions.

Proof of Theorem 4.2. Let $\sqrt{2}\sigma = 1$. Our object of interest is $\mathcal{G} = \mathbb{G}_n(a,b,c) - \mathbb{G}_n(0,0,c)$. We have $|\nabla\epsilon_i| \sim \mathsf{G}$ and $\nabla\epsilon_i \sim \mathsf{H}$. Let $c^* = c + (2n)^{-1/2}ac$. Because

$$1_{(|\nabla\epsilon_i - \nabla x_{in}'b| \le c^*)} = 1_{(\nabla\epsilon_i \le c^* + \nabla x_{in}'b)} - 1_{(\nabla\epsilon_i < -c^* + \nabla x_{in}'b)}$$

and by the definition of the one and two sided empirical process, we have $\mathbb{G}_n(a,b,c) = \mathbb{F}_n(a,b,c) - \lim_{c^\dagger \downarrow c} \mathbb{F}_n(a,b,-c^\dagger)$ for any $c > 0$. Using this and the triangle inequality, we have

$$\mathcal{G} \le |\mathbb{F}_n(a,b,c) - \mathbb{F}_n(0,0,c)| + \lim_{c^\dagger \downarrow c} |\mathbb{F}_n(a,b,-c^\dagger) - \mathbb{F}_n(0,0,-c^\dagger)|.$$

The terms on the right hand side vanish uniformly in a, b, c by Theorem D.1 using Assumption 3.1, 4.1 (i,ii,iva). □

D.2 The bias correction term in Theorem 4.3

As for the proof of Theorem 4.2, we proceed by showing first the one-sided equivalent of Theorem 4.3. Recall the definition of the pseudo compensator $\overline{F}_n$ in (D.2).

Theorem D.4. *Suppose Assumptions 3.1 and 4.1 (ib,iii,ivb) hold. Then, for all $B > 0$, we get*

$$\sup_{c \in \mathbb{R}} \sup_{|a|,|b| \le n^{1/4-\eta}B} \left| n^{1/2}\{\overline{F}_n(a,b,c) - \overline{F}_n(0,0,c)\} - \mathcal{B}_n(a,b,c) \right| = \mathsf{o}_{\mathsf{P}}(n^{-2\eta}),$$

where the bias term, with $c^\dagger = c(1 + n^{-1/2}a/\sigma)$ and $\xi(c)$ as in (20), *is defined as*

$$\mathcal{B}_n(a,b,c) = \frac{ca}{\sigma}\mathsf{h}(c) + \mathsf{h}(c^\dagger)\xi(c^\dagger)b.$$

We prove Theorem D.4 by further considering the two separate cases where $a = 0$ and $b = 0$.

Theorem D.5. *Suppose Assumptions 3.1 and 4.1 (iii,ivb) hold. Then, for all $B > 0$, we get*

$$\sup_{c \in \mathbb{R}} \sup_{|b| \le n^{1/4-\eta}B} \left| n^{1/2}\{\overline{F}_n(0,b,c) - \overline{F}_n(0,0,c)\} - \mathsf{h}(c)\xi(c)b \right| = \mathsf{o}_{\mathsf{P}}(n^{-2\eta}).$$

Proof of Theorem D.5. The object of interest is

$$\mathcal{Q}_n(b,c) = n^{1/2}\{\overline{F}_n(b,c) - \overline{F}_n(0,c)\} - \mathsf{h}(c)\xi(c)b = n^{-1/2}\sum_{i=1}^n \mathcal{P}_i(b,c),$$

where $\mathcal{P}_i(b,c) = \mathsf{E}_{i-1}\{1_{(\nabla\varepsilon_i \le \sqrt{2}\sigma c + \nabla x_{in}'b)} - 1_{(\nabla\varepsilon_i \le \sqrt{2}\sigma c)}\} - \mathsf{h}(c)\xi(c)b$. As the joint density $m_i(y,x)$ is differentiable w.r.t y by Assumption 4.1 (iii), apply Lemma 5.1 to $\mathcal{P}_i(b,c)$ conditional on $\mathcal{F}_{i-1}$ with

$$Y = \frac{\nabla\varepsilon_i}{\sqrt{2}\sigma} \quad \text{and} \quad X = \frac{n^{1/2}\nabla x_{in}'}{\sqrt{2}\sigma}.$$

Note that Y is independent of $\mathcal{F}_{i-1}$. Let the conditional density $\dot{\mathsf{m}}_i$ given $\mathcal{F}_{i-1}$ be defined as in (5) and represent the conditional derivative of the conditional density of $\nabla\varepsilon_i$ given $n^{1/2}\nabla x_{in}$ and conditional on $\mathcal{F}_{i-1}$. We get the bound

$$|\mathcal{P}_i(b,c)| \le 2^{-1}|b|^2 \sup_{y \in \mathbb{R}, x \in \mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)|\, \mathsf{E}_{i-1}|\nabla x_{in}|^2 \le \mathrm{O}(n^{1/2-2\eta}) \sup_{y \in \mathbb{R}, x \in \mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)|\, \mathsf{E}_{i-1}|\nabla x_{in}|^2.$$

Then by the triangular inequality we have

$$|\mathcal{Q}_n(b,c)| \leq \mathrm{O}(n^{-2\eta}) \max_{1\leq i\leq n} \sup_{y\in\mathbb{R},x\in\mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)| \frac{1}{n}\sum_{i=1}^n \mathsf{E}_{i-1}|n^{1/2}\nabla x_{in}|^2.$$

Due to Assumption 4.1 (ivb), $n^{-1}\sum_{i=1}^n \mathsf{E}_{i-1}|n^{1/2}\nabla x_{in}|^2 = \mathrm{O}(1)$ and by Assumption (ii) we have that $\max_{1\leq i\leq n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)| < \infty$. Thus the expression on the left hand side is of order $\mathrm{o}_{\mathsf{P}}(n^{-2\eta})$ uniformly in c and b. □

While Theorem D.5 controlled for variation in b and c, we state in the next paragraphs the theorem to handle variation in a and c.

Theorem D.6. *Suppose Assumption 3.1 and 4.1 (ib) holds. Then, for all $B > 0$, we get*

$$\sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta}B} \left| n^{1/2}\{\overline{F}_n(a,0,c) - \overline{F}_n(0,0,c)\} - \frac{ca}{\sigma}\mathsf{h}(c)\right| = \mathrm{o}_{\mathsf{P}}(n^{-2\eta}).$$

Proof of Theorem D.6. The object of interest is

$$\mathcal{Q}_n(a,c) = n^{1/2}\{\overline{F}_n(a,0,c) - \overline{F}_n(a,0,c)\} - \frac{ca}{\sigma}\mathsf{h}(c).$$

Let $I_i(a,c) = 1_{(\nabla\varepsilon_i \leq \sqrt{2}\sigma c + \sqrt{2}n^{-1/2}ac)} - 1_{(\nabla\varepsilon_i\leq\sqrt{2}\sigma c)}$ and $h(a,c) = (ac)/(n^{-1/2}\sigma)$. Define $\mathcal{P}_i(a,c) = \mathsf{E}_{i-1}I_i(a,c) - h(a,c)\mathsf{h}(c)$ and note that $\nabla\varepsilon_i$ is independent of $\mathcal{F}_{i-1}$. Rewrite the object of interest as $\mathcal{Q}_n(a,c) = n^{-1/2}\sum_{i=1}^n \mathcal{P}_i(a,c)$. Write $\mathcal{P}_i(a,c)$ as an integral

$$\mathcal{P}_i(a,c) = \int_c^{c+h(a,c)} \mathsf{h}(u)du - h(a,c)\mathsf{h}(c).$$

As h is differentiable, we can Taylor expand the integral at c and get $\mathcal{P}_i(a,c) = h^2(a,c)\dot{\mathsf{h}}(\tilde{c})/2$, where $|\tilde{c}-c| \leq |h(a,c)|$. There exists $n_0 > 0$, so for any $n > n_0$ we have $|\sigma^{-1}n^{-1/2}c| \leq 1/2$. We then apply the second inequality in Lemma 1 to obtain $h^2(a,c) \leq 16(n^{-1}a^2\tilde{c}^2\sigma^{-2})$. Exploit the bound $|a| \leq n^{1/4-\eta}B$ to get

$$|\mathcal{P}_i(a,c_\psi)| = \mathrm{O}(n^{-1/2-2\eta})\tilde{c}^2|\dot{\mathsf{h}}(\tilde{c}^2)|.$$

Since $\tilde{c}^2|\dot{\mathsf{h}}(\tilde{c}^2)| \leq \sup_{c\in\mathbb{R}} c^2|\dot{\mathsf{h}}(c)| < \infty$ by Assumption 4.1 (ib) we have $|\mathcal{P}_i(a,c)| = \mathrm{O}(n^{-1/2-2\eta})$ uniformly in ψ, a and i. Then it follows that $|\mathcal{Q}_n(a,c)| = \mathrm{O}(n^{-2\eta})$ uniformly in c and a. □

Proof of Theorem D.4. The interest is $\mathcal{W}_n = n^{1/2}\{\overline{F}_n(a,b,c) - \overline{F}_n(0,0,c)\} - \mathcal{B}_n(a,b,c)$. Notice that $\overline{F}_n(a,b,c) = \overline{F}_n(0,b,c^\dagger)$. Add and subtract $n^{1/2}\overline{F}_n(a,0,c) = n^{1/2}\overline{F}_n(0,0,c^\dagger)$ and apply the triangle inequality to get

$$\begin{aligned}\mathcal{W}_n \leq & \left| n^{-1/2}\{\overline{F}_n(0,b,c^\dagger) - \overline{F}_n(0,0,c^\dagger)\} - \mathsf{h}(c^\dagger)\xi\left(c^\dagger\right)b\right| \\ & + \left| n^{1/2}\{\overline{F}_n(a,0,c) - \overline{F}_n(0,0,c)\} - \frac{ca}{\sigma}\mathsf{h}(c)\right|.\end{aligned}$$

Thus, the problem reduces to showing

$$\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\leq n^{1/4-\eta}B} \left|\overline{F}_n(0,b,c^\dagger) - \overline{F}_n(0,0,c^\dagger) - \mathsf{h}(c^\dagger)\xi\left(c^\dagger\right)b\right| = \mathrm{o}_{\mathsf{P}}(n^{-2\eta}) \tag{D.5}$$

$$\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\leq n^{1/4-\eta}B} \left|\overline{F}_n(a,0,c) - \overline{F}_n(0,0,c) - \frac{ca}{\sigma}\mathsf{h}(c)\right| = \mathrm{o}_{\mathsf{P}}(n^{-2\eta}) \tag{D.6}$$

Then, (D.5) is shown by Theorem D.5 using Assumptions 4.1 (iii,ivb). Further, (D.6) is shown by Theorem D.6 using Assumption 4.1 (ib). □

We can now expand the Theorem D.4 to the absolute empirical processes. In this case we consider the distribution G as the reference distribution to pick our quantile c, as we have absolute differenced error terms.

Proof of Theorem 4.3. The term of interest is $\mathcal{G} = n^{1/2}\{\overline{G}_n(a,b,c) - \overline{G}_n(0,0,c)\} - \mathcal{G}_n(a,b,c)$. Note $|\nabla\varepsilon_i|/(\sqrt{2}\sigma) \sim \mathsf{G}$ and $\nabla\varepsilon_i/(\sqrt{2}\sigma) \sim \mathsf{H}$. We can write the two-sided indicator function as the difference of two one-sided indicator functions as

$$1_{(\nabla\varepsilon_i \le \sqrt{2}\sigma c + \sqrt{2}n^{-1/2}ac + \nabla x'_{in}b)} - 1_{(\nabla\varepsilon_i < -\sqrt{2}\sigma c - \sqrt{2}n^{-1/2}ac + \nabla x'_{in}b)}$$

and we have $\overline{G}_n(a,b,c) = \overline{F}_n(a,b,c) - \lim_{c^\ddagger \downarrow -c}\overline{F}_n(a,b,c^\ddagger)$ for any $c > 0$. By this and the triangle inequality, then,

$$\begin{aligned}|\mathcal{G}| \le& \left| n^{1/2}\{\overline{F}_n(a,b,c) - \overline{F}_n(0,0,c)\} - \frac{ca}{\sigma}\mathsf{h}(c) - \mathsf{h}(c)\xi\left(c + \frac{ac}{n^{1/2}\sigma}\right)\right| \\ &+ \lim_{c^\ddagger \downarrow -c}\left| n^{1/2}\{\overline{F}_n(a,b,c^\ddagger) - \overline{F}_n(0,0,c^\ddagger)\} - \frac{c^\ddagger a}{\sigma}\mathsf{h}(c^\ddagger) - \mathsf{h}(c^\ddagger)\xi\left(c^\ddagger + \frac{ac^\ddagger}{n^{1/2}\sigma}\right)\right|.\end{aligned}$$

In the previous equation, we can replace $c + ac/(n^{1/2}\sigma)$ by c, which relies on continuity of $\xi(c)$. To show continuity, decompose ∇x_i into $\nabla x_i = v_{i-1} + \eta_i$, where v_{i-1} is $\mathcal{F}_{i-1}$ measurable and η_i is independent of $\mathcal{F}_{i-1}$. Then, $\mathsf{E}(\nabla x_i|\nabla\varepsilon_i = \sqrt{2}\sigma c) = v_{i-1} + \mathsf{E}_{i-1}(\eta_i|\nabla\varepsilon_i = \sqrt{2}\sigma c)$. For the latter term we get $\mathsf{E}_{i-1}(\eta_i|\nabla\varepsilon_i = \sqrt{2}\sigma c) = \mathsf{E}(\eta_i|\nabla\varepsilon_i = \sqrt{2}\sigma c) = \int_{\mathbb{R}^p}\eta\mathsf{f}_{\eta|\nabla\varepsilon}(\eta|c)d\eta$, where $\mathsf{f}_{\eta|\nabla\varepsilon} = \mathsf{m}/\mathsf{f}$. The density $\mathsf{f}_{\eta|\nabla\varepsilon}$ is continuous, by Assumption 4.1 (iii) and thus the integral and the conditional expectation is. As $c + ac/(n^{1/2}\sigma) \to c$ for $n \to \infty$, we have by continuity of $\xi(c)$ that $\xi(c + ac/(n^{1/2}\sigma)) \to \xi(c)$.

Due to the symmetry of h, we then get the desired reductions by applying Theorem D.4 with Assumptions 4.1 (ib,iii,ivb). □

D.3 Proof of Theorem 4.4

The idea of the following proof is to transform the problem to format to which our existing Theorems, in particular, Theorem D.2 and D.3 can be applied to.

Proof of Theorem 4.4. Let $\sigma = 1$. Write the empirical distribution function as $n^{1/2}\widehat{\mathsf{H}}_n(a,b,c) = \mathbb{H}_n(a,b,c) + n^{1/2}\overline{H}_n(a,b,c)$ by adding and subtracting the compensator, as defined in (30). Then, it suffices to show for some $\delta_n \to 0$ as $n \to \infty$ that

$$\sup_{|a|,|b|\le n^{1/4-\eta}B}\ \sup_{0\le\psi\le1}\mathbb{H}_n(a,b,c) = \mathsf{o}_\mathsf{P}(1), \qquad \sup_{|a|,|b|\le n^{1/4-\eta}B}\ \sup_{0\le\psi\le1} n^{1/2}\overline{\mathsf{H}}_n(a,b,c) \to 0.$$

Consider first the compensator term. We apply the mean value theorem on the conditional expectation

$$n^{1/2}\overline{\mathsf{H}}_n(a,b,c) = n^{-1/2}\sum_{i=1}^n \mathsf{E}_i 1_{(|\nabla\varepsilon_i - \sqrt{2}c(1+n^{-1/2}a) - \nabla x'_{in}b| \le \sqrt{2}|c|n^{-1/2}\delta_n)} \le 2\sqrt{2}\delta_n|c_\psi|\mathsf{f}(c^*),$$

where $|c^* - \varepsilon_i + \nabla x'_{in}b + \sqrt{2}c(1 + n^{-1/2}a)| \le \sqrt{2}|c|n^{-1/2}\delta_n$. Note that $|X - Y| \le |Z|$ implies $|X| \le |Z| + |Y|$ and let $\tilde{c} = \sqrt{2}c(1 + n^{-1/2}a)$ to get

$$|c^* + \tilde{c}| \le \frac{\sqrt{2}|\tilde{c}|n^{-1/2}\delta_n}{|1 + n^{-1/2}a|} + |\varepsilon_i - \nabla x'_{in}b|.$$

For a sufficiently large n, we have $\sqrt{2}n^{-1/2}\delta_n/|1 + n^{-1/2}a| < 1/2$ and we apply Lemma B.2 to get $|c| \le 2(1 + |c^*| + |\varepsilon_i| + |\nabla x'_{in}b|)$. Use this bound on $|c|$ and get

$$n^{1/2}\overline{\mathsf{H}}_n(a,b,c) \le 4\sqrt{2}\delta_n\sup_{v\in\mathbb{R}}|v|\mathsf{f}(v) + 4\sqrt{2}\delta_n n^{-1}\sum_{i=1}^n\{(|\nabla x'_{in}b| + |\varepsilon_i|)\sup_{v\in\mathbb{R}}\mathsf{f}(v)\}.$$

By Assumption 4.1 (iva), we have $\max_{1\leq i\leq n}|\nabla x_{in}| = \mathsf{o}_\mathsf{P}(n^{\kappa-1/2})$ that is, for all $\epsilon > 0$ there exists a $C_x > 0$ so that the set $(\max_{1\leq i\leq n}|\nabla x'_{in}| \geq C_x n^{\kappa-1/2})$ has probability of less than ϵ.

On the complement of this set, which is $(\max_{1\leq i\leq n}|\nabla x'_{in}| \leq C_x n^{\kappa-1/2})$, we have $n^{-1}\sum_{i=1}^n |\nabla x'_{in} b| = \mathrm{O}(n^{\kappa-\eta-1/4})$ and by $\mathsf{E}|\varepsilon_i| < \infty$ and $\sup_{v\in\mathbb{R}}|v|\mathsf{f}(v) < \infty$, we have with $\delta_n \to 0$ as $n \to \infty$ that $\sup_{|a|,|b|\leq n^{1/4-\eta}B}\sup_{0\leq\psi\leq 1} n^{1/2}\overline{\mathsf{H}}_n(a,b,c_\psi) \to 0$.

Consider next the martingale term. For this part let $\sqrt{2}\sigma = 1$. Rewrite the two sided indicator functions in $\mathbb{H}(a,b,c)$ into two one-sided indicator functions of the form

$$I_i(a,b,c) = 1_{(\nabla\varepsilon_i - \nabla x'_{in}b \leq c(1+2n^{-1/2}(a+\delta_n \mathrm{sign}(c))))} - \mathsf{E}_i 1_{(\nabla\varepsilon_i - \nabla x'_{in}b \leq c(1+2n^{-1/2}(a+\delta_n \mathrm{sign}(c))))}$$

as $n^{-1/2}\sum_{i=1}^n I_i(a+\delta_n\mathrm{sign}(c_\psi),b,c) - \lim_{\delta_n^\dagger\downarrow\delta_n} I_i(a-\delta_n^\dagger\mathrm{sign}(c),b,c)$. Add and subtract $I_i(0,0,c)$ from $\mathbb{H}(a,b,c)$. Then use the triangle inequality to get

$$\mathbb{H}(a,b,c) \leq \left| n^{-1/2}\sum_{i=1}^n \{I_i(a+\delta_n\mathrm{sign}(c),b,c) - I_i(0,0,c)\}\right| + \left| n^{-1/2}\sum_{i=1}^n \left\{ \lim_{\delta_n^\dagger\downarrow\delta_n} I_i(a-\delta_n^\dagger\mathrm{sign}(c),b,c) - I_i(0,0,c)\right\}\right|.$$

Then for large n, we have that δ_n is sufficiently small such that $a + \delta_n\mathrm{sign}(c) \leq 2a$. We define $\mathcal{J}_i(a,b,c) = J_i\{c, c(1+\sqrt{2}n^{-1/2}a) + \nabla x'_{in}b\}$ with the J_i function defined as in (B.1). It suffices to show

$$\sup_{|a|,|b|\leq 2n^{1/4-\eta}B}\sup_{c\in\mathbb{R}}\left| n^{-1/2}\sum_{i=1}^n \{\mathcal{J}_i(a,b,c) - \mathsf{E}_i\mathcal{J}_i(a,b,c)\}\right| = \mathsf{o}_\mathsf{P}(1).$$

Use the argument in Theorem D.1 where we first set $b = 0$ and then $a = 0$ and combine afterwards, such that it suffices to show, with $c^\dagger = c(1+\sqrt{2}n^{-1/2}a)$

$$\mathcal{M}_1 = \sup_{|a|\leq 2n^{1/4-\eta}B}\sup_{c\in\mathbb{R}}\left| n^{-1/2}\sum_{i=1}^n \{J_i(c,c^\dagger) - \mathsf{E}_i J_i(c,c^\dagger)\}\right| = \mathsf{o}_\mathsf{P}(1),$$

$$\mathcal{M}_2 = \sup_{|b|\leq 2n^{1/4-\eta}B}\sup_{c\in\mathbb{R}}\left| n^{-1/2}\sum_{i=1}^n \{J_i(c,c+\nabla x'_{in}b) - \mathsf{E}_i J_i(c,c+\nabla x'_{in}b)\}\right| = \mathsf{o}_\mathsf{P}(1).$$

The first term vanishes as in points 4 to 7 in Theorem D.2, while the second term vanishes as Theorem D.3. □

Appendix E: Proofs of main Theorems 3.3 and 3.4

We first show that the analytical expression of the empirical gauge can be represented in terms of the two-sided empirical distribution function $\widehat{\mathsf{G}}_n$, its compensator $\overline{\mathsf{G}}_n$ and the corresponding empirical process $\mathbb{G}_n$, which were defined in (26), (27) and (28). To do this, we first show that the term for the estimated variance $\hat{\omega}_{j,i}^2$ can be replaced by $\hat{\sigma}_j^2$. We then proceed to apply Theorem 4.2 and 4.3 to derive our results on the asymptotic distribution and convergence of the empirical gauge.

E.1 Absolute empirical process representation of the Gauge

In this section, we decompose the analytical expression for the empirical gauge. We defined the gauge in equation (19) as

$$\hat{\gamma}_n(c) = \frac{1}{n}\sum_{i=1}^n 1_{(|v_i|>c)},$$

where the random variable v_i in the indicator function is defined as

$$v_i = \frac{\nabla y_i - \hat{\beta}_2 \nabla x_i}{\sqrt{2}\hat{\omega}_{2,i}} 1_{(i \in n_1)} + \frac{\nabla y_i - \hat{\beta}_1 \nabla x_i}{\sqrt{2}\hat{\omega}_{1,i}} 1_{(i \in n_2)}$$

and where the estimated variance is $\hat{\omega}_{j,i}^2 = \hat{\sigma}_j^2\{2 + \nabla x_i'(\sum_{k \in I_j} x_k x_k')^{-1} \nabla x_i\}$.

We define the estimation errors $\hat{a}_{j,i} = n_j^{1/2}(\hat{\omega}_{j,i} - \sigma)$, $\hat{a}_j = n_j^{1/2}(\hat{\sigma}_j - \sigma)$ and $\hat{b}_j = N_j^{-1}(\hat{\beta}_j - \beta)$, where j refers to each half-samples. The normalised differenced regressors are $\nabla x_{in} = N_j' \nabla x_i$, where N_j is the normalisation for the respective half sample. Rewrite the empirical gauge in terms of the estimation errors $\hat{a}_{j,i}$ and $\hat{b}_j$. We use

$$\begin{aligned}
\nabla \hat{\varepsilon}_i = \nabla y_i - \hat{\beta}_j \nabla x_i &= \nabla \varepsilon_i - (\hat{\beta}_j - \beta)' \nabla x_i \\
&= \nabla \varepsilon_i - \{N_j^{-1}(\hat{\beta}_j - \beta)\}'(N_j' \nabla x_i) \\
&= \nabla \varepsilon_i - \hat{b}_j' \nabla x_{in}
\end{aligned}$$

and $\sqrt{2}c\hat{\omega}_{j,i} = \sqrt{2}c(\hat{\omega}_{j,i} - \sigma + \sigma) = \sqrt{2}c(n_j^{-1/2}\hat{a}_{j,i} + \sigma)$ to get for the empirical gauge with $I_{j,i}(\hat{a}_{j,i}, \hat{b}_j, c) = 1_{(|\nabla \varepsilon_i - \hat{b}_j' \nabla x_{in}| \leq \sqrt{2}c(\sigma + n_j^{-1/2}\hat{a}_{j,i}))}$ and $\lambda_j = n_j/n$

$$\hat{\gamma}_n(c) = 1 - \frac{\lambda_1}{n_1} \sum_{i \in I_1} I_{1,i}(\hat{a}_{2,i}, \hat{b}_2, c) - \frac{\lambda_2}{n_2} \sum_{i \in I_2} I_{2,i}(\hat{a}_{1,i}, \hat{b}_1, c). \tag{E.1}$$

Note that I_j for $j = 1, 2$ denotes the two subsets of observations created in Algorithm 2.2 and n_j refers to the sample size of the respective subsample. Without loss of generality, consider the normalised first sum of indicator functions

$$\frac{1}{\sqrt{n_1}} \sum_{i \in I_1} 1_{(|\nabla \varepsilon_i - \hat{b}_2' \nabla x_{in}| \leq \sqrt{2}c(\sigma + n_2^{-1/2}\hat{a}_{2,i}))}. \tag{E.2}$$

If the Step-indicator Saturation estimator is implemented by saturating the sample with half-samples of step-indicators, the resulting variance of the step-indicators is of the form $\hat{\omega}_{2,i}^2 = \hat{\sigma}_2^2\{2 + \nabla x_i'(\sum_{k \in I_j} x_k x_k')^{-1} \nabla x_i\}$. The standard error $\hat{a}_{2,i}$ involving the $\hat{\omega}_{2,i}$'s can be handled with Theorem 4.4. By Assumption 3.2 (ivc) we have $\max_{1 \leq i \leq n} n^{1/2}|\hat{\omega}_{2,i} - \hat{\sigma}_2| = o_{\mathsf{P}}(1)$ and by Assumption 3.2 (iva,ivb) we have $|\hat{a}_2|, |\hat{b}_2| = o_{\mathsf{P}}(n^{1/4-\eta})$, so that for all $\epsilon > 0$ there exists a $\delta, B > 0$ such that the event $(\max_{1 \leq i \leq n} |\hat{a}_{2,i} - \hat{a}_2| \leq \delta)$ and $(|\hat{a}_2|, |\hat{b}_2| \leq Bn^{1/4-\eta})$ has probability greater than $1 - \epsilon$. Add and subtract $I_{1,i}(\hat{a}_2, \hat{b}_2, c)$ from (E.2) and use Lemma B.3 with $\max_{1 \leq i \leq n} |\hat{a}_{2,i} - \hat{a}_2| \leq \delta_n$, where $\delta_n \to 0$ as $n \to \infty$, to get

$$\mathcal{H}_{n_1} = \frac{1}{\sqrt{n_1}} \sum_{i \in I_1} \left\{ I_{1,i}(\hat{a}_{2,i}, \hat{b}_2, c) - I_{1,i}(\hat{a}_2, \hat{b}_2, c) \right\} = \sqrt{n_1} \left\{ \widehat{\mathsf{H}}_{n_1}(\hat{a}_2, \hat{b}_2, c) + \widehat{\mathsf{H}}_{n_1}(\hat{a}_2, \hat{b}_2, -c) \right\},$$

where $\widehat{\mathsf{H}}_n(a, b, c)$ is defined as in (29) and where the index n_j signals that only observations of the j's-subsample I_j are used for the sums. Then, on the event $(\max_{1 \leq i \leq n} |\hat{a}_{2,i} - \hat{a}_2| \leq \delta)$ and $(|\hat{a}_2|, |\hat{b}_2| \leq Bn^{1/4-\eta})$ by Theorem 4.4 and for $\delta_n \to 0$ as $n \to \infty$

$$\mathcal{H}_{n_1} \leq \sup_{|a|,|b| \leq n^{1/4-\eta}B} \sup_{0 \leq \psi \leq 1} 2n_1^{1/2} \widehat{\mathsf{H}}_{n_1}(a, b, c_\psi) = o_{\mathsf{P}}(1).$$

Therefore we can consider the estimation errors $\hat{a}_2$ and $\hat{b}_2$ within the indicator functions. Then, rewrite the empirical gauge of (E.1) in terms of the empirical distribution function defined in (26) as

$$\hat{\gamma}_n(c) = 1 - \lambda_1 \widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \lambda_2 \widehat{\mathsf{G}}_{n_2}(\hat{a}_1, \hat{b}_1, c), \tag{E.3}$$

where the index n_j signals that only observations of the j's-subsample I_j are used for the sums.

E.2 Proof of convergence in distribution

Proof of Theorem 3.4. Our object of interest is $G_n = \sqrt{n}\{\hat\gamma_n(c) - \gamma\}$, where $\hat\gamma_n(c)$ is defined as in (E.3). Noting $\psi = 1 - \gamma$, we then have

$$G_n = \sqrt{n}\left\{\psi - \lambda_1 \widehat{G}_{n_1}(\hat a_2, \hat b_2, c) - \lambda_2 \widehat{G}_{n_2}(\hat a_1, \hat b_1, c)\right\},$$

where $\lambda_i = n_i/n$ for $i = 1, 2$. We decompose this term into two components

$$G_n = -\sqrt{\lambda_1} G_{n_1} - \sqrt{\lambda_2} G_{n_2} \tag{E.4}$$

with $G_{n_1} = \sqrt{n_1}\left\{\widehat{G}_{n_1}(\hat a_2, \hat b_2, c) - \psi\right\}$ and $G_{n_2} = \sqrt{n_2}\left\{\widehat{G}_{n_2}(\hat a_1, \hat b_1, c) - \psi\right\}$.

Without loss of generality, consider the first split-sample. Write the empirical gauge of the first split-sample in terms of the empirical distribution function $\widehat{\mathsf{G}}_{n_1}(\hat a_2, \hat b_2, c)$, the compensator $\overline{\mathsf{G}}_{n_1}(\hat a_2, \hat b_2, c)$ and the empirical process $\mathbb{G}_{n_1}(\hat a_2, \hat b_2, c)$ as

$$G_{n_1} \le \sqrt{n_1}\left\{\widehat{\mathsf{G}}_{n_1}(0,0,c) - \psi\right\} + \left|\mathcal{G}_{n_1}(\hat a_2, \hat b_2, c)\right| + \left|\overline{\mathcal{G}}_{n_1}(\hat a_2, \hat b_2, c)\right| + \mathcal{B}_{n_1}(\hat a_2, \hat b_2, c),$$

where the empirical process term is $\mathcal{G}_{n_1}(\hat a_2, \hat b_2, c) = \mathbb{G}_{n_1}(\hat a_2, \hat b_2, c) - \mathbb{G}_{n_1}(0,0,c)$, the compensator term is $\overline{\mathcal{G}}_{n_1}(\hat a_2, \hat b_2, c) = \sqrt{n_1}\left\{\overline{\mathsf{G}}_{n_1}(\hat a_2, \hat b_2, c) - \overline{\mathsf{G}}_{n_1}(0,0,c)\right\} - \mathcal{B}_{n_1}(\hat a_2, \hat b_2, c)$ and the bias term is $\mathcal{B}_{n_1}(\hat a_2, \hat b_2, c) = \mathsf{h}(c)(2c\hat a_2\sigma^{-1} + \xi_c \hat b_2)$ with $\xi_c = \xi(c) - \xi(-c)$ as in (21). As before, the index n_j signals that only observations of the j's-subsample I_j are used for the various sums.

The empirical process term is $\left|\mathcal{G}_{n_1}(\hat a_2, \hat b_2, c)\right| = \mathsf{o}_\mathsf{P}(1)$. Because $(\hat a_2, \hat b_2) = \mathsf{o}_\mathsf{P}(n^{1/4-\eta})$, for all ϵ, there exists a B, so that for large n then $\mathsf{P}\{|(\hat a_2, \hat b_2)| \le n^{1/4-\eta}B\} \ge 1 - \epsilon$. Thus, it suffices to show that $\sup_{|a|,|b|\le n^{1/4-\eta}} |\mathbb{G}_{n_1}(a,b,c) - \mathbb{G}_{n_1}(0,0,c)| = \mathsf{o}_\mathsf{P}(1)$, which follows from 4.2 and Assumptions 3.1 and 4.1(i,ii,iva).

The compensator term is $\left|\overline{\mathcal{G}}_{n_1}(\hat a_2, \hat b_2, c)\right| = \mathsf{o}_\mathsf{P}(1)$. Use Theorem 4.3 with Assumptions 3.1 and 4.1(ib,iii,ivb) and the same truncation argument on the estimation errors $(\hat a_2, \hat b_2)$.

For the bias term $\mathcal{B}_{n_1}(\hat a_2, \hat b_2, c)$, we Taylor expand the estimation errors $\hat a_2, \hat b_2$ such that

$$\sqrt{n_j}(\hat\beta_j - \beta) = \Sigma^{-1}\frac{1}{\sqrt{n_j}}\sum_{i\in I_j} x_i\varepsilon_i + \mathsf{o}_\mathsf{P}(1), \quad \sqrt{n_j}(\hat\sigma_j - \sigma) = \frac{1}{2\sqrt{n_j}}\sum_{i\in I_j}(\varepsilon_i^2/\sigma - \sigma) + \mathsf{o}_\mathsf{P}(1),$$

where $\Sigma = \mathsf{E}(x_i x_i')$. Insert the expansions into the bias term to get

$$\mathcal{B}_{n_1}(\hat a_2, \hat b_2, c) = \frac{\mathsf{h}(c)}{\sqrt{n_2}}\sum_{i\in I_2}\left\{c\left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right) + \xi_c\Sigma^{-1}x_i\varepsilon_i\right\} + \mathsf{o}_\mathsf{P}(1).$$

is n_j normalization sufficient, even for non-stationary regressors?

The term G_n. We plug our subsample expressions for G_{n_1} and G_{n_2} into (E.4) and we study the empirical distribution term and the bias term separately. For the empirical distribution term, we get $-\sqrt{n}\left\{\widehat{G}_n(0,0,c) - \psi\right\}$. For the bias term, we get with h_i defined as in (22) that

$$\sqrt{\lambda_1}\mathcal{B}_{n_1}(\hat a_2, \hat b_2, c) - \sqrt{\lambda_2}\mathcal{B}_{n_2}(\hat a_1, \hat b_1, c) = -\frac{\mathsf{h}(c)}{\sqrt{n}}\sum_{i=1}^n h_i\left\{c\left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right) + \xi_c\Sigma^{-1}x_i\varepsilon_i\right\} + \mathsf{o}_\mathsf{P}(1).$$

□

Proof of Corollary 3.5. We write $S_n = \sum_{i=1}^n X_i$, where

$$X_i = 1_{(|\nabla\varepsilon_i| \le \sqrt{2}\sigma c)} - \gamma + c\mathsf{h}(c)\left(\varepsilon_i^2/\sigma^2 - 1\right)h_i + \xi_c\Sigma^{-1}\mathsf{h}(c)x_i\varepsilon_i h_i.$$

This term consists of a 2-dependent, i.i.d. and m.d.s. term. Note that $n^{1/2}\{\hat{\gamma}_n(c) - \gamma(c)\} = S_n n^{-1/2}$. The sequence $\{X_i\}$ satisfies the condition of a mixingale with ψ_k of size $-1/2$. As we assumed that x_i are stationary and as ε_i are i.i.d., the sequence $\{X_i^2\}$ is uniformly integrable. By Theorem 4.5 and following the dependency structure of X_i, we have with $\min(k,m,n) \to \infty$ that

$$n^{-1}\mathsf{E}\left|\mathsf{E}\{(S_{k+n} - S_k)^2|\mathcal{F}_{k-m}\} - \sigma_X^2\right| \to 0.$$

Therefore we can apply Theorem 4.5 to get the limiting distribution, where we use Lemma B.4 for the covariance terms in $\sigma_X^2 = \mathsf{E}X_1^2 + 2\mathsf{E}(X_1X_2)$. □

The proof of Corollary 3.6 follows along the line of the proof of Corollary 3.5, but simplifies as $\xi_c \to 0$ for random walk and trend stationary regressors. Therefore the m.d.s. component drops out of X_i.

E.3 Proof of mean square convergence of Gauge

Proof of Theorem 3.3. Without loss of generality, consider the first split-sample. Write the empirical gauge of the first split-sample (as in (E.3)) in terms of the empirical distribution function $\widehat{\mathsf{G}}_n(a,b,c)$as

$$\widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c) \leq \widehat{\mathsf{G}}_{n_1}(0,0,c) + \{\widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \widehat{\mathsf{G}}_{n_1}(0,0,c)\}.$$

We argue that $\{\widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \widehat{\mathsf{G}}_{n_1}(0,0,c)\}$ vanishes. Because $(\hat{a}, \hat{b}) = \mathrm{o}_{\mathsf{P}}(n^{1/4-\eta})$, for all $\epsilon > 0$ there exists a $B > 0$ so that for large n then $P\{|(\hat{a}, \hat{b})| \leq n^{1/4-\eta}B\} \geq 1 - \epsilon$. Thus it suffices to show $\sup_{|a|,|b| \leq n^{1/4-\eta}} |\widehat{\mathsf{G}}_{n_1}(a,b,c) - \widehat{\mathsf{G}}_{n_1}(0,0,c)| = \mathrm{o}_{\mathsf{P}}(1)$. This follows, noting $n^{-1/2}\mathcal{B}_n(a,b,c) = \mathsf{h}(c)(2ca\sigma^{-1} + \xi_c b) = \mathrm{o}_{\mathsf{P}}(1)$, from

$$\widehat{\mathsf{G}}_{n_1}(a,b,c) - \widehat{\mathsf{G}}_{n_1}(0,0,c) = n^{-1/2}\{\mathbb{G}_{n_1}(a,b,c) - \mathbb{G}_{n_1}(0,0,c)\} + \overline{G}_{n_1}(a,b,c) - \overline{G}_{n_1}(0,0,c)$$

and Theorem 4.2 and 4.3 with Assumptions 3.1 and 4.1(i), (ii,iii,iv).

Then, split the term $\widehat{\mathsf{G}}_{n_1}(0,0,c)$ with $\lambda^{odd} = n_1^{odd}/n_1$ and $\lambda^{even} = n_1^{even}/n_1$ into

$$\frac{\lambda^{odd}}{n_1^{odd}} \sum_{i \in n_1^{odd}} 1_{(|\nabla\varepsilon_i| \leq \sqrt{2}\sigma c)} + \frac{\lambda^{even}}{n_1^{even}} \sum_{i \in n_1^{even}} 1_{(|\nabla\varepsilon_i| \leq \sqrt{2}\sigma c)}.$$

Apply the weak Law of Large Numbers for i.i.d. random variables so that the terms converge to $\psi = 1 - \gamma$. The same applies to the second split-sample. Therefore, the full sample empirical gauge $\hat{\gamma}_n(c) = 1 - \lambda_1\widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \lambda_2\widehat{\mathsf{G}}_{n_2}(\hat{a}_1, \hat{b}_1, c)$ converges in probability to γ. □

6 References

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1–22.

Bercu, B. & Touati, A. (2008). Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability 18*, 1848–1869.

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

Castle, J.L., Doornik J.A. & Hendry D.F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics 3*, 1–31.

Castle, J.L., Doornik J.A., Hendry D.F. & Pretis F. (2015). Detecting location shifts during model selection by Step-indicator Saturation. *Econometrics 3*, 240–264.

Clements P. & Hendry D.F. (1998). *Forecasting economic time series.* Cambridge: Cambridge University Press.

Doornik, J.A. (2009). Autometrics. In Castle, J.L. and Shephard, N. (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pp. 88–121. Oxford: Oxford University Press.

Doornik, J.A. & Hendry, D.F. (2013). *Empirical Econometric Modelling - PcGive 14, volume 1.*, London: Timberlake Consultants.

Engler, E. & Nielsen, B. (2009). The empirical process of autoregressive residuals. *Econometrics Journal 12*, 367–381.

Freedman, D.A. (1975). On Tail Probabilities for Martingales. *Annals of Probability, 3*, 100–118.

Hansen, B.E. (2001). The new econometrics of structural change: Dating breaks in US labor productivity. *Journal of Economic Perspectives 15*, 117-128.

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*, 341–361. New York: John Wiley.

Hendry D.F. & Johansen, S. & Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics 33*, 317–335. Erratum, 337–339.

Hendry D.F. & Krolzig, H.M. (2005). The properties of automatic Gets modelling. *Economic Journal 115*, C32-C61.

Hendry D.F. & Mizon G. F. (2011). Econometric Modelling of Time Series with Outlying Observations. *Journal of Time Series Econometrics 3, Article 6.*

Hendry, D.F. & Santos, C. (2010). An automatic test of super exogeneity. In Watson, M.W., Bollerslev, T. and Russell, J., (eds.) *In Volatility and Time Series Econometrics*, 164-193. Oxford: Oxford University Press.

Hoover, K.D. & Perez, S.J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *Econometrics Journal 2*, 167–191.

Jiao, X. & Nielsen, B. (2016). Asymptotic analysis of iterated 1-step huber-skip M-estimators with varying cut-offs. In Antoch, J. et al. (eds.) *AMISTAT 2015: Analytical Methods in Statistics, Springer Proceedings in Mathematics & Statistics 193*, 23–52.

Johansen, S. & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J.L. and Shephard, N. (eds.). *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, 1–36. Oxford: Oxford University Press.

Johansen, S. & Nielsen, B. (2013). Asymptotic theory for iterated one-step Huber-skip estimators. *Econometrics 1*, 53–70.

Johansen, S. & Nielsen, B. (2016a). Analysis of the Forward Search using some new results for martingales and empirical processes. *Bernoulli 22*, 1131–1183.

Johansen, S. & Nielsen, B. (2016b). Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics 43*, 321–348.

Koul, H. L. (2012). *Weighted empirical processes in dynamic nonlinear models (Vol. 166)*, Springer Science & Business Media.

Koul, H. L. & Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *The Annals of Statistics*, 540–562.

McLeish, D.L. (1975). Invariance principles for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 32*, 165–178.

Pólya, G. (1923). Herleitung des Gaußschen Fehlergesetzes aus einer Funktionalgleichung. *Mathematische Zeitschrift 18*, 96–108.

Pretis, F., Reade, J. & Sucarrat G. (2016). General-to-specific (GETS) modelling and indicator saturation with the R package gets. To appear in *Journal of Statistical Software.*

Pretis, F., Schneider, L., Smerdon, J. E. & Hendry D. F. (2016). Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-Indicator Saturation *Journal of Economic Surveys 30*, 403–429.