Nuffield College
UNIVERSITY OF OXFORD

ECONOMICS DISCUSSION PAPERS

Paper No: 2023-W03

Asymptotic Properties of the Gauge and Power of Step-Indicator Saturation

By Bent Nielsen and Matthias Qian

November 2023

Asymptotic Properties of the Gauge and Power of Step-Indicator Saturation

Bent Nielsen* & Matthias Qian†

23 November 2023

Abstract

Step-Indicator Saturation (SIS) is an algorithm to address multiple location shifts at unknown dates in time series during model selection. We derive asymptotic theory for tuning parameter choice based on consistency and asymptotic normality of the frequence gauge – the rate of false detections. Simulations suggest that a smaller gauge minimizes bias in post-selection regression estimates. For the small gauge situation, we develop a complementary Poisson theory. We compare the local power of SIS to detect shifts with that of an extant method. We find that SIS excels when breaks are near the sample end or closely spaced. An application to UK labor productivity reveals a growth slowdown after the 2008 financial crisis.

1 Introduction

Step-Indicator Saturation (SIS), suggested by Castle et al. (2015), is a model selection algorithm designed to address location shifts in time series without restrictions on their number, date, and distance to each other or sample boundaries. In its most general form, the initial specification has a k-variate regressor x_i, which can be of the exogenous, (trend-)stationary, or random walk type, and as many step indicators as observations:

$$y_i = \beta' x_i + \sum_{j=1}^{n} \delta_j 1_{(i \leq j)} + \varepsilon_i \qquad \text{for } i = 1, \ldots, n. \tag{1}$$

If the number of $\delta_j \neq 0$ and their location j were small and known, the model could be estimated by least squares. In practice, the nature of location shifts is often unknown,

Comments from David Hendry, Felix Pretis, James Duffy, Jennifer Castle, Jonas Kurle, Jurgen Doornik, Luidas Giraitis, Neil Ericsson, Vanessa Berenguer Rico and Xiyu Jiao are gratefully acknowledged.

*Nuffield College & Department of Economics, University of Oxford. Address for correspondence: Nuffield College, Oxford OX1 1NF, United Kingdom. E-mail: bent.nielsen@nuffield.ox.ac.uk.

†ESMT Berlin. Financial support by the David Walton Memorial Fund (Department of Economics, University of Oxford) and the Studienstiftung des deutschen Volkes are gratefully acknowledged.

and investigators estimate them using block-search algorithms (Doornik, 2009; Hoover & Perez, 1999; Hendry & Krolzig, 2005). Such algorithms depend on a tuning parameter, which can be chosen indirectly by controlling the type I error. Castle et al. (2015) measured type I errors in terms of the frequency of falsely detected shifts, which we will refer to as the gauge. We develop an asymptotic theory to understand the gauge of simplified versions of SIS, and show that for conformable values of the gauge, the procedure maintains power to detect shifts.

Location shifts are a common phenomenon in observed time series (Perron, 1989; Andrews, 1993; Bai & Perron, 1998), and a failure to address them can affect model selection probabilities of variables (Castle & Hendry, 2014), distort parameter estimation (Hendry & Mizon, 2011), and result in forecast failure (Clements & Hendry, 1998). The growing importance of SIS in tackling location shifts is reflected in its applications in fields as varied as economics (Chuffart & Hooper, 2019; Pellini, 2021; Bernstein & Martinez, 2021), climate science (Raggad, 2018; Pretis et al., 2018; Koch et al., 2022; O'Callaghan et al., 2022), and public health (Doornik et al., 2022). However, despite its popularity, no study of its asymptotic properties exists. This study fills the gap using theoretical insights to shed light on four pivotal areas for practitioners: First, the control of its tuning parameter with the gauge can be closely aligned with the investigator's preferences without detailed knowledge of the regressor type. Second, the bias in post-selection regression estimates can be addressed by choosing a small gauge, or switching to the Poisson theory for the gauge when it is vanishing. Third, SIS can detect minor shifts after a short period of upheaval and maintains power near the end of the sample. Fourth, SIS has weak regularity conditions for the regressors.

In this paper, we study the split-half SIS algorithm. This is a simplified version of the SIS algorithm as implemented in tools like EViews (2020), *gets* in R (Pretis et al., 2018; Sucarrat, 2020), and *Autometrics* in Oxmetrics (Doornik, 2009). Simulations by Castle et al. (2015) indicate that the general SIS has the same gauge properties as split-half SIS, but can detect a wider range of shifts with more power. Split-half SIS splits the sample into two subsets with n_1 and $n - n_1$ observations. It then applies stylized SIS to both subsamples. Stylized SIS, when applied to the second subsample, excludes the first set of step-indicators. For example, it is imposed that $\delta_j = 0$ for $j \leq n_1$ in (1). The model is then estimated by OLS to determine which of the coefficients δ_j for $j > n_1$ are significant. An analysis of split-half SIS can shed light on more general versions of the algorithm and provide mathematical tools for examining related algorithms.

Split-half SIS results in n decisions about the inclusion of step-indicators $1_{(i \leq j)}$. This method requires setting a tuning parameter: a common cut-off c for selecting step-indicators. Drawing inspiration from classical test theory, we aim to determine the cut-off c indirectly from a measure of type I error. Classical testing problems focus on single-decision problems in which the critical value – or the cut-off – is chosen from the size of the test, which is the probability of a type I error of falsely rejecting the hypothesis. In multiple-decision problems, there are many alternative ways of measuring type I error. We study the gauge, which is based on a count of the false rejections. The gauge is also referred to as the expected error rate (Miller, 1981) or the per-comparison error rate (Dudoit & van der Laan, 2010). A concept similar to the gauge was introduced by Hoover & Perez (1999). The term gauge originates in Hendry & Santos (2010) and

Castle et al. (2011), see also Hendry & Doornik (2014, p. 122).

When comparing the gauge to alternative measures of type I error in the context of multiple-decision testing problems, we note that the gauge is more amenable to asymptotic analysis. These alternatives include the probability error rate (Miller, 1981; Dudoit & van der Laan, 2010), also called the family-wise error rate (Dudoit & van der Laan, 2010), and the false discover rate (Benjamini & Hochberg, 1995). The probability error rate is the probability of at least one false detection. It requires a detailed assessment of the dependence of the individual decisions. In contrast, the gauge ignores this dependence structure. The false discovery rate is the expected value of the proportion of type I errors among the rejected hypotheses. Under our null hypothesis of no location shift in the data generating process, the false discovery rate equals unity.

To formalize the notion of the gauge, consider two equivalent approaches to formulate stylized SIS. We introduced this algorithm by imposing $\delta_j = 0$ for $j \leq n_1$ in (1), estimating the model by least squares and then investigating the significance of the remaining step-indicators. An equivalent alternative formulation is to first regress y_i on x_i and an intercept for $i \leq n_1$. This yields least squares estimators $\hat{\beta}_1$ and $\hat{\sigma}_1^2$. These estimators will be consistent if there are no location shifts in the first subsample. We then compute the scaled residuals in the second subsample. As pointed out by Castle et al. (2015) and as shown in Section 2, we can then inspect the forward differenced residuals for outliers. That is, if there are n observations of (1), compute

$$(\nabla y_i - \hat{\beta}_1' \nabla x_i)/\sqrt{2}\hat{\sigma}_1 \qquad \text{for } i = n_1 + 1, \dots, n-1, \tag{2}$$

with $\nabla y_i = y_i - y_{i+1}$, and where the $\sqrt{2}$-factor arises since the variance of $\nabla \varepsilon_i$ is twice the variance of ε_i. A location shift is declared if the absolute value of the forward differenced residual exceeds a cut-off, c. The frequency of declared location shifts in the stylized SIS algorithm is the frequence gauge:

$$\hat{\gamma}_n = \frac{1}{n - n_1 - 1} \sum_{i=n_1+1}^{n-1} 1_{(|\nabla y_i - \hat{\beta}_1' \nabla x_i| \geq \sqrt{2}\hat{\sigma}_1 c)}. \tag{3}$$

If the data generating process has no location shifts, then all declarations of shifts are false, so that $\hat{\gamma}_n$ is the average type I error. We show the consistency

$$\hat{\gamma}_n \overset{\mathsf{P}}{\to} \gamma = \mathsf{P}(|\nabla \varepsilon_i| \geq \sqrt{2}\sigma c), \tag{4}$$

for a wide range of time series regressors x_i including stationary and non-stationary regressors. We can then choose the cut-off c from the limiting gauge γ. In simulations, we confirm the consistency result and provide some further analysis.

The consistency of the frequence gauge for a variety of time series regressors shows that it is possible to control the type I error of SIS without prior knowledge of the detailed time series structure. The regressors do have a second-order effect on this consistency result, which we investigate through an asymptotic expansion of the normalized frequence gauge $n^{1/2}(\hat{\gamma}_n - \gamma)$. We find that it is asymptotically zero mean normal, but its variance depends on the correlation structure of ∇x_i and $\nabla \varepsilon_i$. Numerical approximations confirm that the asymptotic variance of the frequence gauge is strictly larger for split-half SIS than for split-half IIS. In contrast to split-half IIS, the asymptotic variance of split-half SIS depends on the temporal persistence of the time series. A small gauge substantially reduces its asymptotic variance.

A challenge to the asymptotic analysis of the frequence gauge for SIS is the temporal and cross-sectional correlation due to the forward differencing of x_i and ε_i in (2). For instance, in the autoregression $x_{i+1} = \rho x_i + \varepsilon_i$ with independent ε_i and x_i, we get that $\nabla\varepsilon_i = \varepsilon_i - \varepsilon_{i+1}$ is temporally and cross-sectionally correlated with $\nabla x_i = x_i - x_{i+1} = (1-\rho)x_i - \varepsilon_i$. In the related asymptotic analysis of IIS by Hendry et al. (2008) and Johansen & Nielsen (2009, 2013, 2016a,b), the use of impulse-indicators of the form $1_{(i=j)}$ avoids the temporal and cross-sectional correlation structure. Therefore, IIS can be analyzed using a version of the empirical process theory of (Koul & Ossiander, 1994), see also Giraitis et al. (2012). Our analysis of the SIS overcomes the correlation problem by combining the empirical process theory with mixingale theory of McLeish (1977).

A simulation study shows that split-half SIS can introduce a bias in the updated estimates for β in (1) that does not vanish asymptotically. The bias is largest when regressors are lagged dependent variables with an autoregressive coefficient close to unity, and when the frequence gauge is large. For split-half SIS, the empirical setting after the selection over the step-indicators resembles an unbalanced panel regression with a small temporal and large cross-sectional range. Each interval in between two consecutive retained step-indicators can be interpreted as another i in the panel that introduces a new individual fixed effect. The incidental parameter problem arises because with a non-zero frequence gauge γ, the number of breaks is approximately $n\gamma$, so that the number of observations in each interval is on average $1/\gamma$ and therefore finite even as the sample size increases. This matches the situation of a panel data model with large cross-sectional dimension and finite time series dimension, in which biases arise for the dynamic parameter estimators. We conjecture that the bias is due to a combination of the incidental parameter problem (Lancaster, 2000, 2002) and the correlation of the retained step-indicators with the innovations (Arellano & Bond, 1991).

We suggest two different approaches address the bias in the estimation of β under split-half SIS. First, simulations suggest investigators to use a small frequence gauge, as a smaller frequence gauge is associated with a smaller bias. In a sample of 100 observations, we would recommend a frequence gauge of 1% if one would normally conduct inference at the 5% level. Second, we develop a theory for shrinking the gauge with increasing sample sizes. For this, we consider the absolute gauge

$$\hat{\Gamma}_n = \sum_{i=n_1+1}^{n-1} 1_{(|\nabla y_i - \hat{\beta}_1' \nabla x_i| \geq \sqrt{2}\hat{\sigma}_1 c_n)}, \tag{5}$$

for increasing sequences of the cut-off c_n that satisfy, for some $\lambda > 0$,

$$\mathsf{P}(|\nabla\varepsilon_i| > \sqrt{2}\sigma c_n) = \lambda/n. \tag{6}$$

As the c_n increases with the sample size, the absolute gauge is smaller than the frequence gauge as the sample grows. By modifying the theory of Johansen & Nielsen (2016b), we show that the absolute gauge $\hat{\Gamma}_n$ is asymptotically Poisson distributed. The asymptotic result is the same whether the regressors are stationary or non-stationary. In the proof, we encounter the same dependence issue between ∇x_i and $\nabla\varepsilon_i$. We address this using the Poisson limit theorem of Chen (1975).

An alternative to SIS is the Bai & Perron (1998) procedure. It builds on the Andrews (1993) breaks test and provides estimates for timing and location of breaks. Comparing the power properties of SIS and the Bai-Perron procedure is challenging due to the inherent complexity of both methods. Instead, we compare stylized SIS with the Andrews test. We consider two types of scenarios: Scenarios where the Andrews test is consistent in that power approaches unity while stylized SIS has trivial power approaching the gauge. And scenarios where the Andrews test has trivial power while stylized SIS is consistent. On balance, we find that the Andrews test is preferable if there is one break or two well-separated breaks in the middle of the sample. SIS is preferable for a break near the end of the sample. Such a break is important to discover and address in forecasting contexts (Clements & Hendry, 1998). SIS is also preferable if two close breaks offset each other, for instance if the growth rate moves from one level to a slightly different level through a short period of upheavel, see Castle et al. (2023) and the empirical illustration. We argue that the results carry over to a comparison with the Bai-Perron procedure.

The proof of the local power results uses convergence on the $D[0, 1]$ space of discontinuous functions. We handle the one-break case by the Skorokhod (1956) J_1-metric discussed by Billingsley (1968). However, in order to establish convergence in the two-close-breaks case, we use Skorokhod's M_1 metric in line with Whitt (2002).

Our theory for simplified versions of SIS requires knowledge of the innovation distribution. The normal distribution is the standard choice. Just as in a standard regression, the normality assumption will be testable from the residuals once the model has been fitted. With a finite cut-off, the standard cumulant based normality test may have to be adjusted. Indeed, this is the case when applying outlier detection with finite cut-off (Berenguer-Rico & Nielsen, 2023). In contrast, standard heteroscedasticity tests remain valid after outlier detection with finite cut-off (Berenguer-Rico & Wilms, 2021). It should be noted that other procedures, such as the Andrews (1993) only require distributional assumptions that are sufficient to apply a Central Limit Theorem. In turn, SIS requires weaker assumptions to the regressors.

We apply our split-half SIS theory to analyze the UK labor productivity from 1980 to 2021. While there is a growing consensus about the decline of productivity growth in the UK (Chadha, 2022), a simple autoregressive model is not rejected by the standard diagnostic tests. This indicates that location shifts are not always obvious to the investigator. Using a 1% gauge, the split-half SIS algorithm identified multiple shifts in UK productivity growth: 0.56% before 2000, 0.37% up to 2008 and 0.04% up to 2020. These findings also illustrate the ability of SIS to find minor shifts around episodes of upheaval, in our case the 2008 financial crisis and the 2020 Covid pandemic.

Section 2 outlines the model and the SIS algorithm. Sections 3, 4 present the asymptotic results on the frequence gauge for the stylized and split-half SIS respectively, while section 5 presents the Poisson theory for the absolute gauge. Power analysis is found in Section 6. Simulation results are given in Section 7. An empirical illustration follows in Section 8. Section 9 concludes. Proofs follow in an appendix.

2 Model and algorithms

We begin by presenting the linear regression model to which we apply the SIS algorithm. Subsequently, we introduce two simplified versions of the SIS: stylized SIS and split-half SIS. Lastly, given that the decisions rules on the retaining of step-indicators pertain to differenced innovations, we discuss of their notable properties.

2.1 The model

Step-Indicator Saturation (SIS) aims to detect location shifts within the model:

$$y_i = \mu + \beta' x_i + \varepsilon_i \qquad \text{for } i = 1, \ldots, n. \tag{7}$$

By saturating with step-indicators of the type $1_{(i \le j)}$, we obtain equation (1) with $\delta_n = \mu$. In practice, one would expect that only a few of the δ_j parameters in (1) are non-zero, but their number and location are unknown. The regressor x_i is a k-vector, which does not include an intercept. It can include stationary, trend-stationary and random walk variables, but excludes explosive regressors. The innovations ε_i are independent, identically, distributed with a continuous distribution that is known up to the scale. Further, the innovations are independent of current and past regressors x_j for $j \le i$. The coefficient to the intercept is identified when $\mathsf{E}\varepsilon_i = 0$, but the asymptotic theory does not depend on this constraint.

As a model selection algorithm, the idea of SIS is grounded in the general-to-specific approach to regressor selection of Hoover & Perez (1999). Its core mechanism revolves around iterative backward elimination: in each step, a regression is estimated, the least significant regressor is eliminated, and the smaller model is re-estimated. The iteration stops when the fit of the model deteriorates too much. While a single backward elimination has poor properties for correlated regressors, Hoover & Perez (1999) found that multiple backward eliminations with different starting points have better properties in recovering the original data generating process. Algorithms such as *PcGets* (Hendry & Krolzig, 2005) and *Autometrics* (Doornik, 2009) adopt this multi-path approach but search over many more paths to get closer to evaluating all possible subsets of regressors. *Autometrics* allows situations with more regressors than observations by searching over blocks of regressors. This permits saturation with indicators for each observation as in Impulse-Indicator Saturation (IIS) and Step-Indicator-Saturation (SIS). The saturation approach allows a simultaneous search over regressors x_i and indicator variables. The simultaneous search is helpful when there is a high sample correlation between regressors and indicator variables, see Hendry & Doornik (2014). SIS is implemented in the R package *gets* (Pretis et al., 2018; Sucarrat, 2020), in EViews (2020), and within a structural time series model in Marczak & Proietti (2016). It is worth noting that *Autometrics* employs indicators of the form $1_{(i \le j)}$ as here, while *gets* utilizes $1_{(i \ge j)}$. A related algorithm based on sensitivity analysis was presented by Becker et al. (2021).

2.2 Split-half estimation and forward differencing

While regression equations with more variables than observations cannot be estimated as a single equation, they can be approached by using a subset – or blocks – of those variables. The strategy involves experimenting with various blocks to find the relevant regressors. The simplest block search algorithm is the *stylized* SIS. We apply it to (1). It begins by dividing the observations into two parts: the first n_1 observations and the remaining $n_2 = n - n_1$ observations. For the first half, we keep only an intercept and otherwise drop the indicator variables. This gives the model equation

$$y_i = \beta' x_i + \mu 1_{(i \le n_1)} + \sum_{j=n_1+1}^{n} \delta_j 1_{(i \le j)} + \varepsilon_i \qquad \text{for } i = 1, \ldots, n. \tag{8}$$

Since the second half-sample is saturated with indicators, that half will have perfect fit. The consequence of this observation is best seen through reparameterization. Multiply x_i by unity, written as a sum of indicators for the first half $(i \le n_1)$ and for the impulses $(i = \ell)$ for $n_1 < \ell \le n$. Decompose the indicator for $(i \le j)$ likewise. This gives

$$y_i = \beta' x_i 1_{(i \le n_1)} + \Big(\mu + \sum_{j=n_1+1}^{n} \delta_j\Big) 1_{(i \le n_1)} + \sum_{\ell=n_1+1}^{n} \beta' x_i 1_{(i=\ell)} + \sum_{j=n_1+1}^{n} \delta_j \sum_{\ell=n_1+1}^{j} 1_{(i=\ell)} + \varepsilon_i.$$

Interchanging summation order in the last δ-term gives the reparameterization

$$y_i = \beta' x_i 1_{(i \le n_1)} + \nu 1_{(i \le n_1)} + \sum_{\ell=n_1+1}^{n} \eta_\ell 1_{(i=\ell)},$$

where

$$\nu = \mu + \sum_{j=n_1+1}^{n} \delta_j, \qquad \eta_\ell = \beta' x_\ell + \sum_{j=\ell}^{n_1} \delta_j. \tag{9}$$

As the indicators are orthogonal, the least squares estimators for β, ν are found by standard multiple regression on x_i and the intercept using the first sample, while η_ℓ is estimated by $\hat{\eta}_\ell = y_\ell$. Solving the expression for η_ℓ in (9) for δ_ℓ shows that

$$\hat{\delta}_\ell = \big(\hat{\eta}_\ell - \hat{\beta}' x_\ell\big) - \big(\hat{\eta}_{\ell+1} - \hat{\beta}' x_{\ell+1}\big) = \nabla y_\ell - \hat{\beta}' \nabla x_\ell \qquad \text{for } n_1 < \ell < n, \tag{10}$$

while $\hat{\delta}_n = \hat{\eta}_n - \hat{\beta}' x_n = y_n - \hat{\beta}' x_n$. In the subsequent analysis, we will analyze the gauge, which is the count of the significant estimated δ_ℓ coefficients. Apart from the last estimate, these are all based on forward differencing. In an asymptotic analysis of the gauge, we can ignore the last estimator without affecting the asymptotic result.

2.3 Step-Indicator Saturation algorithms

We present two simplified SIS algorithms in a more formal way. The algorithms involve splitting the sample in two consecutive parts for the n_1 first observations and the $n_2 = n - n_1$ last observations. When working with differenced variables, one observation is lost from each sub-sample. We define index sets

$$I_1 = (i \le n_1), \quad I_1^\circ = (i < n_1), \quad I_2 = (n_1 < i \le n), \quad I_2^\circ = (n_1 < i < n), \tag{11}$$

and counts $n_1^\circ = n_1 - 1$, $n_2^\circ = n_2 - 1$ and $n^\circ = n_1^\circ + n_2^\circ = n - 2$. For each sub-sample I_j, for $j = 1, 2$, we estimate the constant intercept regression model $y_i = \mu + \beta' x_i + \varepsilon_i$ by least squares regression and get the estimators

$$\overline{x}_j = n_j^{-1} \sum_{i \in I_j} x_i, \qquad \hat{\beta}_j = \Big\{ \sum_{i \in I_j} (x_i - \overline{x}_j)(x_i - \overline{x}_j)' \Big\}^{-1} \sum_{i \in I_j} (x_i - \overline{x}_j) y_i, \tag{12}$$

$$\overline{y}_j = n_j^{-1} \sum_{i \in I_j} y_i, \qquad \hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i \in I_j} \big\{ (y_i - \overline{y}_j) - \hat{\beta}_j'(x_i - \overline{x}_j) \big\}^2. \tag{13}$$

We will use the estimates from the first sub-sample to predict location shifts in the second sub-sample. This corresponds to predicting outliers for the differenced series using $\nabla y_i - \hat{\beta}_1 \nabla x_i$. This gives the forecast correction factors

$$\omega_{1,i}^2 = 1 + (\nabla x_i)' \Big\{ 2 \sum_{k \in I_1^\circ} (x_k - \overline{x}_j)(x_k - \overline{x}_j)' \Big\}^{-1} \nabla x_i \qquad \text{for } i \in I_2, \tag{14}$$

and we define $\omega_{2,i}^2$ vice versa when replacing the index sets I_2°, I_2 by I_1°, I_1. The factors arise as follows. First, rewrite $\nabla y_i - \hat{\beta}_1' \nabla x_i = \nabla \varepsilon_i - (\hat{\beta}_1 - \beta)' \nabla x_i$ by applying equation (7). Then, assuming fixed regressors and independent normal $\mathsf{N}(0, \sigma^2)$ innovations we get that $\nabla y_i - \hat{\beta}_1' \nabla x_i$ is normal $\mathsf{N}(0, 2\sigma^2 \omega_{1,i}^2)$. Later we show that under mild regularity conditions $\omega_{2,i}^2$ is uniformly close to unity and it can indeed be replaced by unity for asymptotic purposes. We define the stylized SIS algorithm, which searches for location shifts in the second sub-sample.

Algorithm 2.1. ***The stylized Step-Indicator Saturation algorithm.***
1. Choose a cut-off value $c > 0$ to select breakpoints.
2. Calculate the least squares estimators $(\hat{\beta}_1, \hat{\sigma}_1^2)$ based on sample I_1.
3. Calculate forecast correction factors $\omega_{1,i}^2$ for $i \in I_2$.
4. Declare a location shift at $i + 1$ if

$$\big| \nabla y_i - \hat{\beta}_1' \nabla x_i \big| \geq \sqrt{2} \hat{\sigma}_1 \omega_{1,i} c \qquad \textit{for } i \in I_2^\circ. \tag{15}$$

The frequency of location shifts declared by Algorithm 2.1 is

$$\hat{\gamma}_n^{stylized} = \frac{1}{n_2^\circ} \sum_{i \in I_2^\circ} 1_{(|\nabla y_i - \hat{\beta}_1' \nabla x_i| \geq \sqrt{2} \hat{\sigma}_1 \omega_{1,i} c)}. \tag{16}$$

When the data generating process has no location shifts, so that $\mu_i = \mu$, the expression $\hat{\gamma}_n$ is the frequence gauge of the algorithm, which is the object of interest in this paper.

Castle et al. (2015) refer to a split-half SIS algorithm, which is a symmetrized version of the above algorithm. For reference, we define that algorithm including a statement on how to update the estimators for β, σ^2 in light of the identified location shifts. We allow the sub-samples to be of unequal size, but retain the split-half descriptor.

Algorithm 2.2. ***The split-half Step-Indicator Saturation algorithm.***
1. Choose a cut-off value $c > 0$ to select breakpoints.
2. Calculate the least squares estimators $(\hat{\beta}_j, \hat{\sigma}_j^2)$ based on sample I_j for $j = 1, 2$.

3. Calculate forecast correction factors $\omega_{j,i}^2$ for $i \notin I_j$ and $j = 1, 2$.
4. Declare a location shift at $i + 1$ if

$$|\nabla y_i - \hat{\beta}_j' \nabla x_i| \geq \sqrt{2}\hat{\sigma}_j \omega_{j,i} c \qquad \text{for } i \in I_{3-j}^{\circ} \text{ and } j = 1, 2. \tag{17}$$

For notational simplicity, we do not consider the possibility of a location shift from $i = n_1$ to $i = n_1 + 1$. The split-half SIS algorithm of Castle et al. (2015) continues to re-estimate β, σ on the full sample while taking the detected location shifts into account.

The frequency of declared location shifts by Algorithm 2.2 is

$$\hat{\gamma}_n^{split} = \frac{1}{n^{\circ}} \Big\{ \sum_{i \in I_1^{\circ}} 1_{(|\nabla y_i - \hat{\beta}_2' \nabla x_i| \geq \sqrt{2}\hat{\sigma}_2 \omega_{2,i} c)} + \sum_{i \in I_2^{\circ}} 1_{(|\nabla y_i - \hat{\beta}_1' \nabla x_i| \geq \sqrt{2}\hat{\sigma}_1 \omega_{1,i} c)} \Big\}. \tag{18}$$

2.4 Properties of the differenced innovations

The scaled innovations ε_i/σ have density f. In applications, we often assume f to be the normal density. The forward differenced innovations are denoted

$$\nabla \varepsilon_i = \varepsilon_i - \varepsilon_{i+1}, \qquad \chi_i = \nabla \varepsilon_i / (\sqrt{2}\sigma). \tag{19}$$

The scaled forward differenced innovations χ_i have the convolution density

$$\mathsf{h}(x) = \sqrt{2} \int_{-\infty}^{\infty} \mathsf{f}(y)\mathsf{f}(\sqrt{2}x + y)\, \mathrm{d}y, \tag{20}$$

and distribution function H. Following (4) let

$$\gamma = \mathsf{P}(|\chi_i| \geq c). \tag{21}$$

We highlight four properties of the density h.

Theorem 2.3. *Assume ε_i/σ are i.i.d. and continuous with density f. The density h then satisfies the following properties:*
(a) Symmetry: $\mathsf{h}(x) = \mathsf{h}(-x)$;
(b) Suppose f has second moment. Then $\mathsf{f} = \mathsf{h}$ if and only if f is standard normal;
(c) for $k \in \mathbb{N}_0$: $\sup_{v \in \mathbb{R}} |v|^k \mathsf{f}(v) < \infty \Rightarrow \sup_{v \in \mathbb{R}} |v|^k \mathsf{h}(v) < \infty$;
(d) for $k \in \mathbb{N}_0$: $\sup_{v \in \mathbb{R}} (1 + |v|^k)|\dot{\mathsf{f}}(v)| < \infty$ and $\mathsf{E}|\varepsilon_i^k| < \infty \Rightarrow \sup_{v \in \mathbb{R}} |v^k \dot{\mathsf{h}}(v)| < \infty$.

Theorem 2.3 implies that when the reference distribution f for ε is standard normal so is the distribution h for χ_i. Thus, the gauge γ is associated with a cut-off c chosen as the normal $(1 - \gamma/2)$ quantile.

3 The main results for stylized SIS

We present an asymptotic theory for the frequence gauge of stylized SIS. The first-order result is consistency. This allows us to choose the cut-off c indirectly from the gauge. We obtain consistency for a wide range of stationary and non-stationary regressors. We will also develop a second-order expansion of the gauge with a view to understand how uniform the consistency result is. In this section, we give an asymptotic expansion, which is developed into an asymptotic theory for split-half SIS in the subsequent section. We then find that the asymptotic distribution is normal for a wide range of regressors, but with an asymptotic variance depending on the type of regressors.

We require the following time series structure for innovations ε_i and regressors x_i.

Assumption 3.1. *Let $\mathcal{F}_i$ be a filtration so that ε_{i-1} and x_i are $\mathcal{F}_{i-1}$-adapted, and ε_i/σ has unit variance and is independent of $\mathcal{F}_{i-1}$ with distribution function F and positive density f on $\mathbb{R}$ with derivative $\dot{\mathsf{f}}$.*

Assumption 3.1 rules out endogeneity of the form, $\mathsf{Cov}(x_i, \varepsilon_i) \neq 0$, but allows predetermined time series regressors. The innovations need not have zero mean as Theorem 2.3(a) implies $\mathsf{E}\nabla\varepsilon_i = 0$ even if $\mathsf{E}\varepsilon_i \neq 0$. Intriguingly, Jiao (2019) exploits the techniques developed here to analyze situations with endogeneity.

The theory results allow for stationary and non-stationary regressors. For this purpose, we introduce normalization matrices N_j for each sub-sample $j = 1, 2$. This yields normalized regressors

$$x_{in} = N_j' x_i, \quad \nabla x_{in} = N_j'(x_i - x_{i+1}) \quad \text{for } i \in I_j^\circ, \tag{22}$$

where we have suppressed the index j in the definition of the normalized regressor x_{in}. We choose the normalizations depending on the stochastic properties of x_i so that

$$\widehat{\Sigma}_{jn} = \sum_{i \in I_j} N_j'(x_i - \bar{x}_j)(x_i - \bar{x}_j)' N_j \quad \text{where} \quad \widehat{\Sigma}_{jn}^{-1} = \mathrm{O}_{\mathsf{P}}(1). \tag{23}$$

In the asymptotic theory we will require that

$$\widehat{V}_{jn} = \sum_{i \in I_j} N_j'(x_i - \bar{x}_j)(\varepsilon_i - \mathsf{E}\varepsilon_i) = \mathrm{O}_{\mathsf{P}}(1); \qquad \mathsf{E}\sum_{i \in I_j^\circ} |\nabla x_{in}|^2 = \mathrm{O}(1). \tag{24}$$

For the practitioner it will be possible to choose the cut-off c without precise knowledge of the type of regressors and hence the normalization. The knowledge of the type is only needed for the second-order theory.

We give some examples of normalizations. If x_i is stationary, then ∇x_i is also stationary. Thus, we let $N_j = n_j^{-1/2} I_{\dim x}$ and find that $\widehat{\Sigma}_{1n}$, $\widehat{V}_{1n}$ and $\mathsf{E}\sum_{i \in I_2^\circ} |\nabla x_{in}|^2$ converge under mild regularity conditions. If x_i is a random walk, then ∇x_i is i.i.d. and we let $N_j = n_j^{-1} I_{\dim x}$. Then, under mild regularity conditions, $\widehat{\Sigma}_{1n}$, $\widehat{V}_{1n}$ converge, while $\mathsf{E}\sum_{i \in I_2^\circ} |\nabla x_{in}|^2$ vanishes. Thus, the asymptotic expansions simplify in the latter case. As an example of cointegrated regressors, we could have

$$N_1 = \begin{pmatrix} n^{-1/2} & 0 \\ 0 & n^{-1} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad \text{if} \quad x_i = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \sum_{j=1}^{i-1} \varepsilon_j + z_i$$

for some stationary, bivariate process z_i. We note that this N_1 is non-diagonal.

In most applications, the density of the innovations ε_i will be normal. However, the density needs neither be centered at zero nor be symmetric as the theory results will only depend on the implied density for the differenced innovations $\nabla\varepsilon_i = \varepsilon_i - \varepsilon_{i+1}$.

Our theory does require that the density f of the innovations ε_i and its derivative are bounded. The condition is satisfied for a wide range of densities, including the normal density. Moreover, the differenced innovations' conditional density, given the differenced regressors and the past, should also be bounded. If the regressors are pre-determined, this reduces to boundedness of the density of the differenced innovations and follows from the boundedness of the density f of the innovations ε_i due to Theorem 2.3.

Assumption 3.2. *Suppose that*
(i) the density f *satisfies (a)* $\sup_{v\in\mathbb{R}} \mathsf{f}(v) < \infty$, *(b)* $\sup_{v\in\mathbb{R}}(1+v^2)|\dot{\mathsf{f}}(v)| < \infty$*;*
(ii) the conditional density $\mathsf{m}_i(y|x)$ *of* χ_i *given* ∇x_i *and* $\mathcal{F}_{i-1}$ *exists for* $i = n_1+1,\ldots,n$, *it is differentiable in* y *and satisfies* $\max_{n_1+1\le i\le n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}(1+|y|)\,|\dot{\mathsf{m}}_i(y|x)| < \infty$*;*
(iii) the regressors x_i *satisfy, with* $\widehat{\Sigma}_{1n}$, $\widehat{V}_{1n}$ *defined in (23), (24):*
(a) $\widehat{\Sigma}_{1n}^{-1} = \mathrm{O}_{\mathsf{P}}(1)$, *(b)* $\widehat{V}_{1n} = \mathrm{O}_{\mathsf{P}}(1)$, *(c)* $\mathsf{E}\sum_{i\in I_2^\circ}|\nabla x_{in}|^2 = \mathrm{O}(1)$*;*
(iv) the sub-sample lengths satisfy $(n_2/n_1)^{1/2}$, $N_2^{-1}N_1 = \mathrm{o}(n_2^{1/4-\eta})$ *for some* $\eta > 0$.

We start by showing that the forecast correction factor $\omega_{1,i}^2$ can be replaced by unity with negligible asymptotic consequences.

Theorem 3.3. *Consider the gauge of the stylized SIS Algorithm 2.1. Suppose Assumptions 3.1, 3.2(ia, iii, iv) apply and that no locations shifts are present so that* $\mu_i = \mu$. *Then, we get for fixed* $c \in \mathbb{R}$ *that*

$$\hat{\gamma}_n^{stylized} = \frac{1}{n_2^\circ}\sum_{i\in I_2^\circ} 1_{(|\nabla y_i - \hat{\beta}_1'\nabla x_i| \ge \sqrt{2}\hat{\sigma}_1 c)} = \frac{1}{n_2^\circ}\sum_{i\in I_2^\circ} 1_{(|\nabla y_i - \hat{\beta}_1'\nabla x_i| \ge \sqrt{2}\hat{\sigma}_1\omega_{1,i} c)} + \mathrm{o}_{\mathsf{P}}(n_2^{-1/2}).$$

The next result presents the expansion for the frequence gauge $\hat{\gamma}_n^{stylized}$ of stylized SIS as defined in (16) around the population gauge $\gamma = \mathsf{P}(|\chi_i| \ge c) = \mathsf{P}(|\nabla\varepsilon_i| \ge \sqrt{2}\sigma c)$. The data generating process is assumed to have no location shifts.

Theorem 3.4. *Consider the gauge of the stylized SIS Algorithm 2.1. Suppose Assumptions 3.1, 3.2 apply and that no locations shifts are present so that* $\mu_i = \mu$. *Let*

$$\xi_{2n}(c) = n_2^{-1/2}\sum_{i\in I_2^\circ}\mathsf{E}_{i-1}(\nabla x_{in} \mid \chi_i = c) = \mathrm{O}_{\mathsf{P}}(1).$$

Then, we get for fixed $c \in \mathbb{R}$ *that*

$$\begin{aligned} n_2^{1/2}(\hat{\gamma}_n^{stylized} - \gamma) \;=\;& n_2^{-1/2}\sum_{i\in I_2^\circ}\{1_{(|\chi_i|\ge c)} - \mathsf{E}1_{(|\chi_i|\ge c)}\} \\ & - c\mathsf{h}(c)(n_2/n_1)^{1/2}n_1^{-1/2}\sum_{i\in I_1}(\varepsilon_i^2/\sigma^2 - 1) \\ & - \mathsf{h}(c)(\sqrt{2}\sigma)^{-1}\{\xi_{2n}(c) - \xi_{2n}(-c)\}'N_2^{-1}N_1\widehat{\Sigma}_{1n}^{-1}\widehat{V}_{1n} + \mathrm{o}_{\mathsf{P}}(1). \end{aligned} \quad (25)$$

Finally, $\hat{\gamma}_n^{stylized}$ *is consistent in that* $\hat{\gamma}_n^{stylized} \to \gamma$ *in probability and in mean.*

The consistency statement in Theorem 3.4 for the stylized SIS algorithm is nuisance parameter-free. It can be used for calibrating the SIS algorithm. The result provides the rationale for choosing c to match the desired population gauge γ: We specify our tolerance for false positives expressed by γ. Given the innovation density f we obtain a selection quantile c. For example, if the innovations ε_i are normal, then the forward differenced innovations χ_i are standard normal by Theorem 2.3. If the sample is $n = 100$ and $\gamma = 1\%$, we choose $c = 2.58$, which is the normal 99.5% quantile.

The expansion in Theorem 3.4 has three components. The first component is a binomial term. The next two components relate to the estimation uncertainty from the initial estimation. They involve factors n_2/n_1 and $N_2^{-1}N_1$, respectively, where $N_2^{-1}N_1$ is an increasing function of n_2/n_1. These factors are allowed to diverge at an $\mathrm{o}(n^{1/4-\eta})$ rate. This means that the expansion would apply if we choose, in a stationary context, $n_1 = n^{7/8}$ and $n_2 = n - n_1$, so that $n_2/n_1 = \mathrm{O}(n^{1/8})$, which requires that $\eta < 1/8$ in Assumption 3.2(iv). In other words, the length of the sub-sample used for the initial estimation may be of a lower order of magnitude than the sub-sample used to search for location shifts. This feature is implicitly exploited in more complicated versions of the algorithm, which search for small sub-sets of observations without location shifts.

The third term in the Theorem 3.4 expansion involves the nuisance quantity $\xi_{2n}(c)$. It vanishes in two distinct cases. First, if the regressors are strictly exogenous, then $\mathsf{E}_{i-1}(\nabla x_i \mid \chi_i = c) = \mathsf{E}_{i-1}\nabla x_i$ does not depend on c so that $\xi_{2n}(c)-\xi_{2n}(-c) = 0$. Second, for random walk type regressors with stationary ∇x_i the normalization is $N_2 = n^{-1}$ so that $\xi_{2n}(c)$ vanishes. The third term simplifies if the sequence $\nabla x_i, \chi_i$ is stationary. In this case, we let $N_2 = n_2^{-1/2}$ and get $\xi_{2n}(c) = n_2^{-1}\sum_{i\in I_2^\circ} \mathsf{E}_{i-1}(\nabla x_i|\chi_i = c)$. Under regularity conditions, this converges in probability to $\mathsf{E}\mathsf{E}_0(\nabla x_1|\chi_1 = c) = \mathsf{E}(\nabla x_1|\chi_1 = c)$. Under a normality assumption, this can be computed explicitly. Thus, suppose that $(\nabla x_1, \chi_1)$ are normal given $\mathcal{F}_0$ with conditional mean $(v_0, 0)$, where v_0 is $\mathcal{F}_0$-measurable with expectation $\mathsf{E}v_0 = 0$. Noting that χ_1 has unit variance, we have

$$\begin{pmatrix}\nabla x_1\\ \chi_1\end{pmatrix} \mid \mathcal{F}_0 \overset{\mathsf{D}}{=} \mathsf{N}\left\{\begin{pmatrix}v_0\\ 0\end{pmatrix}, \begin{pmatrix}\sigma_{\nabla\nabla} & \sigma_{\nabla\chi}\\ \sigma_{\chi\nabla} & 1\end{pmatrix}\right\}. \tag{26}$$

Then, $\xi_{2n}(c) \to \mathsf{E}v_0 + c\sigma_{\nabla\chi} = c\sigma_{\nabla\chi}$ in probability, while $\xi_{2n}(c) - \xi_{2n}(-c) \to 2c\sigma_{\nabla\chi}$. For example, in the autoregression $y_i = \mu + \alpha y_{i-1} + \varepsilon_i$ so that $x_i = y_{i-1}$, we find that $\sigma_{\nabla\chi} = \mathsf{E}_0(x_1 - x_2)(\varepsilon_1 - \varepsilon_2)/(\sqrt{2}\sigma) = \mathsf{E}_0(y_0 - y_1)(\varepsilon_1 - \varepsilon_2)/(\sqrt{2}\sigma) = -\sigma/\sqrt{2}$.

In the statement of Theorem 3.4, the initial least squares estimation is based on observations with indices $I_1 = (i \leq n_1)$, while the search for location shifts is based on observations with indices $I_2 = (i > n_1)$. The consecutive nature of the sets I_1 and I_2 are convenient in the proof to simplify notation. However, the result extends to situations where the sets I_1 and I_2 are more complicated. Indeed, this is possible because Theorem 3.4 is derived under the hypothesis of no location shifts. It would be possible to choose I_1 as all odd and I_2 as all even indices. In that case, all observations will be involved when computing the forward differences arising from the set I_2.

4 The main results for split-half SIS

We provide an asymptotic expansion for split-half SIS and analyze the asymptotic distribution of the frequence gauge for stationary and for random walk regressors.

4.1 Expansion of the gauge for split-half SIS

We expand the gauge for split-half SIS by applying Theorem 3.4 to each sub-sample. This requires a symmetrized version of Assumption 3.2.

Assumption 4.1. *Suppose that*
(i) the density f *satisfies* $\sup_{v\in\mathbb{R}}\mathsf{f}(v)<\infty$, $\sup_{v\in\mathbb{R}}(1+v^2)|\dot{\mathsf{f}}(v)|<\infty$*;*
(ii) the conditional density $\mathsf{m}_i(y|x)$ *of* χ_i *given* ∇x_i *and* $\mathcal{F}_{i-1}$ *exists for* $i=1,\dots,n$*, it is differentiable in* y *and satisfies* $\max_{1\le i\le n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}(1+|y|)\,|\dot{\mathsf{m}}_i(y|x)|<\infty$*;*
(iii) the regressors x_i *satisfy for* $j=1,2$*, with* $\widehat{\Sigma}_{jn}$*,* $\widehat{V}_{jn}$ *defined in* (23), (24)*:*
(a) $\widehat{\Sigma}_{1n}^{-1}=\mathrm{O}_{\mathsf{P}}(1)$*, (b)* $\widehat{V}_{jn}=\mathrm{O}_{\mathsf{P}}(1)$*, and (c)* $\mathsf{E}\sum_{i\in I_j^\circ}|\nabla x_{in}|^2=\mathrm{O}(1)$*;*
(iv) the sub-sample lengths satisfy $(n_2/n_1)^{1/2}$*,* $N_2^{-1}N_1=\mathrm{o}(n_2^{1/4-\eta})$*, and* $(n_1/n_2)^{1/2}$*,* $N_1^{-1}N_2=\mathrm{o}(n_1^{1/4-\eta})$ *for some* $\eta>0$*.*

Theorem 4.2. *Consider the gauge of the split-half SIS Algorithm 2.2. Suppose Assumptions 3.1, 4.1 apply and that no locations shifts are present so that* $\mu_i=\mu$*. Let* $\xi_{jn}(c)=n_j^{-1/2}\sum_{i\in I_j}\mathsf{E}_{i-1}(N_j'\nabla x_i\mid\chi_i=c)$ *for* $j=1,2$*. Then, we get for fixed* $c\in\mathbb{R}$ *that*

$$\begin{aligned}\sqrt{n}(\hat{\gamma}_n^{split}-\gamma)&=n^{-1/2}\sum_{i=1}^{n-1}\big\{1_{(|\chi_i|\ge c)}-\mathsf{E}1_{(|\chi_i|\ge c)}\big\}\\&\quad-c\mathsf{h}(c)n^{-1/2}\sum_{i=1}^{n}\big\{n_2{n_1}^{-1}1_{(i\in I_1)}+n_1{n_2}^{-1}1_{(i\in I_2)}\big\}(\varepsilon_i^2\sigma^{-2}-1)\\&\quad-\mathsf{h}(c)(\sqrt{2}\sigma)^{-1}\big[(n_1/n)^{1/2}\{\xi_{1n}(c)-\xi_{1n}(-c)\}'N_1^{-1}N_2\widehat{\Sigma}_{2n}^{-1}\widehat{V}_{2n}\\&\qquad+(n_2/n)^{1/2}\{\xi_{2n}(c)-\xi_{2n}(-c)\}'N_2^{-1}N_1\widehat{\Sigma}_{1n}^{-1}\widehat{V}_{1n}\big]+\mathrm{o}_{\mathsf{P}}(1).\end{aligned}$$

Finally, $\hat{\gamma}_n^{split}$ *is consistent in that* $\hat{\gamma}_n^{split}\to\gamma$ *in probability and in mean.*

Once again, the consistency statement in Theorem 4.2 for the split-half SIS algorithm is nuisance parameter-free.

4.2 Asymptotic distribution in the stationary case

We now consider the expansion of split-half SIS when the regressors x_j are stationary. We start by introducing some notations for various moments for the innovations ε_i:

$$\varkappa_1=\mathsf{E}\varepsilon_i/\sigma,\qquad\varkappa_2=\mathsf{E}\varepsilon_i^2/\sigma^2=1,\qquad\varkappa_4=\mathsf{E}\varepsilon_i^4/\sigma^4,\tag{27}$$

$$\varsigma_0=\mathsf{E}\{1_{(|\chi_i|\ge c)}1_{(|\chi_{i+1}|\ge c)}\},\tag{28}$$

$$\varsigma_2=\mathsf{E}\{1_{(|\chi_i|\ge c)}(\varepsilon_{i+1}^2/\sigma^2-1)\}=\mathsf{E}\{1_{(|\chi_i|\ge c)}(\varepsilon_i^2/\sigma^2-1)\}.\tag{29}$$

Further, for the stationary regressor x_i, we denote

$$\mu_x=\mathsf{E}x_i,\qquad\Sigma_x=\mathsf{Var}x_i,\tag{30}$$

and finally, for a cross moment for innovations and regressors, we denote

$$\varsigma_{1x}=\mathsf{E}\{\nabla x_i(1_{(|\chi_i|\ge c)}-\gamma)(\varepsilon_i/\sigma-\varkappa_1)\},\tag{31}$$

$$\xi_c=\mathsf{E}(\nabla x_i\mid\chi_i=c)=\mathsf{E}(\nabla x_i\mid\chi_i=-c).\tag{32}$$

Then, the vector $s_i = \{1_{(|\chi_i|\geq s)} - \gamma, \varepsilon_i^2/\sigma^2 - 1, (\varepsilon_i/\sigma - \varkappa_1)(x_i - \mu_x)'\Sigma_x^{-1}\}'$ has variance and first-order autocovariance of the form

$$\Omega_0 = \begin{pmatrix} \gamma(1-\gamma) & \varsigma_2 & 0 \\ \varsigma_2 & \varkappa_4 - 1 & 0 \\ 0 & 0 & \Sigma_x^{-1}(1-\varkappa_1^2) \end{pmatrix}, \qquad \Omega_1 = \begin{pmatrix} \varsigma_0 - \gamma^2 & \varsigma_2 & \varsigma_{1x}'\Sigma_x^{-1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{33}$$

Finally, we define long-run variances for the summands of the frequence gauge in (18). Let (j,k) be $(1,2)$ or $(2,1)$ and define with $n_j/n \to \lambda_j > 0$ for $j = 1, 2$

$$d_j = \begin{pmatrix} 1 \\ -c\mathsf{h}(c)(\lambda_k/\lambda_j) \\ -\mathsf{h}(c)\xi_c(\lambda_k/\lambda_j)/\sqrt{2} \end{pmatrix}, \qquad \omega_j^2 = d_j'\Omega_0 d_j + 2e_1'\Omega_1 d_j. \tag{34}$$

The long-run variances ω_j^2 will be assumed to be positive in order to exploit the Functional Central Limit Theorem for non-stationary mixingales in McLeish (1977).

Example 4.1. *If ε_i/σ has standard normal density φ and distribution function Φ, then $\mathsf{h}(x) = \varphi(x)$, while $\varkappa_1 = 0$ and $\varkappa_4 = 3$. It is argued in Appendix A.9 that*

$$\varsigma_0 = 2\gamma - 4\{T(c, 1/\sqrt{3}) + T(c, \sqrt{3})\}, \qquad \varsigma_2 = c\varphi(c). \tag{35}$$

Here, $T(c,a) = \int_c^\infty \varphi(x)\int_0^{ax}\varphi(y)dydx$ following Owen (1980, 2.2; 2.8). In particular, $T(c,a)$ is positive and decreasing in c with $T(0, 1/\sqrt{3}) = 1/12$ and $T(0,\sqrt{3}) = 1/6$. Finally, if $\nabla x_1, \chi_1$ are jointly normal given $\mathcal{F}_0$ as in (26) then $\xi_c = 2c\sigma_{\nabla\chi}$.

Assumption 4.3. *Suppose*
(i) the density f satisfies $\sup_{v\in\mathbb{R}} |v|\mathsf{f}(v) < \infty$ and $\int_{\mathbb{R}} v^{4+}\mathsf{f}(v)dv < \infty$;
(ii) the pairs x_i, ε_i are stationary with $\mathsf{E}|x_i^{2+}| < \infty$;
(iii) $\omega_1^2, \omega_2^2 > 0$;
(iv) let z_i be either of x_i, $x_i x_i'$ or $\nabla x_i 1_{(|\chi_i|\geq c)}(\varepsilon_i/\sigma - \varkappa_1)$ and suppose
$$\mathsf{E}|\mathsf{E}_{k-m} n^{-1}\textstyle\sum_{i=k+1}^{k+n}(z_i - \mathsf{E}z_i)| \to 0 \text{ as } \min(k,m,n)\to\infty;$$
(v) $n^{-1}\sum_{i=1}^n x_i = \mu_x + \mathrm{o}_\mathsf{P}(1)$.

Theorem 4.4. *Consider the gauge of the split-half SIS Algorithm 2.2 with $n_j/n \to \lambda_j > 0$ for $j = 1, 2$, so that $\lambda_1 + \lambda_2 = 1$. Suppose Assumptions 3.1, 4.1, 4.3 apply and that no locations shifts are present so that $\mu_i = \mu$. Then, for fixed $c \in \mathbb{R}$, we get $n^{1/2}(\hat{\gamma}_n^{split} - \gamma) \overset{\mathsf{D}}{\to} \mathsf{N}(0, B)$, where*

$$\begin{aligned} B &= \lambda_1\omega_1^2 + \lambda_2\omega_2^2 \\ &= \gamma(1-\gamma) + 2(\varsigma_0 - \gamma^2) - 4c\mathsf{h}(c)\varsigma_2 - \sqrt{2}\mathsf{h}(c)\varsigma_{1x}'\Sigma_x^{-1}\xi_c \\ &\quad + (\lambda_1^2/\lambda_2 + \lambda_2^2/\lambda_1)\mathsf{h}^2(c)\{c^2(\varkappa_4 - 1) + (1-\varkappa_1^2)\xi_c'\Sigma_x^{-1}\xi_c/2\}. \end{aligned} \tag{36}$$

Example 4.2. *Let $y_i = \mu + \alpha y_{i-1} + \varepsilon_i$ be stationary so that $|\alpha| < 1$ and ε_i/σ is standard normal. Then $x_i = y_{i-1}$ has mean $\mu_x = \mu/(1-\alpha)$ and variance $\Sigma_x = \sigma^2/(1-\alpha^2)$. It is argued in Appendix A.9 that $\sigma_{\nabla\chi} = -\sigma/\sqrt{2}$ in (26), that $\varsigma_{1x} = -\sigma\varsigma_2$ and that condition (iv) of Assumption 4.3 holds.*

Example 4.3. *We consider the asymptotic distribution of the gauge for standard normally distributed error terms ε_i/σ, so that $\varkappa_1 = 0$ and $\varkappa_4 = 3$. Further, assume that the sample size in the two sub-samples is equal and that the regressors x_i are stationary. The asymptotic variance (36) in Theorem 4.4 then reduces to*

$$\begin{aligned} B &= \gamma(1-\gamma) + 2(\varsigma_0 - \gamma^2) - 4c\mathsf{h}(c)\varsigma_2 - \sqrt{2}\mathsf{h}(c)\varsigma_{1x}'\Sigma_x^{-1}\xi_c \\ &\quad + \mathsf{h}^2(c)(2c^2 + \xi_c'\Sigma_x^{-1}\xi_c/2). \end{aligned} \tag{37}$$

Recall that if, in addition, $\nabla x_1, \chi_1$ are normal given F_0 as in (26) then $\xi_c = 2c\sigma_{\nabla\chi}$. Further, in a first-order autoregression $y_i = \mu + \alpha y_{i-1} + \varepsilon_i$ the conditional covariance $\sigma_{\nabla\chi}$ equals $-\sigma/\sqrt{2}$, while $\varsigma_{1x} = -\varsigma_2$ and $\Sigma_x = \mathsf{Var}x_i = \sigma^2/(1-\alpha^2)$.

4.3 Distribution of split-half SIS when ξ_{nj} vanishes

The Theorem 4.2 expansion for the split-half SIS's frequence gauge simplifies when the term ξ_{nj} vanishes so that the third term in the expansion falls away. As remarked after Theorem 3.4, this happens for strictly exogenous or random walk regressors. The limiting long-run variance simplifies so that

$$\tilde{\omega}_j^2 = \gamma(1-\gamma) + 2(\varsigma_0 - \gamma^2) - 4c\mathsf{h}(c)(\lambda_k/\lambda_j)\varsigma_2 + c^2\mathsf{h}^2(c)(\lambda_k/\lambda_j)^2(\varkappa_4 - 1). \tag{38}$$

We will require that $\tilde{\omega}_j^2 > 0$.

Theorem 4.5. *Consider the gauge of the split-half SIS Algorithm 2.2 with $n_j/n \to \lambda_j > 0$ for $j = 1, 2$, so that $\lambda_1 + \lambda_2 = 1$. Let $\xi_{jn} = \mathsf{o}_{\mathsf{P}}(1)$. Suppose Assumptions 3.1, 4.1, 4.3(i) apply, $\tilde{\omega}_j^2 > 0$ for $j = 1, 2$ and that no locations shifts are present so that $\mu_i = \mu$. Then, for fixed $c \in \mathbb{R}$, we get $n^{1/2}(\hat{\gamma}_n^{split} - \gamma) \overset{\mathsf{D}}{\to} \mathsf{N}(0, \tilde{B})$, where*

$$\begin{aligned} \tilde{B} &= \lambda_1\tilde{\omega}_1^2 + \lambda_2\tilde{\omega}_2^2 \\ &= \gamma(1-\gamma) + 2(\varsigma_0 - \gamma^2) - 4c\mathsf{h}(c)\varsigma_2 + c^2\mathsf{h}^2(c)(\lambda_1^2/\lambda_2 + \lambda_2^2/\lambda_1)(\varkappa_4 - 1). \end{aligned} \tag{39}$$

5 Poisson approximation

We present a theory for a vanishing frequence gauge. We set the cut-off so as to control the absolute gauge, the number of falsely discovered outliers. This gives a Poisson exceedance theory. For stylized and split-half SIS, the absolute gauges are

$$\hat{\Gamma}_n^{stylized} = \sum_{i \in I_2^\circ} 1_{(|\nabla y_i - \hat{\beta}_1'\nabla x_i| \geq \sqrt{2}\hat{\sigma}_1\omega_{1,i}c_n)}, \tag{40}$$

$$\hat{\Gamma}_n^{split} = \sum_{i \in I_2^\circ} 1_{(|\nabla y_i - \hat{\beta}_1'\nabla x_i| \geq \sqrt{2}\hat{\sigma}_1\omega_{1,i}c_n)} + \sum_{i \in I_1^\circ} 1_{(|\nabla y_i - \hat{\beta}_2'\nabla x_i| \geq \sqrt{2}\hat{\sigma}_2\omega_{2,i}c_n)}. \tag{41}$$

Here, we choose the cut-off c_n so that, for some $\lambda > 0$,

$$\mathsf{P}(|\nabla\varepsilon_i| > \sqrt{2}\sigma c_n) = \mathsf{P}(|\chi_i| > c_n) = \lambda/n. \tag{42}$$

The analysis builds on the Poisson exceedance theory for IIS, (Johansen & Nielsen, 2016b). The analysis has two part. The first part is a Poisson limit theorem for the case without estimation errors. For IIS, the standard Poisson limit theorem could be used. For SIS, we have that the forward differenced innovations are 1-dependent. We can then apply the Chen (1975) Poisson limit theorem. The second part is an argument that the estimation errors do not matter for the asymptotic theory. This argument is similar to that of the IIS analysis. For the analysis, we need the following high-level assumptions.

Assumption 5.1. *Suppose that*
(i) the innovations ε_i are i.i.d., so that $\chi_i = \nabla\varepsilon_i/(\sqrt{2}\sigma)$ has continuous distribution function H with density h satisfying
(a) $\mathsf{E}|\chi|^r < \infty$ for some $r > 4$;
(b) $\mathsf{h}(c_n)/[c_n\{1 - \mathsf{H}(c_n)\}] = \mathrm{O}(1)$;
(c) $\mathsf{h}(c_n - n^{-1/4}A)/\mathsf{h}(c_n) = \mathrm{O}(1)$ for all $A > 0$;
(d) given $\lambda > 0$ choose c_n so that for all i then $\mathsf{P}(|\chi_i| > c_n) = \lambda/n$ and suppose $n\{\mathsf{E}1_{(|\chi_i|>c_n)}1_{(|\chi_{i+1}|>c_n)}\} \to 0$;
(ii) the regressors x_i satisfy, with $j = 1, 2$ and $\widehat{\Sigma}_{jn}$, $\widehat{V}_{jn}$ defined in (23), (24):
(a) $\widehat{\Sigma}_{jn}^{-1} = \mathrm{O}_{\mathsf{P}}(1)$, (b) $\widehat{V}_{jn} = \mathrm{O}_{\mathsf{P}}(1)$, (c) $\mathsf{E}\sum_{i\in I_j^\circ}|\nabla x_{in}|^4 = \mathrm{O}(n^{-1})$;
(iii) the sub-sample lengths satisfy $N_2^{-1}N_1$, $N_1^{-1}N_2 = \mathrm{O}_{\mathsf{P}}(1)$.

Remark 5.1. *Assumption 5.1(i) is satisfied when the innovations ε_i are normal. For parts (a)-(c), this follows from Lemma A.14 in the Appendix A.10. For part (d), this follows from Lemma A.13.*

Theorem 5.2. *Suppose Assumption 5.1, that $n_2/n \to \psi$ for $0 < \psi < 1$ and that the cut-off is chosen through $\mathsf{P}(|\chi_i| > c_n) = \lambda/n$ for all i. Then,*
(a) $\hat{\Gamma}_n^{stylized} = \sum_{i\in I_2^\circ} 1_{(|\chi_i|>\sigma c_n)} + \mathrm{o}_{\mathsf{P}}(1) \overset{\mathsf{D}}{\to} \mathsf{Poisson}(\lambda\psi)$;
(b) $\hat{\Gamma}_n^{split} = \sum_{i=1}^{n-1} 1_{(|\chi_i|>\sigma c_n)} + \mathrm{o}_{\mathsf{P}}(1) \overset{\mathsf{D}}{\to} \mathsf{Poisson}(\lambda)$.

The Poisson result shows that the absolute gauge is not consistent for the target λ. Rather, it has a Poisson variation around the target. The asymptotic Poisson variation depends neither on the regressors nor on the estimation error.

6 Power

We consider local power for stylized SIS and for the Andrews (1993) test and argue that the results carry over to the Bai & Perron (1998) procedure. Proofs are given in Appendix A.11.

6.1 Power of stylized SIS

The power properties of the SIS algorithm are discussed by Castle et al. (2015). We give further discussion for the stylized SIS algorithm. For simplicity, we focus on the case without regressors, so the model in (8) reduces to

$$y_i = \sigma\mu 1_{(i\le n_1)} + \sum_{j=n_1+1}^{n} \sigma\delta_j 1_{(i\le j)} + \varepsilon_i \qquad \text{for } i = 1, \dots, n, \tag{43}$$

with independent, normal $\mathsf{N}(0,\sigma^2)$ innovations and where μ and δ_j are reparameterized using the scale σ. This data generating process allows up to $n-n_1-1$ breaks.

The stylized SIS Algorithm 2.1 estimates the error variance from the first sample-half and used forward differences throughout the second sample-half to detect location shifts, see §2.2. Thus, stylized SIS declares step-shifts for any observation in the second sample half, $n_1<i<n$, if

$$\big|\nabla y_i\big| \geq \sqrt{2}\hat{\sigma}_1 c. \tag{44}$$

Theorem 3.4 analyzes the gauge of the procedure. Under normality, we choose the cut-off from the equation $\gamma = 2\{1-\Phi(c)\}$, see (21); e.g. $\gamma = 1\%$ corresponds to $c = 2.58$.

By the temporal independence, then y_i for $i>n_1$ is independent of the variance estimator $\hat{\sigma}_1^2$, which is asymptotically $\sigma^2\chi^2_{n_1-1}/(n_1-1)$-distributed. Assuming also normality, then the t-statistics defined from (44) are non-central t-distributed (Johnson et al., 1993). We note that for an index i in the second sample-half, then (43) can be written as $y_i=\sum_{j=i}^n \sigma\delta_j+\varepsilon_i$. Thus, we find with $\chi_i=\nabla\varepsilon_i/(\sqrt{2}\sigma)$ that

$$z_i = \frac{\nabla y_i}{\sqrt{2}\hat{\sigma}_1} = \frac{\chi_i+\delta_i/\sqrt{2}}{\hat{\sigma}_1/\sigma} \overset{\mathsf{D}}{=} \mathsf{t}_{n_1-1}\Big(\frac{\delta_i}{\sqrt{2}}\Big) \qquad \text{for } n_1<i<n. \tag{45}$$

A single step-shift at time $\tau+1$ of size δ comes about in model (43) if $\mu=\delta_\tau=-\delta$, with δ_n indicating the post-break level, while all other δ_i are zero. If χ represents a standard normal variable then the power to detect such a shift is

$$\begin{aligned}\mathsf{P}\big\{|z_i|>c\big\} &= \mathsf{P}\big\{\mathsf{t}_{n_1-1}\big(-\delta/\sqrt{2}\big)\big\}\\ &\to \mathsf{P}\big(\big|\chi-\delta/\sqrt{2}\big|>c\big) = \Phi\big(-c+\delta/\sqrt{2}\big)+\Phi\big(-c-\delta/\sqrt{2}\big).\end{aligned} \tag{46}$$

We learn a number of properties from this result. First, the power does not depend on the sign of the shift. Second, the power of the difference decision rule (44) is invariant to time τ. The power stays the same even in the boundaries of the sample. Third, the t-tests are *only* consistent, i.e. approach unit power, when $|\delta|$ is increasing. Fourth, two decisions are dependent if they concern consecutive time periods. Otherwise, they are independent. Thus, the power is invariant to the number, magnitude, and timing of other shifts as long as they are at least two periods away. SIS can detect shifts, even if their number is large. Fifth, a slight location shift can be detected with high probability if the two episodes are separated by a short period of upheaval. For analytic simplicity, this short period is at least two periods long. Thus, suppose there is one level until τ, a location shift of size δ at $\tau+1$ followed by an opposite location shift of size $\nu-\delta$ at $\tau+3$, to a new level that is ν larger than the first level and where ν may be small. In terms of the model (43) this comes about through $\mu=\delta_\tau=-\delta$ and $\delta_{\tau+2}=\delta-\nu$ while δ_n gives the post-break level. The joint probability of correct detection is

$$\begin{aligned}\mathsf{P}\big\{|z_\tau|>c,|z_{\tau+2}|>c\big\} &\to \big\{\Phi\big(-c+\delta/\sqrt{2}\big)+\Phi\big(-c-\delta/\sqrt{2}\big)\big\}\\ &\times\big[\Phi\big\{-c+(\delta-\nu)/\sqrt{2}\big\}+\Phi\big\{-c-(\delta-\nu)/\sqrt{2}\big\}\big].\end{aligned} \tag{47}$$

Thus, for large n and small ν, we find

$$\mathsf{P}\big\{|z_\tau|>c,|z_{\tau+2}|>c\big\} = \big\{\Phi\big(-c+\delta/\sqrt{2}\big)+\Phi\big(-c-\delta/\sqrt{2}\big)\big\}^2+\mathrm{O}(\nu).$$

As a consequence, a small location shift can be discovered consistently, if the upheavel δ is large. Once it has been established that there is, for instance, a shift of this type and no other shifts, it can be tested whether $\nu = 0$. This test will be consistent for finite ν.

While the fifth case may seem contrived, it occurs empirically. Castle et al. (2023) find that the UK annual real wage growth rate increases from 0.8% prior to World War II to 1.7% after the war, with a large impulse during the war. Similarly, the UK annual productivity per worker increases from 1.2% prior to World War I to 1.7% after a huge deflation episode in the wake of the war. Such changes have profound implications for the economy, even if they are small relative to the residual standard error.

6.2 Local power for Andrews test

We consider the Andrews (1993) test for a single break at an unknown time in the central part of the sample. This test is consistent for a shift of fixed magnitude that is not at the ends of the sample. We consider local power for various alternatives. The test is based on the simple one-shift model

$$y_i = \sigma\mu + \sigma\delta 1_{(i\leq\tau)} + \varepsilon_i \qquad \text{for } i = 1, \dots, n,$$

with independent, normal $\mathsf{N}(0, \sigma^2)$ innovations. If the break point is known, we can form the t-statistic, Z_τ say, for the hypothesis $\delta = 0$, see (A.28) for a detailed expression. For the case of an unknown break point, τ, we may suppose $\underline{n} \leq \tau \leq \overline{n}$ for some user-chosen bounds satisfying $0 < \underline{n} \leq \overline{n} < n$. The likelihood ratio test is then formed by maximizing the squared t-statistic over location. This gives the test statistic

$$LR_{\max} = \max_{\underline{n}\leq t\leq\overline{n}} Z_t^2. \tag{48}$$

Distribution under hypothesis. Critical values are found from the distribution of the test statistic under the hypothesis of no break. There are two relevant limits. We note two differences to stylized SIS. On the one hand, the Andrews test asymptotics applies for unknown error distributions while SIS requires a known error distribution. On the other hand, the Andrews test generalizes to the case of stationary regressors, but in contrast to SIS, it does not generalize to the case of non-stationary regressors.

First, when there are no restrictions on the search range, so that $1 = \underline{n}$ and $\overline{n} = n-1$, then the likelihood ratio statistic diverges at a rate of $2\log\log n$ due to the behavior of a Brownian motion near the origin as described by the law of iterated logarithms. With an appropriate logarithmic normalization, the statistic converges to an extreme value distribution (Yao & Davis, 1986; Hidalgo & Seo, 2013). This test is not so common. Perhaps because it is felt that too much power is lost by the additional normalization.

Second, when the search range is trimmed, the likelihood ratio statistic converges to a supremum of a standardized Brownian bridge (Andrews, 1993). That is, if $\mathbb{B}_u$ is a standard Brownian bridge for $0 \leq u \leq 1$, which has variance $u(1-u)$, then for large n and with $\underline{n}/n \to \underline{\lambda} > 0$ and $\overline{n}/n \to \overline{\lambda} < 1$, we get

$$LR_{\max} = \max_{\underline{n}\leq t\leq\overline{n}} Z_t^2 \overset{\mathsf{D}}{\to} \sup_{\underline{\lambda}\leq u\leq\overline{\lambda}} \frac{\mathbb{B}_u^2}{u(1-u)}.$$

The critical values increase with decreasing trimming, reaching the extreme value asymptotics when there is no trimming. Andrews provided simulated critical values. A 15% trimming is commonly used with critical value 12.35 for a 1% sized test. Bai & Perron (1998) preferred 5% trimming. The test is known to be consistent for a central break of finite magnitude δ. This contrasts with SIS. We investigate local power in various cases.

A single break. We consider the power against an alternative with a shift of vanishing magnitude at time $\tau = \lambda n$. We allow $0 < \lambda < 1$, while noting that the Andrews test is aimed at the trimmed interval $0 < \underline{\lambda} \leq \lambda \leq \overline{\lambda} < 1$. Local power is found when the magnitude of the break vanishes as $\delta = \phi/\sqrt{n}$ for fixed ϕ. We find in Appendix A.11 that, for fixed $0 < \lambda < 1$,

$$LR_{\max} \overset{\mathsf{D}}{\to} \sup_{\underline{\lambda} \leq u \leq \overline{\lambda}} \frac{(\mathbb{B}_u + \phi s_u^\lambda)^2}{u(1-u)}, \tag{49}$$

where the function s_u^λ increases linearly from 0 at $u = 0$ to $\lambda(1-\lambda)$ at $u = \lambda$ after which it decreases linearly to 0 at $u = 1$ as given by

$$s_u^\lambda = (1-\lambda)u1_{(u \leq \lambda)} + \lambda(1-u)1_{(u > \lambda)}. \tag{50}$$

The non-centrality term is largest for $u = \lambda$, taking the value $\phi\{\lambda(1-\lambda)\}^{1/2}$. Thus, the Andrews test has local power for this alternative, whereas asymptotically, stylized SIS has trivial power. For a finite sample, we compare the maximal pointwise non-centrality for the Andrews test of $\phi\{\lambda(1-\lambda)\}^{1/2} = \delta\{n\lambda(1-\lambda)\}^{1/2}$ with the SIS non-centrality of $\delta/\sqrt{2}$ arising from (45). Notably, the magnitude of the break δ will give neither method an advantage in the power comparison. Instead, the positioning λ and the sample size n determine the comparative performance. We compare the two non-centralities, while ignoring the simultaneity of decisions within the two procedures. The Andrews test with 15% trimming and 1% size has critical value $12.35 = (3.51)^2$, while stylized SIS has 1% critical value $2.57 = (6.63)^{1/2}$. Dividing the non-centralities with 3.51 and 2.57, respectively, equating and solving gives $n = (12.35/6.63)/\{2\lambda(1-\lambda)\}$, with SIS being advantageous for n smaller than those values. The implied n-values for central values $\lambda = (0.5, 0.75, 0.85)$ are $n = (4, 5, 7)$ so that the Andrews test is favourable. However, this changes when the break occurs in the trimmed period. The largest u considered by the test statistic is $\overline{\lambda}$, so that the Andrews test has maximal pointwise non-centrality of $\delta\{n\overline{\lambda}/(1-\overline{\lambda})\}^{1/2}(1-\lambda)$. Proceeding as before, we find $n = (12.35/6.63)(1/2)\{(1-\overline{\lambda})/\overline{\lambda}\}/(1-\lambda)^2$. Thus, for $\overline{\lambda} = 0.85$, the implied n-values for $\lambda = (0.9, 0.95, 0.99)$ are $n = (16, 66, 1650)$. The comparison indicates that stylized SIS may be competitive in small samples with a late break.

Next, consider the consequence of a break close to the sample boundaries. The above derivation can be modified to the case where $\delta(1-\tau/n) = \psi/\sqrt{n}$ while $\tau/n \to 1$ and fixed ψ. These constraints imply $|\delta|/\sqrt{n} \leq |\psi|$ with equality when $\tau = n-1$. Thus, we let $\delta/\sqrt{n} \to \eta$ where $0 \leq |\eta| \leq |\psi|$ while $\eta\psi \geq 0$. For large n, we get

$$LR_{\max} \overset{\mathsf{D}}{\to} \sup_{\underline{\lambda} \leq u \leq \overline{\lambda}} \frac{(\mathbb{B}_u + \psi u)^2}{u(1-u)(1+\eta\psi)} \tag{51}$$

To see that (51) conforms with (49), note that $u \leq \overline{\lambda} < 1$ and $\tau/n \to 1$ imply that $u < \tau/n$ so that $s_u^{\tau/n} = (1-\tau/n)u$ for large n, while a small δ corresponds to $\eta = 0$.

The result (51) shows that when δ diverges, then the Andrews test has local power, while the stylized SIS is consistent, see (45). In particular, we can let δ diverge a slow rate with τ sufficiently close to n to achieve $\psi = 0$, so that the Andrews test has trivial power, while stylized SIS is consistent.

Two breaks. Let $y_i = \sigma\mu + \sigma\delta_1 1_{(i \leq \tau_1)} + \sigma\delta_2 1_{(i \leq \tau_2)} + \varepsilon_i$ where ε_i is i.i.d. $\mathsf{N}(0, \sigma^2)$ so that the level is changed twice at $\tau_1 < \tau_2$. Again, this alternative is outside those the Andrews test is optimized against, but relevant in practice. We consider the situation where two large location shifts are close and nearly offset each other so that $\tau_2 - \tau_1$ and $\delta_1 + \delta_2$ are close to zero. This is an empirically relevant situation where SIS performs well. Thus, we investigate local power when $\delta_1 + \delta_2 = \xi/\sqrt{n}$ and $\delta_2(\tau_2 - \tau_1) = \psi\sqrt{n}$ while $\tau_1/n = \lambda$ and $(\tau_2 - \tau_1)/n \to 0$ for fixed ξ, ψ, λ. These constraints imply $|\delta_2|/\sqrt{n} \leq |\psi|$ with equality when $\tau_2 = \tau_1 + 1$. Thus, we let $\delta_2/\sqrt{n} \to \eta$ where $0 \leq |\eta| \leq |\psi|$ while $\eta\psi \geq 0$. We find in Appendix A.11 using the Skorokhod (1956) M_1-metric that

$$LR_{\max} \overset{\mathsf{D}}{\to} \sup_{\underline{\lambda} \leq u \leq \overline{\lambda}} \frac{\left[\mathbb{B}_u + \xi s_u^\lambda + \psi\{1_{(u \geq \lambda)} - u\}\right]^2}{u(1-u)(1+\eta\psi)}. \tag{52}$$

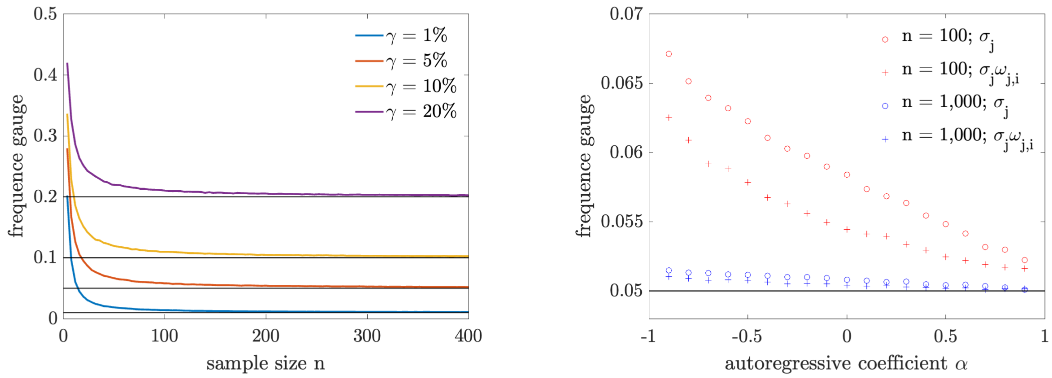
Again, when δ_2 and hence δ_1 diverge, then the Andrews test has local power while split-half SIS is consistent. In particular, when δ_2 diverges slowly while τ_2 and τ_1 are so close that $\psi = 0$, then the Andrews test has trivial power while stylized SIS is consistent.

6.3 Discussion of Bai and Perron procedure

We first summarize the findings for the Andrews test. This test is consistent for a fixed-sized central break in contrast with stylized SIS which only has local power in that situation. Otherwise, SIS can be competitive. We found that SIS is consistent, while the Andrews test has trivial power in two situations. The first case has a break near the end point of the sample. Detecting such a break is highly relevant when forecasting (Clements & Hendry, 1998). The second case is when two breaks are close and nearly offsetting. This can reveal small but important changes in, for instance, growth series (Castle et al., 2023). Thus, the Andrews test is preferable if one is content that there is only one central break or perhaps two well-separated central breaks. With more complicated series, SIS will be competitive.

The Bai & Perron (1998) (BP) procedure is developed for the situation where there is an unknown, but bounded, number of multiple well-separated breaks. This procedure provides estimates of the number of breaks and their timing. This requires trimming between breaks and at the end points of the sample and a maximal number of breaks. The usual 15% trimming eliminates too much of the sample and a 5% trimming is recommended. The above analysis suggests that the BP procedure will consistently detect fixed-sized breaks that are not too close. But, with many breaks or with close breaks, the BP procedure may have near trivial power, while SIS could have high power.

As a further point of comparison, we note that the BP procedure allows an unknown error distribution and it generalizes to stationary, but not non-stationary regressors. The SIS procedure requires a known error distribution, but allows both stationary and non-stationary regressors. We note that for many macro-economic time series, normality is not unreasonable, but assuming stationarity of the regressors may not be appropriate.



(a) DGP1 with varying γ and n

(b) DGP2 with varying α and n

Figure 1: Finite sample properties of the frequence gauge

Finally, the general SIS algorithm is designed to work jointly with regression selection, whereas the BP procedure requires a fixed set of regressors.

7 Simulations & Numerical Approximations

We complement the asymptotic analysis of split-half SIS with simulations and numerical approximations. These results confirm the validity of the asymptotic theory, allow comparisons to other algorithms, and inform us about the small sample properties of SIS. First, we confirm the consistency of the frequence gauge and characterize its small sample bias. Second, we use numerical approximations to decompose the components of the asymptotic variance. Third, we confirm with simulations the distributional convergence of the frequence gauge. Fourth, we consider the bias of an updated regression estimator. Fifth, we compare the power of split-half SIS with the Andrews (1993).

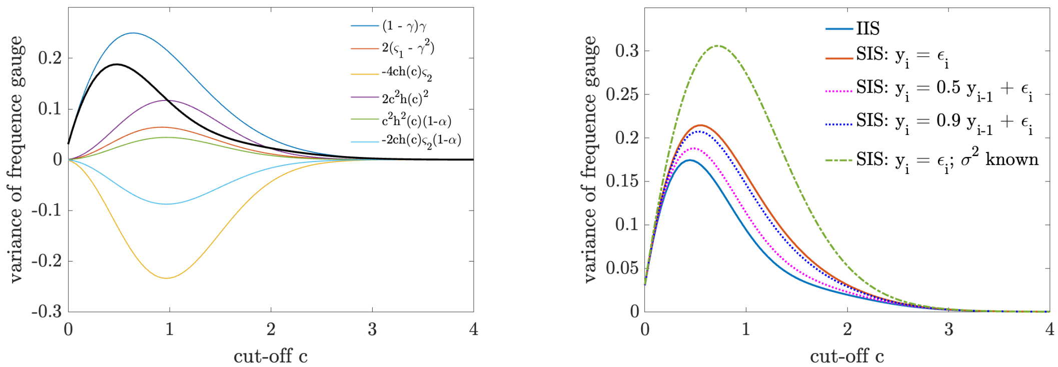
All simulations have 10^4 repetitions. Each time we increase the sample size, we redraw all n observations. The simulations have been coded in MATLAB using the MFE toolbox (Sheppard, 2018). When we do not explicitly mention otherwise, we set $\hat{\omega}^2_{j,i} = 1$ for simplification, as we are mainly concerned with evaluating the asymptotic distributions. Given a target frequence gauge γ, we choose the cut-off c in the SIS algorithm as the normal $(1 - \gamma/2)$ quantile.

7.1 Analysis of consistency of frequence gauge

We validate the consistency of the frequence gauge of split-half SIS as analyzed in Theorem 4.2. We consider two data generating processes. In both cases, the algorithm is based on the model (7) with one univariate regressor x_i and $n_1 = n_2$.

DGP1 includes an exogenous regressor $y_i = \beta x_i + \varepsilon_i$, so that x_i and ε_i are independent standard normal. The DGP1 is white noise, if $\beta = 0$, in which case y_i is also independent standard normal. As the regressor x_i is strictly exogenous Theorem 4.5 applies.

DGP2 is a first-order auto-regression $y_i = \alpha y_{i-1} + \varepsilon_i$, where $|\alpha| < 1$, ε_i is independent standard normal and $y_0 = 0$. Thus, $\delta_j = 0$ for all j and $\beta = \alpha$ in (1) while $x_i = y_{i-1}$.



(a) Variance decomposition for DGP2 (b) Variance comparison for IIS and SIS

Figure 2: Analysis of the asymptotic variance of the frequence gauge for varying c

Figure 1(a) uses DGP1 with exogenous regressor and coefficient $\beta = 0$. It shows the frequence gauge γ for an increasing sample size and different gauges γ. We use the white noise version of DGP1. We find that the small sample bias of the gauge is positive. The bias vanishes quickly with growing samples and it is modest for $n = 100$.

Figure 1(b) uses the autoregressive DGP2. It considers different values of the first-order autoregressive coefficient α for two sample sizes $n = 100$ and $1,000$. We also consider the effect of including the weights $\omega^2_{j,i}$. For constant n, the small sample bias appears to decrease for increasing α. This could reflect that as the autoregressive coefficient α increases, the sample correlation between the retained step-indicators and the autoregressive process increases. Consistent with theory, the small sample bias vanishes asymptotically. The rescaling of the estimated variance using the forward correction factors $\omega^2_{j,i}$ reduces the small sample bias by about one-third.

7.2 Analysis of asymptotic distribution of frequence gauge

We decompose the asymptotic variance of the frequence gauge of split-half SIS as a function of the cut-off c to understand the contributions of the various terms, and compare the variance to IIS. We continue to use DGP1 and DGP2.

Figure 2(a) presents a decomposition of the individual terms of the asymptotic variance of the gauge as functions of the cut-off c for the autoregressive DGP2 as given by Theorem 4.4 and Example 4.3. The terms that do not depend on estimation errors are $(1-\gamma)\gamma$ and $2(\varsigma_1 - \psi^2)$; the terms that depend on the scale estimation error are $-4c\mathsf{h}(c)\varsigma_2$ and $2c^2\mathsf{h}(c)^2$ and the terms that depend on the location estimation error are $c^2\mathsf{h}^2(c)(1-\alpha)$ and $-2c\mathsf{h}(c)\varsigma_2(1-\alpha)$. Some terms increase the asymptotic variance one of the location terms and one of the scale terms decrease the asymptotic variance.

Figure 2(b) compares the asymptotic variance of the gauge of the split-half Impulse-Indicator Saturation (IIS) to split-half SIS.

(Johansen & Nielsen, 2016b, Corollary 5) gives the asymptotic distribution of the IIS gauge as

$$n^{1/2}\{\hat{\gamma}^{\text{IIS}}_n(c) - \gamma\} \overset{\mathsf{D}}{\to} \mathsf{N}\{0, \gamma(1-\gamma) + 2c\mathsf{h}(c)\tilde{\varkappa}_2 + 2c^2\mathsf{h}^2(c)\}, \tag{53}$$

where $\tilde{\varkappa}_2 = \int_{-c}^{c}(u^2-1)\mathsf{f}(u)du$ is a truncated moment. Figure 2(b) displays the different asymptotic variance curves of the gauge as functions of the cut-off c for IIS and SIS for different DGPs. For IIS, we consider white noise DGP1. For SIS, we first consider the same DGP1, and second consider the autoregressive DGP2 with $\alpha = 0.5$ and $\alpha = 0.9$. Finally, we reconsider DGP1, but assume the error variance is known, so that $\hat{\omega}_{j,i}^2 = \sigma^2 = 1$ and two components of the asymptotic variance become zero.

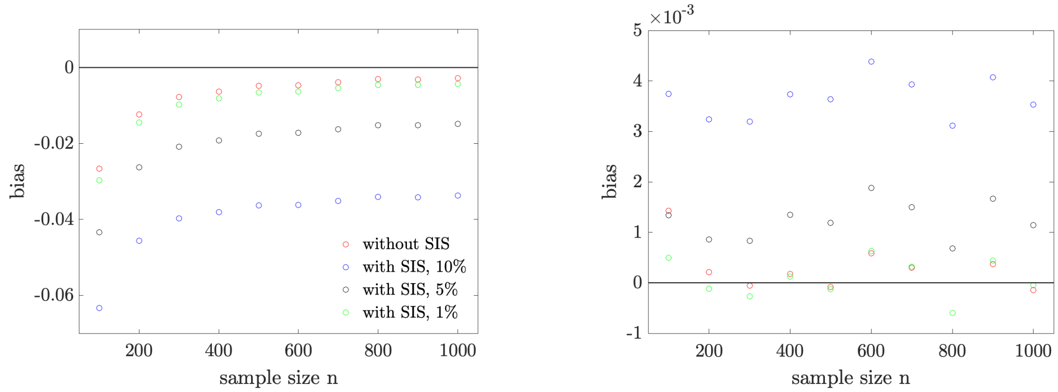
We make the following observations. First, for all c, the asymptotic variance of the gauge in IIS is lower than for all four competing SIS models. Second, running SIS knowing the variance σ^2 results in a higher asymptotic variance of the frequence gauge. Third, in the autoregressive model, the α coefficient changes the asymptotic variance. The asymptotic variance is larger with $\alpha = 0.9$ than $\alpha = 0.5$. This is different from IIS, where the asymptotic variance does not include regressor-dependant terms. Finally, we observe that the asymptotic variance of the gauge falls rapidly for growing c. This motivates the choice of a large c in empirical applications, corresponding to a gauge of 1% or lower, as recommended by Castle et al. (2015).

	γ vs. n	100	400	1600	∞
DGP1	5%	0.0516	0.0399	0.0363	0.0347
$\beta = 0$	1%	0.0160	0.0104	0.0094	0.0089
$\hat{\omega}_{j,i}^2 = 1$	0.5%	0.0093	0.0057	0.0051	0.0047
	0.1%	0.0025	0.0013	0.0011	0.0010
DGP2	5%	0.0411	0.0284	0.0261	0.0249
$\alpha = 0.5$	1%	0.0149	0.0094	0.0085	0.0079
$\hat{\omega}_{j,i}^2 = 1$	0.5%	0.0089	0.0056	0.0044	0.0044
	0.1%	0.0024	0.0013	0.0011	0.0010
DGP2	5%	0.0425	0.0348	0.0331	0.0323
$\alpha = 0.9$	1%	0.0134	0.0097	0.0087	0.0086
$\hat{\omega}_{j,i}^2 = 1$	0.5%	0.0075	0.0052	0.0049	0.0046
	0.1%	0.0019	0.0012	0.0010	0.0010
DGP1	5%	0.0649	0.0627	0.0606	0.0610
$\beta = 0$	1%	0.0132	0.0124	0.0118	0.0117
$\hat{\sigma}_j^2 = \sigma^2$	0.5%	0.0063	0.0060	0.0057	0.0057
	0.1%	0.0013	0.0011	0.0011	0.0010

Table 1: Simulated and asymptotic variance of the frequence gauge of split-half SIS

7.3 Analysis of distribution convergence of frequence gauge

We now verify the asymptotic distribution results of the frequence gauge of split-half SIS and evaluate small sample properties. Table 1 tabulates the simulated variance and computed asymptotic variance of the frequence gauge of split-half SIS for the target gauges γ = 5%, 1%, 0.5%, and 0.1% and sample sizes n = 100, 400, and 1600. We consider the same models for split-half SIS as in Figure 2(b). Overall, the finite sample variance is quite close to the asymptotic variance when $n = 400$ and not too bad when $n = 100$. Our findings are consistent with the results in Figure 2.



(a) Autoregressive model

(b) Exogeneous regressor model

Figure 3: Bias of updated regression estimator as function of sample size

7.4 Updating estimation of regression coefficients

In this section, we use simulation to show that split-half SIS can introduce a bias when updating the estimates for β in (1). We conjecture that this bias can persist asymptotically with a fixed frequence gauge.

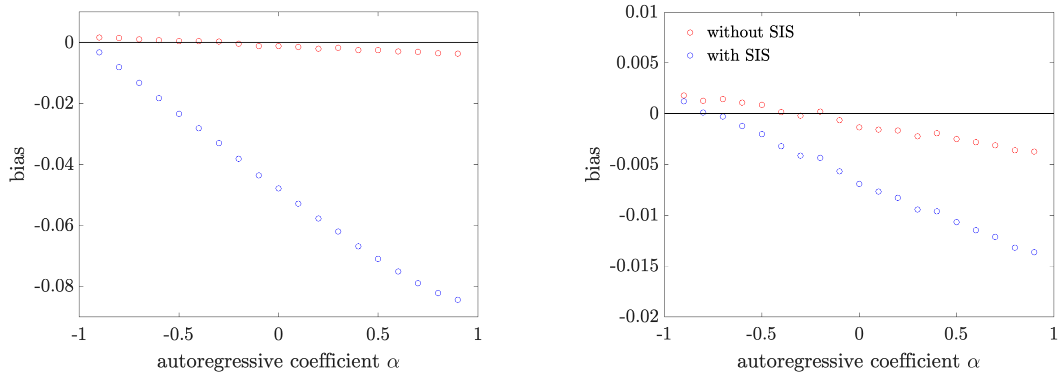
Suppose the split-half SIS Algorithm 2.1 is applied to data generated from an autoregressive model $y_i = \alpha y_{i-1} + \varepsilon_i$. This may result in $m-1$ level shifts at locations $\tau_0 = 0 < \tau_1 < \cdots < \tau_{m-1} < \tau_m = n$. We update the α estimate by the regression

$$y_i = \mu_i + \alpha y_{i-1} + u_i \qquad \text{for } \tau_{j-1} < i \leq \tau_j \text{ and } j = 1, \ldots, m. \tag{54}$$

With a frequence gauge of γ we will have approximately $m \approx \gamma n$ breaks so that the sub-sample lengths are approximately $n/m \approx 1/\gamma$. Thus, estimation of (54) corresponds to estimation of an unbalanced dynamic panel model, with a (random) increasing cross-sectional dimension and a (random) finite time dimension. It seems like we are faced with the same issues as in panel data of a incidental parameter problem (Lancaster, 2000, 2002) and a correlation of the retained (random) step-indicators with the dynamic regressors (Arellano & Bond, 1991). As with panel data, we would expect the bias to disappear asymptotically in a model with strictly exogenous regressors.

Figure 3 shows simulated biases of the updated estimator of the regression coefficients as a function of sample length n for different frequence gauges. Panel (a) uses the autoregressive DGP2 with $\alpha = 0.5$. As a baseline, we estimate the AR(1) model without split-half SIS. This shows the well-known negative finite sample bias that disappears asymptotically (Marriott & Pope, 1954). Then we use split-half SIS with the frequence gauge at 1% (green), 5% (black), and 10% (blue). We find that a larger frequence gauge is associated with a larger bias that does not appear to vanish asymptotically. When we repeat this exercise in Panel b for exogenous regressors, we find that the bias is an order of magnitude smaller than before.

Figure 4 uses the autoregressive DGP2 and shows simulated biases as a function of the autoregressive coefficient α when the sample size is $n = 1,000$. Both panels use a standard autoregressive estimation without SIS as a benchmark along, with split-half SIS estimation results. The frequence gauge is 10% in panel (a) and 1% in panel (b).



(a) Frequence gauge of $\gamma = 10\%$ (b) Frequence gauge of $\gamma = 1\%$

Figure 4: Bias of updated regression estimator as function of autoregressive coefficient

We find a bias across all values of α, and it grows together with the value of α. The bias is much larger with the frequence gauge at 10% than at 1%.

Overall, the simulations provide evidence towards the presence of an incidental parameter bias when applying SIS with dynamic regressors and calibrated through the frequence gauge. The bias increases with gauge and with the autoregressive coefficients.

7.5 Analysis of power

We compare the power of split-half SIS and the Andrews (1993) test. We consider a one-shift data generating process with a view to validate the asymptotic theory in (45) for SIS and (49) and (51) for the Andrews test.

DGP3 has one location shift and is given by

$$y_i = \alpha y_{i-1} + \delta 1_{(i \geq \lambda n)} + \varepsilon_i \qquad \text{for } i = 1, \dots, n, \tag{55}$$

with independent standard normal innovations. We will vary α, δ, λ and n.

We subject the model (55) to split-half SIS and the Andrews test. For SIS, we use a 1% gauge and compute the retention frequency for the indicator at λn. The Andrews F test for detecting a single location shift with 15% trimming has a 1% critical value of 12.35. We report the power for the (maximum) test.

Table 2 shows the simulation results. The magnitude δ of the location shift is explored along columns. The location λ is explored along rows. Panels 1 and 2 consider a non-dynamic process $\alpha = 0$ for $n = 100$ and 66. Panel 3 considers a dynamic process $\alpha = 0.5$ for $n = 66$. The value 66 is chosen to find the δ where Andrews and SIS have equal power for $\lambda = 0.95$ as discussed in theory Section 6.

The columns marked $\delta = 0$ show the finite sample size and frequence gauge. We notice that the Andrews size is always larger than the SIS gauge. We note that the distortion is larger for $\alpha = 0.5$ than for $\alpha = 0$. The power simulations are not size corrected and are therefore favourable to the Andrews test.

The theory suggests that the power increases with δ. We see that the SIS power is always increasing in δ. The Andrews power is also increasing in $\delta = 0$, 2, and 4, but it

	λ	$\delta=0$ A	$\delta=0$ SIS	$\delta=2$ A	$\delta=2$ SIS	$\delta=4$ A	$\delta=4$ SIS	$\delta=8$ A	$\delta=8$ SIS
$n=100$	0.90	1.3%	1.0%	88.6%	12.0%	100.0%	57.0%	100.0%	99.9%
$\alpha=0$	0.95	1.1%	1.1%	19.2%	11.9%	51.7%	58.3%	39.5%	99.8%
	0.99	1.1%	1.2%	2.1%	11.3%	3.4%	58.1%	0.6%	99.9%
$n=66$	0.90	1.4%	1.1%	76.5%	12.8%	99.9%	58.4%	100.0%	99.8%
$\alpha=0$	0.95	1.2%	1.1%	11.0%	12.9%	24.9%	57.1%	13.0%	99.7%
	0.99	1.3%	1.2%	3.0%	12.3%	3.7%	58.2%	0.6%	99.7%
$n=66$	0.90	2.3%	0.4%	19.6%	8.5%	74.0%	55.6%	99.8%	99.9%
$\alpha=0.5$	0.95	2.4%	0.4%	4.3%	8.3%	7.4%	56.0%	6.4%	99.9%
	0.99	2.3%	0.3%	2.7%	8.8%	2.6%	56.4%	0.8%	99.9%

Table 2: Simulated power for the Andrews (A) test and split-half SIS

declines at $\delta = 8$ for $\lambda = 0.95$ and 0.99. For $\lambda = 0.99$, it even dips below the size. This may be a finite sample effect.

The theory suggests that the power of split-half SIS is invariant to the location λ, whereas the the power of the Andrews test declines as λ approaches unity. This is confirmed in the simulations.

Further, the theory suggests that the Andrews test has higher power than split-half SIS when λ is away from 1 while SIS is more powerful for λ is close to zero. Indeed, simulations are in favour of the Andrews test for $\lambda = 0.9$ and in favour of SIS for $\lambda = 0.99$. For the inbetween case $\lambda = 0.95$, the results are mixed with SIS being more powerful except in the first panel with $n = 100$ for $\delta = 2$.

Finally, we see that the power declines with increasing temporal persistency by looking at the panels 2 and 3 where $n = 66$, but the autoregressive coefficient is $\alpha = 0$ and $\alpha = 0.5$, respectively. There is an indication that the decline in performance is larger for the Andrews test than for SIS.

8 Empirical illustration

As an empirical example on the use of stylized SIS, consider the log UK labor productivity, y_i, from the first quarter of 1980 to the third quarter of 2021. This gives a sample of length of $n = 167$ plus initial values. The labor productively is measured by the UK's Office of National Statistics as a chain volume measure of gross value added at basic prices divided by the number of hours worked. We used PcGive in OxMetrics 8 for the analysis (Doornik & Hendry, 2013).

Figure 5(a) shows the log labor productivity y with a marked decline in its growth rate after the 2008 financial crisis. There is considerable movement through the Covid pandemic from 2020. The post-2008 decline has been of concern in the political debate for some years, see for example Chadha (2022), and the submission to the Treasury Committee in October 2021 by the Bank of England's Chief Economist Huw Pill:

> "Before the global financial crisis, UK productivity growth averaged over two per cent per year. Since then, labor productivity (growth) has fallen

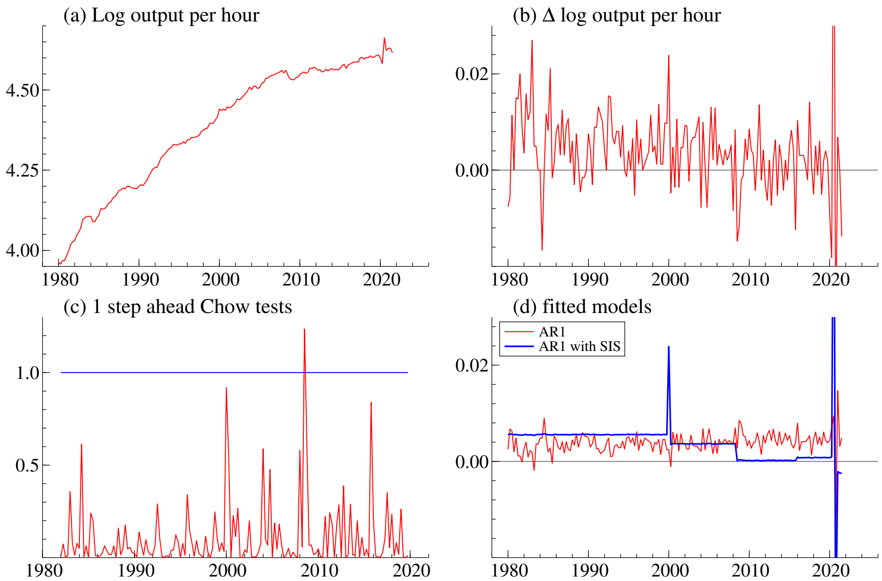


Figure 5: UK labor productivity

considerably."

Panel (b) shows the labor productivity growth rate measured as the log difference $\Delta y_i = y_i - y_{i-1}$. Note that the y-axis has been truncated to better visualize the pre-Covid periods. We make the following observations. The series is very noisy, and one can just about visually discern a gradual decline over time. We will model the growth rate as a first-order autoregression, thus imposing that the series in levels has a unit root. We will show how SIS can help in capturing the declining level of the growth rate.

We start by fitting a first-order autoregression to the growth rate for the whole period. While not reported here, the results point to a very mis-specified model, and diagnostics point to difficulties matching movements through the Covid period. An investigator may, therefore, drop that period and focus on the period until 2019:4. We then find the model:

$$\underset{(se)}{\widehat{\Delta y_i}} = \underset{(0.079)}{0.104}\Delta y_{i-1} + \underset{(0.0007)}{0.0035} \tag{56}$$

$$\widehat{\sigma} = 0.0069, \qquad n = 160, \qquad RSS = 0.0074, \tag{57}$$

$$\chi^2_{norm}[2] = 5.00\,(p = 0.082), \qquad \mathsf{F}_{ar(1-5)}[5, 153] = 1.96\,(p = 0.088) \tag{58}$$

$$\max C^2 = 8.49\,(p = 0.482)\,\{\arg\max = 2008:3\}. \tag{59}$$

$$\max F = 3.52\,(p = 0.01)\,\{\arg\max = 2004:1\}. \tag{60}$$

The fitted model reported in (56) and in Figure 5(d). The fit indicates an overall constant level for the quarterly growth rate of $0.0035/(1 - 0.104) = 0.39\%$.

We subjected the model (to 2019:4) to various misspecification tests. These do not tend to reject the model. A normality test based on cumulants (Doornik & Hansen, 2008) and a test for residual autocorrelation (Godfrey, 1978; Nielsen, 2006) are reported in (58). Figure 5(c) shows a one-step recursive Chow test with pointwise 1% critical values. This indicates a slight rejection in 2008:3, but the practitioner may not wish to give too much attention to this, given that about 144 tests were conducted (Hendry & Nielsen, 2007). Indeed, a joint test as shown in (59) does not reject the model (Nielsen & Whitby, 2015). The Andrews test reported in (60), used for detecting a single location shift, gives a marginal decision indicating a possible break in 2004:1. It appears that minor location-shifts are not reliably detected by conventional misspecification tests. Yet, Figure 5(a) does show a marked decline in the log labor productivity y_i since 2008.

We now apply the stylized SIS algorithm to the full sample until 2021:3. First, we fit the first-order autoregression to the first sample-half until 1999:4. This is the same as fitting the autoregression to the full sample combined with step-indicators for each observation from 2000:1 to 2021:3. We get

$$\underset{(se)}{\widehat{\Delta y}_i} = \underset{(0.111)}{0.201}\Delta y_{i-1} + \underset{(0.0010)}{0.0045} + \sum_{j=81}^{167} \widehat{\delta}_j 1_{(i\geq j)} \tag{61}$$

$$\widehat{\sigma} = 0.0068, \qquad n = 167, \qquad RSS = 0.0036, \tag{62}$$

$$\chi^2_{norm}[2] = 3.78\,(p = 0.151), \qquad \mathsf{F}_{ar(1-5)}[5, 73] = 1.39\,(p = 0.237). \tag{63}$$

This fit indicates a constant quarterly growth rate of $0.0045/(1 - 0.201) = 0.56\%$ prior to 2000. Test for normality and residual autocorrelation do not reject, see (63).

There are 87 estimated coefficients for the step-indicators. Computing the t-statistics for these 87 estimates, we find that the most extreme t-statistics are: 10.4 for 2020:3, -9.57 for 2020:4, 4.28 for 2021:1, -3.21 for 2000:2, -2.69 for 2008:3, 2.65 for 2016:1, 2.53 for 2000:1, 2.17 for 2004:2 and 2.00 for 2008:2. Using the 1% cut-off for the normal distribution of 2.576, we keep the six most significant step-indicators. Rerunning the model gives

$$\begin{aligned}\underset{(se)}{\widehat{\Delta y}_i} &= -\underset{(0.074)}{0.008}\Delta y_{i-1} + \underset{(0.0009)}{0.0056} \\ &\quad + \underset{(0.0068)}{0.0183}I_{(i\geq 00:1)} - \underset{(0.0069)}{0.0202}I_{(i\geq 00:2)} - \underset{(0.0015)}{0.0033}I_{(i\geq 08:3)} \\ &\quad + \underset{(0.007)}{0.080}I_{(i\geq 20:3)} - \underset{(0.012)}{0.119}I_{(i\geq 20:4)} + \underset{(0.010)}{0.036}I_{(i\geq 21:1)}\end{aligned} \tag{64}$$

$$\widehat{\sigma} = 0.0068, \qquad n = 167, \qquad RSS = 0.0072, \tag{65}$$

$$\chi^2_{norm}[2] = 5.39\,(p = 0.068), \qquad \mathsf{F}_{ar(1-5)}[5, 154] = 2.39\,(p = 0.041). \tag{66}$$

The autoregressive coefficient is now insignificant. Adding up the constant terms and correcting for the modest autoregressive coefficient gives long-run means of 0.56% prior to 2000, then 0.37% until 2008, then 0.043% until 2020.

We identify two significant drops in productivity in 2000 and 2008, corresponding to the burst of the dot-com bubble and the financial crises, respectively. Both are characterized by pairs of offsetting step indicators. However, the split half-SIS retains only one of the two step-indicators from 2008, which results in less accurate tracking of the series during the financial crisis.

The more comprehensive SIS algorithm in OxMetrics yields a similar model to split-half SIS, but it manages to retain two offsetting step-indicators for 2008 instead of just one. In the updated OxMetrics model, these indicators have larger t-statistics than the single 2008 indicator found in (64). It appears that the split-half SIS is too simple to track the somewhat protracted upheaval during the financial crisis.

9 Conclusion

In this paper, we investigated the properties of the SIS algorithm that addresses location shifts in time series in the context of model selection. The growing importance of SIS in tackling location shifts is reflected in its applications in fields as varied as economics (Chuffart & Hooper, 2019; Pellini, 2021; Bernstein & Martinez, 2021), climate science (Raggad, 2018; Pretis et al., 2018; Koch et al., 2022; O'Callaghan et al., 2022), and public health (Doornik et al., 2022). In this section, we summarize the insights gained through a study of SIS with asymptotic analysis, simulations, and numerical approximations.

The first insight is that the frequence gauge is consistent for a wide range of both stationary and non-stationary regressors. This means that even without detailed knowledge of the regressor types, an investigator can choose the cut-off of SIS from the limiting gauge. To address the sensitivity of this result, we demonstrated that the variation of the frequence gauge around its limit follows a normal distribution. However, its variance depends on the type of regressors. Simulations revealed that this variation remains limited, even in small samples. As a result, the sole tuning parameter of the SIS algorithm can be finely adjusted to align with the investigator's preferences.

The second insight concerns the link between the frequence gauge and the bias in the updated regression estimator after selecting over step-indicators. This bias appears to emerge in the presence of dynamic regressors when searching for location shifts. This contrasts with the theory of Impulse Indicator Saturation, where there is no such bias (Johansen & Nielsen, 2016b). The bias diminishes as the gauge decreases, suggesting that the gauge should be chosen small and possibly vanishing with sample size. For that purpose, we developed a Poisson theory for the absolute gauge. For a sample size of $n = 100$ observations, we recommend setting the absolute gauge to 1, which is equal to the frequence gauge of 1%, in line with Castle et al. (2015). In larger samples, we advise targeting the absolute gauge rather than the frequence gauge, so that the cut-off drifts slowly to infinity.

The third insight pertains to the circumstances in which stylized SIS demonstrates higher statistical power compared to the Andrews (1993) test. We developed a local power theory for stylized SIS and the Andrews test. Our findings suggest that the Andrews test maintains consistency when faced with one or two well-separated, central location shifts, whereas the SIS shows trivial power. Conversely, for location shifts near the end of the sample or for two offsetting location shifts close to each other, the SIS maintains power, while the power of the Andrews test goes down to its size. In time series observed over extended periods, major upheavals like the 2008 financial crisis and the 2020 Covid pandemic might recur. Consequently, we anticipate multiple breaks in the data. These breaks may occur closely together or towards the end of the sample. In such scenarios, SIS appears to be preferable to the Andrews test. The same conclusions hold for the Bai & Perron (1998) procedure that allows more breaks but inherits the power trade-offs from the Andrews test.

The fourth insight relates to the regularity conditions of SIS compared to the Andrews test. SIS assumes a known error distribution, while the Andrews test does not. The assumption is testable and contributes to the power of SIS to detect breaks that occur closely together. The first-order theory for SIS applies to a variety of stationary and non-stationary regressors. In order to do this, the present theory is formulated in terms of normalization matrices. This implies that the theory works regardless of the choice of the normalization matrix. In contrast, the asymptotic theory for the Andrews test requires stationary regressors, introducing an additional risk of mistakes, as the investigator must carefully determine the appropriate normalization of the regressors. Furthermore, SIS is designed to be implemented along with regressor selection, which is useful when there is uncertainty about the choice of regressors.

The theory for SIS is complicated because SIS operates on the differenced residuals which are temporally dependent even for well-behaved errors. We found various technical solutions that may be useful elsewhere. The empirical process theory was developed using ideas from the McLeish (1977) mixingale theory. The Poisson theory requires the Chen (1975) Poisson limit theorem for dependent binary variables. In addition, to allow two close breaks in the power theory, we relied on the Skorokhod (1956) M_1-metric favoured by Whitt (2002) rather than the J_1-metric favoured by Billingsley (1968).

A potential further development is to develop a test for the presence of location shifts along the lines of the IIS test for outliers of Jiao & Pretis (2022). The techniques for dealing with correlation between (differenced) errors and regressors turns out to be useful for analyzing instrumental variable estimation (Jiao, 2019).

A Proofs

In section A.1, we prove Theorem 2.3. In section A.2-A.6, we state and prove some auxiliary results. Finally, the main results for the gauge are proven in sections A.7-A.9.

A.1 Properties of differenced innovations

Proof of Theorem 2.3. (a) *Symmetry.* $\mathsf{h}(x)$ is the density of the difference of two i.i.d. variables ε_i, ε_j. Symmetry follows since $\varepsilon_i - \varepsilon_j$, $\varepsilon_j - \varepsilon_i$ are identically distributed.
(b) *Normal distribution.* If ε/σ has a standard normal density f, then $\nabla\varepsilon/(\sqrt{2}\sigma)$ has density $\mathsf{h} = \mathsf{f}$. Conversely, if $\mathsf{h} = \mathsf{f}$ then h, f are symmetric by part (a), so that $\varepsilon_i - \varepsilon_j$ and $\varepsilon_i + \varepsilon_j$ are identically distributed. In particular, ε_i, ε_j and $(\varepsilon_i + \varepsilon_j)/\sqrt{2}$ are identically distributed. Pólya (1923) shows that if that distribution is continuous with finite variance, then it must be normal with zero mean.
(c) *Bounded densities.* The inequality $|x-y|^k \le C_k(|x|^k + |y|^k)$ for some $C_k > 0$ implies

$$\begin{aligned} |v|^k \mathsf{h}(v) &= \sqrt{2}\int_{-\infty}^{\infty} |v+y-y|^k \mathsf{f}(y)\mathsf{f}(v+y)dy \\ &\le \sqrt{2}C_k\Big\{\int_{-\infty}^{\infty} |y|^k \mathsf{f}(y)\mathsf{f}(v+y)dy + \int_{-\infty}^{\infty} |v+y|^k \mathsf{f}(y)\mathsf{f}(v+y)dy\Big\}. \end{aligned}$$

In the first integral, bound $|y|^k\mathsf{f}(y)$ by its supremum and change variable from y to $s = v + y$. In the second integral, bound $|v+y|^k\mathsf{f}(v+y)$ by its supremum. We get

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

(d) *Bounded derivatives.* By the Leibniz rule for improper integrals

$$v^k\dot{\mathsf{h}}(v) = v^k\sqrt{2}\frac{\partial}{\partial v}\int_{-\infty}^{\infty} \mathsf{f}(y)\mathsf{f}(v+y)dy = v^k\sqrt{2}\int_{-\infty}^{\infty} \mathsf{f}(y)\dot{\mathsf{f}}(v+y)dy.$$

Then proceed as in part (c) to get

$$\begin{aligned} |v^k\dot{\mathsf{h}}(v)| &\le C_k\sqrt{2}\int_{-\infty}^{\infty} \mathsf{f}(y)(|y|^k + |v+y|^k)|\dot{\mathsf{f}}(v+y)|dy \\ &\le C_k\sqrt{2}\Big\{\sup_{v\in\mathbb{R}} |\dot{\mathsf{f}}(v)|\int_{-\infty}^{\infty} |y|^k \mathsf{f}(y)dy + \sup_{v\in\mathbb{R}} |v^k\dot{\mathsf{f}}(v)|\int_{-\infty}^{\infty} \mathsf{f}(y)dy\Big\}. \end{aligned}$$

This is finite when $\mathsf{E}|\varepsilon_i|^k < \infty$ and $\sup_{v\in\mathbb{R}}(1 + |v|^k)|\dot{\mathsf{f}}(v)| < \infty$. □

A.2 Expanding distribution function for residuals

The compensators in the empirical process theory will be quite complicated due to the temporal dependence arising from forward differencing. Their analysis will be facilitated by the following expansion of the distribution function for a single residual when the estimation error can be assumed constant.

Theorem A.1. *Let $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$ be random with density $\mathsf{m}_{Y,X}(y,x)$ with respect to the product of the Lebesgue measure and some measure υ on $\mathbb{R}^p$. Suppose (i) there exists densities so that $\mathsf{m}_{Y,X}(y,x) = \mathsf{m}_{Y|X}(y|x)\mathsf{m}_X(x) = \mathsf{m}_{X|Y}(x|y)\mathsf{m}_Y(y)$;*
(ii) $\mathsf{m}_{Y|X}(y|x)$ has y-derivative $\dot{\mathsf{m}}_{Y|X}(y|x)$;
(iii) $C_\mathsf{m} = \sup_{y\in\mathbb{R},x\in\mathbb{R}^p}(1+|y|)|\dot{\mathsf{m}}_{Y|X}(y|x)| < \infty$;
(iv) $\mathsf{E}|X|^2 < \infty$.
Then, for $|a| \leq 1/2$, $b \in \mathbb{R}^p$ and $c \in \mathbb{R}$ and with $c^\dagger = c(1+a)$, we get
(a) $|\mathsf{P}(Y - b'X \leq c) - \mathsf{P}(Y \leq c) - \mathsf{m}_Y(c)\mathsf{E}(b'X|Y=c)| \leq |b|^2 C_\mathsf{m}\mathsf{E}|X|^2/2$.
(b) $|\mathsf{m}_Y(c^\dagger)\mathsf{E}(b'X|Y=c^\dagger) - \mathsf{m}_Y(c)\mathsf{E}(b'X|Y=c)| \leq 2|ab|C_\mathsf{m}\mathsf{E}|X|$.

Lemma A.2. *(Jiao & Nielsen (2017), Lemma 1.1) If $|c^* - c| \leq |Ac+B|$ and $|A| \leq 1/2$, then $|c| \leq 2(|c^*| + |B|)$ and $(Ac+B)^2 \leq 16\{A^2(c^*)^2 + B^2\}$.*

Proof of Theorem A.1. (a) Write $\mathcal{P} = \mathsf{P}(Y - b'X \leq c) - \mathsf{P}(Y \leq c)$ as an integral:

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

Apply the Mean Value Theorem and the identity $\mathsf{m}_{Y,X} = \mathsf{m}_{Y|X}\mathsf{m}_X = \mathsf{m}_{X|Y}\mathsf{m}_Y$ to get

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

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

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

The first term is $\mathcal{P}_1 = \mathsf{m}_Y(c)\mathsf{E}(b'X|Y=c)$. For the second term, the triangle inequality gives

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

By assumption, $C_\mathsf{m} = \sup_{c^*\in\mathbb{R},x\in\mathbb{R}^p}(1+|c^*|)|\dot{\mathsf{m}}_{Y|X}(c^*|x)| < \infty$. The norm inequality gives $(b'x)^2 \leq |x|^2|b|^2$. Thus, we get uniformly in c, that $|\mathcal{P}_2| \leq |b|^2 C_\mathsf{m}\mathsf{E}|X|^2/2$.

(b) Consider the difference term $|\mathsf{q}(c^\dagger) - \mathsf{q}(c)|$, where

$$\mathsf{q}(y) = \mathsf{m}_Y(y)\mathsf{E}(b'X|Y=y) = \mathsf{m}_Y(y)\int_{\mathbb{R}^p} b'x\mathsf{m}_{X|Y}(x|y)d\upsilon(x).$$

Let $c^\dagger = c(1+a)$. Apply the Mean Value Theorem and $\mathsf{m}_{Y|X}\mathsf{m}_X = \mathsf{m}_{X|Y}\mathsf{m}_Y$ to get

$$\mathsf{q}(c^\dagger) - \mathsf{q}(c) = (c^\dagger - c)\int_{-\infty}^{\infty}(b'x)\dot{\mathsf{m}}_{Y|X}(c^*|x)\mathsf{m}_X(x)d\upsilon(x),$$

where $|c^* - c| \leq |c^\dagger - c| \leq |ac|$. Lemma A.2 shows that $|c| \leq 2|c^*|$ since $|a| \leq 1/2$ by assumption. By assumption $C_\mathsf{m} = \sup_{c^*\in\mathbb{R},x\in\mathbb{R}^p}(1+|c^*|)|\dot{\mathsf{m}}_{Y|X}(c^*|x)| < \infty$, we have $|\mathsf{q}(c^\dagger) - \mathsf{q}(c)| \leq 2|ab|C_\mathsf{m}\mathsf{E}|X|$. □

A.3 Exponential martingale inequalities

The subsequent empirical process theory relies on a linear chaining argument. The chaining argument uses a new iterated exponential martingale inequality. Our inequality is related to that of Johansen & Nielsen (2016a, Theorem 5.1), which iterates the Bercu & Touati (2008) exponential inequality for unbounded martingales. Here, a simpler result suffices, which uses the Freedman (1975) exponential inequality for bounded martingales.

We present two versions. The first version is an exact tail probability bound.

Theorem A.3. *For $1 \leq \ell \leq L$, let $M_{\ell n} = \sum_{i=1}^{n}(z_{i\ell} - \mathsf{E}_{i-1} z_{i\ell})$ denote a martingale, where $z_{i\ell}$ is $\mathcal{F}_i$-adapted and $|z_{i\ell} - \mathsf{E}_{i-1} z_{i\ell}| \leq 1$. Then, for all $\kappa_0, \kappa_1 > 0$, we have*

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

The second version is an asymptotic tail probability bound.

Theorem A.4. *For $1 \leq \ell \leq L$, let $M_{\ell n} = \sum_{i=1}^{n}(z_{i\ell n} - \mathsf{E}_{i-1} z_{i\ell n})$ denote a martingale array, where $z_{i\ell n}$ is $\mathcal{F}_{in}$-adapted and $|z_{i\ell n}| \leq 1$. Suppose $\exists \varsigma, \lambda \geq 0$ so that $\mathsf{E} \max_{1\leq\ell\leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} z_{i\ell n}^2 = \mathrm{O}(n^{\varsigma})$ and $L = \mathrm{O}(n^{\lambda})$. Then, $\forall \nu > \varsigma/2, \kappa > 0$ we get*

$$\lim_{n\to\infty} \mathsf{P}\left\{\max_{1\leq\ell\leq L} |M_{\ell n}| > \kappa n^{\nu}\right\} = 0.$$

Lemma A.5. *(Freedman (1975), Theorem 1.6) Let $M_n = \sum_{i=1}^{n}(z_i - \mathsf{E}_{i-1} z_i)$ denote a martingale, where z_i is $\mathcal{F}_i$-adapted with $|z_i - \mathsf{E}_{i-1} z_i| \leq 1$. Let $T_n = \sum_{i=1}^{n} \mathsf{Var}(z_i|\mathcal{F}_{i-1})$. For $a, b > 0$ we get $\mathsf{P}(M_n \geq a, T_n \leq b) \leq \exp[-a^2/\{2(a+b)\}]$.*

Proof of Theorem A.3. Let $m_{i\ell n} = z_{i\ell n} - \mathsf{E}_{i-1} z_{i\ell n}$. Let $A_\ell = \sum_{i=1}^{n} m_{i\ell}$ and $\mathcal{A} = (\max_{1\leq\ell\leq L} |A_\ell| > \kappa_0)$. Let $B_\ell = \sum_{i=1}^{n} \mathsf{E}_{i-1} m_{i\ell}^2$ and $\mathcal{B} = (\max_{1\leq\ell\leq L} B_l \leq \kappa_1)$. 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 $\mathcal{A}_\ell = (|A_\ell| > \kappa_0)$ and $\mathcal{B}_\ell = (|B_\ell| \leq \kappa_1)$. Note $\mathcal{A} = \bigcup_{\ell=1}^{L} \mathcal{A}_\ell$ and $\mathcal{B} \subset \mathcal{B}_\ell$ and apply Boole's inequality to get

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

Apply Lemma A.5, noting that $|m_{i\ell}| \leq 1$, to get

$$\mathsf{P}(\mathcal{A} \cap \mathcal{B}) \leq \sum_{l=1}^{L} \mathsf{P}\big\{(A_\ell > \kappa_0) \cap \mathcal{B}_\ell\big\} + \mathsf{P}\left\{(-A_\ell > \kappa_0) \cap \mathcal{B}_\ell\right\} \leq 2L \exp\Big\{-\frac{\kappa_0^2}{2(\kappa_1+\kappa_0)}\Big\}.$$

Bounding $\mathsf{P}(\mathcal{B}^c)$. The Markov inequality gives

$$\mathsf{P}(\mathcal{B}^c) = \mathsf{P}(\max_{1\leq\ell\leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} m_{i\ell}^2 > \kappa_1) \leq \frac{1}{\kappa_1} \mathsf{E} \max_{1\leq\ell\leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} m_{i\ell}^2.$$

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

Proof of Theorem A.4. We consider the probability $\mathcal{P}_n = \mathsf{P}\{\max_{1\leq\ell\leq L}|M_{\ell n}| > \kappa n^\nu\}$, where the martingale $M_{\ell n}$ has differences $|z_{i\ell n} - \mathsf{E}_{i-1}z_{i\ell n}| \leq 2$, since it is assumed that $|z_{i\ell n}| \leq 1$, but otherwise $M_{\ell n}$ is of the form studied in Theorem A.3. Thus, for some $\kappa > 0$, apply Theorem A.3 with $\kappa_0 = \kappa n^\nu/2$ and $\kappa_1 = \kappa^2 n^{2\nu}\{4(1+\lambda)\log n\}^{-1}/2^2$. Let n be fixed and sufficiently large such that $\kappa_0 < \kappa_1$. Use the bound

$$\exp[-\kappa_0^2/\{2(\kappa_1+\kappa_0)\}] \leq \exp\{-\kappa_0^2/(4\kappa_1)\} = n^{-1-\lambda},$$

and the assumptions $\mathsf{E}\max_{1\leq\ell\leq L}\sum_{i=1}^n \mathsf{E}_{i-1}z_{i\ell n}^2 = \mathrm{O}(n^\varsigma)$ and $L = \mathrm{O}(n^\lambda)$ to get that $\mathcal{P}_n = \mathrm{O}(n^\varsigma)/\kappa_1 + n^\lambda n^{-1-\lambda}$. Note that $n^\varsigma/\kappa_1 \to 0$ when $2v > \varsigma$ to get that $\mathcal{P}_n = \mathrm{o}(1)$. $\square$

A.4 The one-sided empirical process

We establish some empirical process results for differenced residuals. For this purpose, we simplify the setup relative to that of SIS. In SIS, estimation of β, σ is done on one subsample, and the evaluation of residuals is done on another sample. Here, the random estimation error is replaced by a deterministic error. We can therefore avoid the division of the sample into subsamples. We will therefore avoid reference to subsamples.

We consider the model (7). Recall the definition (19) and modify definition (22) as

$$\chi_i = (\varepsilon_i - \varepsilon_{i+1})/(\sqrt{2}\sigma), \qquad \nabla x_{in} = N'(x_i - x_{i+1}) \text{ for } i = 1,\ldots,n. \tag{A.1}$$

Thus, N is a normalization matrix similar to those considered before, but applied to the full sample. Let a and b represent estimation errors in the scale and the location. Define the empirical distribution function

$$\widehat{\mathsf{F}}_n(a,b,c) = n^{-1}\sum_{i=1}^n 1_{(\chi_i \leq c + n^{-1/2}ac + b'\nabla x_{in})}. \tag{A.2}$$

Here, χ_i is $\mathcal{F}_{i+1}$-adapted and ∇x_{in} is $\mathcal{F}_i$-adapted. Thus, we will refer to

$$\overline{\mathsf{F}}_n(a,b,c) = n^{-1}\sum_{i=1}^n \mathsf{E}_{i-1}1_{(\chi_i \leq c + n^{-1/2}ac + b'\nabla x_{in})} \tag{A.3}$$

as a pseudo-compensator for $\widehat{\mathsf{F}}_n$. The empirical process

$$\mathbb{F}_n(a,b,c) = n^{1/2}\{\widehat{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(a,b,c)\} \tag{A.4}$$

satisfies the following result, which will be proved by linear chaining.

Theorem A.6. *Suppose Assumption 3.1 holds and that*
(i) the marginal density f *is bounded:* $\sup_{v\in\mathbb{R}}\mathsf{f}(v) < \infty$*;*
(ii) the regressors x_i *satisfy* $\mathsf{E}\sum_{i=1}^n|\nabla x_{in}| = \mathsf{O}(n^{1/2})$.
Then, for all $B > 0$, $0 < \eta < 1/4$ *,and* $c \in \mathbb{R}$ *we have*

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

Lemma A.7. *Let* $\chi_i = (\varepsilon_i - \varepsilon_{i+1})/(\sqrt{2}\sigma)$, *where* ε_i/σ *is* $\mathcal{F}_i$*-adapted and has density* f. *Let* $c_i \leq \overline{c}_i$ *be* $\mathcal{F}_{i-1}$*-adapted random variables. Then*

$$\mathsf{E}_i 1_{(c_i < \chi_i \leq \overline{c}_i)} \leq \sqrt{2}(\overline{c}_i - c_i) \sup_{v \in \mathbb{R}} \mathsf{f}(v).$$

Proof of Lemma A.7. Write the indicator as $1_{(c_i \leq \chi_i \leq \overline{c}_i)} = 1_{(\varepsilon_i/\sigma - \sqrt{2}\overline{c}_i \leq \varepsilon_{i+1}/\sigma < \varepsilon_i/\sigma - \sqrt{2}c_i)}$. Only ε_{i+1} is varying when conditioning on $\mathcal{F}_i$. Thus, the Mean Value Theorem gives

$$\mathsf{E}_i 1_{(\varepsilon_i/\sigma - \sqrt{2}\overline{c}_i \leq \varepsilon_{i+1}/\sigma < \varepsilon_i/\sigma - \sqrt{2}c_i)} = \int_{\varepsilon_i/\sigma - \sqrt{2}\overline{c}_i}^{\varepsilon_i/\sigma - \sqrt{2}c_i} \mathsf{f}(x)dx = \sqrt{2}(\overline{c}_i - c_i)\mathsf{f}(v^*),$$

where $|v^* - c_i| \leq \sqrt{2}(\overline{c}_i - c_i)$. Finally, note that $\mathsf{f}(v^*) \leq \sup_{v \in \mathbb{R}} \mathsf{f}(v)$. □

Proof of Theorem A.6. Combine the two estimation errors as $u = (a, b')'$ and create an expanded vector of regressors $w_{in} = (n^{-1/2}c, \nabla x_{in}')'$ so that $n^{-1/2}ac + b'\nabla x_{in} = u'w_{in}$.

Recall the definition of $\mathbb{F}_n$ from (A.4) and write our object of interest as

$$\begin{aligned} R_n(u, c) &= \mathbb{F}_n(a, b, c) - \mathbb{F}_n(0, 0, c) \\ &= n^{-1/2} \sum_{i=1}^n [\{1_{(\chi_i \leq c + u'w_{in})} - 1_{(\chi_i \leq c)}\} - \mathsf{E}_{i-1}\{1_{(\chi_i \leq c + u'w_{in})} - 1_{(\chi_i \leq c)}\}]. \end{aligned}$$

We show $\mathcal{R}_n = \sup_{|u| \leq n^{1/4 - \eta}B} |R_n(u, c)| = \mathrm{o}_{\mathsf{P}}(1)$.

This proof has three parts. First, we chain over u by introducing grid points u_m. Second, we show that our empirical process vanishes on the grid points u_m,

$$\mathcal{R}_{n,1} = \max_{1 \leq m \leq M} |R_n(u_m, c)| = \mathrm{o}_{\mathsf{P}}(1). \tag{A.5}$$

Third, we show that our empirical process vanishes in-between our grid points u_m,

$$\mathcal{R}_{n,2} = \max_{1 \leq m \leq M} \sup_{|u - u_m| \leq \delta} |R_n(u, c) - R_n(u_m, c)|. \tag{A.6}$$

1. The chaining setup. We chain over u, by covering them with balls of radius $\delta > 0$. We will choose δ independently of the sample size n in point 3.5 below.

1.1. Cover. For $\delta, n > 0$, cover the set $|u| \leq n^{1/4-\eta}B$ with balls of radius δ which centre in grid points u_m. Thus, for any u there exists a u_m so that $|u - u_m| \leq \delta$. The minimum cover has $M \sim (n^{1/4-\eta}B/\delta)^{\dim x + 1} \sim n^{(1/4-\eta)\dim x}/\delta^{\dim x + 1}$ balls.

1.2. Apply chaining. Write $R_n(u, c) = R_n(u_m, c) + \{R_n(u, c) - R_n(u_m, c)\}$, where $R_n(u_m, c)$ is a discrete point term and $R_n(u, c) - R_n(u_m, c)$ is a local oscillation term. By the triangle inequality, $\mathcal{R}_n \leq \mathcal{R}_{n,1} + \mathcal{R}_{n,2}$ where $\mathcal{R}_{n,1}$, $\mathcal{R}_{n,2}$ are given in (A.5), (A.6).

2. The discrete point term $\mathcal{R}_{n,1}$ is $\mathrm{o}_{\mathsf{P}}(1)$. We decompose R_n into martingales. Then, we apply Theorem A.4 on the constructed martingales.

2.1. Martingale decomposition. Let $z_{im} = 1_{(\chi_i \leq c + u_m'w_{in})} - 1_{(\chi_i \leq c)}$. Define martingales

$$R_n^a(u_m, c) = n^{-1/2} \sum_{i=1}^n (z_{im} - \mathsf{E}_i z_{im}), \quad R_n^b(u_m, c) = n^{-1/2} \sum_{i=1}^n (\mathsf{E}_i z_{im} - \mathsf{E}_{i-1} z_{im}),$$

so that $R_n = R_n^a + R_n^b$. Thus, it suffices to show that

$$\mathcal{R}_{n,1}^a = \max_{1\le m\le M} |R_n^a(u_m, c)| = \mathsf{o}_\mathsf{P}(1), \quad \mathcal{R}_{n,1}^b = \max_{1\le m\le M} |R_n^b(u_m, c)| = \mathsf{o}_\mathsf{P}(1).$$

2.2. The martingale $\mathcal{R}_{n,1}^a$. We show that $\mathcal{R}_{n,1}^a = \mathsf{o}_\mathsf{P}(1)$, by applying Theorem A.4 to it. We set $\nu_a = 1/2$, let index $\ell = m$, and consider $z_{i\ell,a} = z_{im}$, which is $\mathcal{F}_{i+1}$-adapted. Note that $|z_{i\ell,a}| \le 1$. We verify the conditions of Theorem A.4.

The parameter λ_a. The set of indices ℓ has size $L = M$. Since $M \sim n^{(1/4-\eta)(\dim x+1)}$ as δ is fixed then $L \sim n^{\lambda_a}$ where $\lambda_a = (1/4-\eta)(\dim x + 1)$.

The parameter ς_a. We show $\mathcal{E}_a = \mathsf{E}\max_{1\le\ell\le L}\sum_{i=1}^n \mathsf{E}_i z_{i\ell,a}^2 = \mathrm{O}(n^{3/4-\eta})$. First note $\mathsf{E}_i z_{i\ell,a}^2 = \mathsf{E}_i|z_{i\ell,a}|$. Further $\mathsf{E}_i|z_{i\ell,a}| \le \mathsf{E}_i 1_{(c-|u_m' w_{in}|<\chi_i\le c+|u_m' w_{in}|)}$. Apply Theorem A.7 with $c_i = c - |u_m' w_{in}|$ and $\bar{c}_i = c + |u_m' w_{in}|$. Since $|u_m| \le n^{1/4-\eta}B$, we get, uniformly in ℓ,

$$\mathsf{E}_i|z_{i\ell,a}| \le 2\sqrt{2}n^{1/4-\eta}B|w_{in}|\sup_{v\in\mathbb{R}}\mathsf{f}(v), \tag{A.7}$$

where only $|w_{in}|$ is random and depends on i. Apply the Law of Iterated Expectations to get $\mathsf{E}\sum_{i=1}^n \mathsf{E}_i|w_{in}| = \mathsf{E}\sum_{i=1}^n |w_{in}|$. Since $w_{in} = (n^{-1/2}c, \nabla x_{in}')'$, we get the further bound $n^{1/2}|c| + \mathsf{E}\sum_{i=1}^n |\nabla x_{in}|$, which is $\mathrm{O}(n^{1/2})$ since c is fixed and by condition (ii). Further, $\sup_{v\in\mathbb{R}}\mathsf{f}(v) < \infty$ by condition (i). Therefore $\mathcal{E}_a = \mathrm{O}(n^{\varsigma_a})$ where $\varsigma_a = 3/4 - \eta$.

The condition $\varsigma_a < 2\nu_a$. Since $0 < \eta$ and $\nu_a = 1/2$, we have $\varsigma_a = 3/4 - \eta < 1 = 2\nu_a$.

2.3. The martingale $\mathcal{R}_{n,1}^b$. We show that $\mathcal{R}_{n,1}^b = \mathsf{o}_\mathsf{P}(1)$ by applying Theorem A.4 to it. We set $\nu_b = 1/2$, let index $\ell = m$, and consider $z_{i\ell,b} = \mathsf{E}_i z_{i\ell,a} = \mathsf{E}_i z_{im}$, which is $\mathcal{F}_i$-adapted. Note that $|z_{i\ell,b}| \le 1$ as $|z_{im}| \le 1$. We verify the conditions of Theorem A.4.

The parameter λ_b is $\lambda_b = (1/4-\eta)(\dim x+1)$ as in point 2.2.

The parameter ς_b. We show that $\mathcal{E}_b = \mathsf{E}\max_{1\le\ell\le L}\sum_{i=1}^n \mathsf{E}_{i-1} z_{i\ell,b}^2 = \mathrm{O}(n^{3/4-\eta})$. Note that $z_{i\ell,b}^2 = \mathsf{E}_i^2 z_{i\ell,a} \le \mathsf{E}_i z_{i\ell,a}^2 = \mathsf{E}_i|z_{i\ell,a}|$ by Jensen's inequality. In (A.7) we found that $\mathsf{E}_i|z_{i\ell,a}| \le 2\sqrt{2}n^{1/4-\eta}B|w_{in}|\sup_{v\in\mathbb{R}}\mathsf{f}(v)$. Therefore $\mathcal{E}_b$ has the same bound as $\mathcal{E}_a$. Thus, by point 2.2 we get $\mathcal{E}_b = \mathrm{O}(n^{\varsigma_b})$ where $\varsigma_b = \varsigma_a = 3/4 - \eta$.

The condition $\varsigma_b < 2\nu_b$ is satisfied as in point 2.2, since $(\nu_b, \lambda_b, \varsigma_b) = (\nu_a, \lambda_a, \varsigma_a)$.

3. The oscillation term $\mathcal{R}_{n,2}$. We show that $\mathcal{R}_{n,2}$ is $\mathsf{o}_\mathsf{P}(1)$. The proof relies on bounding $S_n(u_m, u, c) = R_n(u, c) - R_n(u_m, c)$ uniformly in u. We then apply a martingale decomposition and use Theorem A.4.

3.1 The term S_n. Write

$$S_n(u_m, u, c) = n^{-1/2}\sum_{i=1}^n \{s_i(u_m, u, c) - \mathsf{E}_{i-1}s_i(u_m, u, c)\},$$

where, due to a cancellation of two indicator functions $1_{(\chi_i\le c)}$, we have

$$s_i(u_m, u, c) = 1_{(\chi_i\le c+u'w_{in})} - 1_{(\chi_i\le c+u_m'w_{in})}.$$

Therefore $\mathcal{R}_{n,2} = \max_{1\le m\le M}\sup_{|u-u_m|\le\delta}|S_n(u_m, u, c)|$.

3.2. Bounding $s_i(u_m, u, c)$. Write $c + u'w_{in} = c + u_m' w_{in} + (u-u_m)'w_{in}$. Noting that $|u - u_m| \le \delta$, we introduce bounds,

$$c_{im} = c + u_m' w_{in} - \delta|w_{in}|, \quad \bar{c}_{im} = c + u_m' w_{in} + \delta|w_{in}|, \tag{A.8}$$

which do not depend on u. Thus, we can bound

$$|s_i(u_m, u, c)| \leq z_{im} = 1_{(c_{im} < \chi_i \leq \overline{c}_{im})}. \tag{A.9}$$

3.3. Bounding $|S_n(u_m, u, c)|$. The triangle inequality gives that $|\sum_{i=1}^n (s_i - \mathsf{E}_{i-1} s_i)| \leq \sum_{i=1}^n (|s_i| + \mathsf{E}_{i-1}|s_i|)$. Using that $|s_i| \leq z_{im}$ by (A.9) in point 3.2 leads to the bound

$$|S_n(u_m, u, c)| \leq M_{mn} = n^{-1/2} \sum_{i=1}^n (z_{im} + \mathsf{E}_{i-1} z_{im}),$$

uniformly in u. Thus, $\mathcal{R}_{n,2} = \mathrm{o}_{\mathsf{P}}(1)$ if $\max_{1 \leq m \leq M} M_{mn} = \mathrm{o}_{\mathsf{P}}(1)$.

3.4. Martingale decomposition. We decompose M_{mn} into two martingale and a compensator term. Add and subtract $n^{-1/2} \sum_{i=1}^n \mathsf{E}_i z_{im}$ twice to M_{mn} and write $M_{mn} = \widetilde{M}^c_{mn} + \widetilde{M}^d_{mn} + 2\overline{M}_{mn}$, where we have two *martingale terms*

$$\widetilde{M}^c_{mn} = n^{-1/2} \sum_{i=1}^n (z_{im} - \mathsf{E}_i z_{im}), \quad \widetilde{M}^d_{mn} = n^{-1/2} \sum_{i=1}^n (\mathsf{E}_i z_{im} - \mathsf{E}_{i-1} z_{im}),$$

and a *compensator term*

$$\overline{M}_{mn} = n^{-1/2} \sum_{i=1}^n \mathsf{E}_{i-1} z_{im}.$$

Thus, it suffices to show that $\widetilde{\mathcal{M}}^c_n = \max_{1 \leq m \leq M} \widetilde{M}^c_{mn}$, $\widetilde{\mathcal{M}}^d_n = \max_{1 \leq m \leq M} \widetilde{M}^d_{mn}$, and $\overline{\mathcal{M}}_n = \max_{1 \leq m \leq M} \overline{M}_{mn}$ are $\mathrm{o}_{\mathsf{P}}(1)$.

3.5. The compensator term $\overline{\mathcal{M}}_n$. We show that $\overline{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$. Recall from (A.9) that $z_{im} = 1_{(c_{im} < \chi_i \leq \overline{c}_{im})}$, where $c_{im} = c + u_m' w_{in} - \delta |w_{in}|$ and $\overline{c}_{im} = c + u_m' w_{in} + \delta |w_{in}|$. The Law of Iterated Expectations gives $\mathsf{E}_{i-1} z_{im} = \mathsf{E}_{i-1} \mathsf{E}_i z_{im}$. Apply Lemma A.7 to get

$$\mathsf{E}_{i-1} z_{im} = \mathsf{E}_{i-1} \mathsf{E}_i z_{im} \leq \mathsf{E}_{i-1} 2\sqrt{2} \delta |w_{in}| \sup_{v \in \mathbb{R}} \mathsf{f}(v) = 2\sqrt{2} \delta \mathsf{E}_{i-1} |w_{in}| \sup_{v \in \mathbb{R}} \mathsf{f}(v), \tag{A.10}$$

uniformly in m and where only w_{in} depends on i.

Turning to the expression for $\overline{\mathcal{M}}_n$, we note that $\mathsf{E} \sum_{i=1}^n \mathsf{E}_{i-1} |w_{in}| = \mathrm{O}(n^{1/2})$ as argued in point 2.2 using condition (ii). Further, condition (i) shows that $\sup_{v \in \mathbb{R}} \mathsf{f}(v) < \infty$. In combination, we get that $\mathsf{E} \overline{\mathcal{M}}_n = \delta \mathrm{O}(1)$ where the $\mathrm{O}_{\mathsf{P}}(1)$-term does not depend on δ. Thus, by the Markov inequality, $\overline{\mathcal{M}}_n = \delta \mathrm{O}_{\mathsf{P}}(1)$.

To show $\overline{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$ we need to show that for any $\gamma > 0$ then $\mathsf{P}(\overline{\mathcal{M}}_n > \gamma)$ vanishes for large n. We are still free to choose δ which will be exploited now. Since $\overline{\mathcal{M}}_n = \delta \mathrm{O}_{\mathsf{P}}(1)$, we can find a constant C not depending on δ so that $\overline{\mathcal{M}}_n \leq \delta C$ with large probability. Choosing $\delta = \gamma / C$ we get $\overline{\mathcal{M}}_n \leq \gamma$ with large probability. Hence, $\overline{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$.

3.6. The martingale $\widetilde{\mathcal{M}}^c_n$. We show $\widetilde{\mathcal{M}}^c_n = \mathrm{o}_{\mathsf{P}}(1)$, using Theorem A.4. We set $\nu_c = 1/2$ and index $\ell = m$ and consider $z_{i\ell,c} = z_{im} = 1_{(c_{im} < \chi_i \leq \overline{c}_{im})}$ defined in (A.9), which is $\mathcal{F}_{i+1}$-adapted. Note that $0 \leq z_{i\ell,c} \leq 1$. We verify the conditions of Theorem A.4.

The parameter λ_c is $(1/4 - \eta)(\dim x + 1)$ as in point 2.2.

The parameter ς_c is $1/2$. We show that $\mathcal{E}_c = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_i z^2_{i\ell,c} = \mathrm{O}(n^{\varsigma_c})$. Since $z^2_{im} = z_{im}$ while $\mathsf{E}_i z_{im} = \mathsf{E}_i \mathsf{E}_{i-1} z_{im}$ we have that $\mathcal{E}_c = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_i \mathsf{E}_{i-1} z_{im}$.

Applying the bound to $\mathsf{E}_{i-1}z_{im}$ in (A.10), which is uniform in m, we see that $\mathcal{E}_c$ has the same bound as $n^{1/2}\mathsf{E}\overline{\mathcal{M}}_n$. Now, use that $\mathsf{E}\overline{\mathcal{M}}_n = \mathrm{O}(1)$ by point 3.5.

The condition $\varsigma_c < 2\nu_c$. Since $0 < \eta$ and $\nu_c = 1/2$ we have $\varsigma_c = 1/2 < 1 = 2\nu_c$.

3.7. The martingale $\widetilde{\mathcal{M}}_n^d$. We show $\widetilde{\mathcal{M}}_n^d = \mathrm{o}_{\mathsf{P}}(1)$ using Theorem A.4. We set $\nu_d = 1/2$ and index $\ell = m$ and consider $z_{i\ell,d} = \mathsf{E}_i z_{im}$, which is $\mathcal{F}_i$-adapted. Note that $0 \le z_{i\ell,c} \le 1$. We verify the conditions of Theorem A.4.

The parameter λ_c is $(1/4-\eta)(\dim x + 1)$ as in point 2.2.

The parameter ς_d is 1/2. We show that $\mathcal{E}_d = \mathsf{E}\max_{1\le\ell\le L}\sum_{i=1}^n \mathsf{E}_{i-1}z_{i\ell,d}^2 = \mathrm{O}_{\mathsf{P}}(n^{\varsigma_d})$. Note that $z_{i\ell,d}^2 = \mathsf{E}_i^2 z_{im} \le \mathsf{E}_i z_{im}^2 = \mathsf{E}_i z_{im}$ by Jensen's inequality. Since $\mathcal{F}_{i-1} \subset \mathcal{F}_i$, we get $\mathsf{E}_{i-1}z_{i\ell,d}^2 \le \mathsf{E}_{i-1}\mathsf{E}_i z_{im} = \mathsf{E}_{i-1}z_{im}$. Thus, $\mathcal{E}_d = n^{1/2}\mathsf{E}\overline{\mathcal{M}}_n$, where $\mathsf{E}\overline{\mathcal{M}}_n = \mathrm{O}(1)$ by point 3.5.

The condition $\varsigma_d < 2\nu_d$ is satisfied as in point 2.2, since $(\nu_d, \lambda_d, \varsigma_d) = (\nu_c, \lambda_c, \varsigma_c)$.

4. Conclusion. We have shown that $\widetilde{\mathcal{M}}_n^c = \widetilde{\mathcal{M}}_n^d = \overline{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$ so that $\mathcal{R}_{n,2} = \mathrm{o}_{\mathsf{P}}(1)$. In point 2 it was shown that $\mathcal{R}_{n,1} = \mathrm{o}_{\mathsf{P}}(1)$. In combination, $\mathcal{R}_n = \mathrm{o}_{\mathsf{P}}(1)$. □

A.5 The compensator

We provide a linearization of the pseudo-compensator $\overline{\mathsf{F}}_n$ defined in (A.3).

Theorem A.8. *Suppose Assumption 3.1 holds and that*
(i) the marginal density f *has bounded derivative:* $\sup_{v\in\mathbb{R}}(1+v^2)|\dot{\mathsf{f}}(v)| < \infty$*;*
(ii) the conditional density $\mathsf{m}_i(y|x)$ *of* χ_i *given* ∇x_i *and* $\mathcal{F}_{i-1}$ *exists, it is differentiable in* y *and satisfies* $\max_{1\le i\le n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}(1+|y|)\,|\dot{\mathsf{m}}_i(y|x)| < \infty$*;*
(iii) the regressors x_i *satisfy* $\mathsf{E}\sum_{i=1}^n |\nabla x_{in}|^2 = \mathrm{O}(1)$*.*
Let $\xi_n = n^{-1/2}\sum_{i=1}^n \mathsf{E}_{i-1}(\nabla x_{in}|\chi_i = c)$*. Then, for all* $B > 0$ *and* $0 < \eta < 1/4$

$$\sup_{|a|,|b|\le n^{1/4-\eta}B}\ \sup_{c\in\mathbb{R}} |n^{1/2}\{\overline{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(0,0,c)\} - \mathsf{h}(c)\{ac + b'\xi_n(c)\}| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}).$$

Proof of Theorem A.8. The quantity of interest is

$$Q_n(a,b,c) = n^{1/2}\{\overline{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(0,0,c)\} - \mathsf{h}(c)\{ac + b'\xi_n(c)\}.$$

Define $c_a = c + n^{-1/2}ac$ and note that $\overline{\mathsf{F}}_n(a,b,c) = \overline{\mathsf{F}}_n(0,b,c_a)$. Add and subtract $\mathsf{h}(c_a)b'\xi_n(c_a)$ and $n^{1/2}\overline{\mathsf{F}}_n(a,0,c) = n^{1/2}\overline{\mathsf{F}}_n(0,0,c_a)$ to Q_n to get

$$Q_n(a,b,c) = Q_n(0,b,c_a) + Q_n(a,0,c) + R_n(a,b,c), \tag{A.11}$$

where

$$R_n(a,b,c) = \mathsf{h}(c_a)b'\xi_n(c_a) - \mathsf{h}(c)b'\xi_n(c). \tag{A.12}$$

We show that each of the terms of the right hand side of (A.11) vanish uniformly in a, b, c.

1. The term $Q_n(0,b,c_a)$**.** We note that $\sup_{a,b,c}|Q_n(0,b,c_a)| = \sup_{b,c}|Q_n(0,b,c)|$ and consider the latter. Write $Q_n(0,b,c) = n^{-1/2}\sum_{i=1}^n q_i(b,c)$, where

$$q_i(b,c) = \mathsf{E}_{i-1}\{1_{(\chi_i - b'\nabla x_{in}\le c)} - 1_{(\chi_i\le c)}\} - \mathsf{h}(c)b'\mathsf{E}_{i-1}(\nabla x_{in}|\chi_i = c).$$

Here, q_i is an expression of the form considered in Theorem A.1(a), noting that the expectation of an indicator is a probability. The probability measure P in the theorem is a conditional measure given $\mathcal{F}_{i-1}$, while the variables are $Y = \chi_i$ and $X = \nabla x_{in}$. Moreover, $\mathsf{m}_i(y,x)$ is the joint density of Y_i, X_i conditional on $\mathcal{F}_{i-1}$. By condition (ii), $\mathsf{m}_i(y|x)$ is differentiable with respect to y. The derivative has the bound $C_\mathsf{m} = \max_{1\le i\le n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}(1+|y|)\,|\dot{\mathsf{m}}_i(y|x)| < \infty$. By condition (iii), $\mathsf{E}|X|^2$ exists. We bound $|q_i(b,c)| \le 2^{-1}|b|^2 C_\mathsf{m}\mathsf{E}_{i-1}|\nabla x_{in}|^2$, Since $|b| \le n^{1/4-\eta}B$ and using the triangular inequality we get $|Q_n(b,c)| \le \mathrm{O}(n^{-2\eta})C_\mathsf{m}\sum_{i=1}^n \mathsf{E}_{i-1}|\nabla x_{in}|^2$. By condition (iii) and the Markov inequality then $\sum_{i=1}^n \mathsf{E}_{i-1}|\nabla x_{in}|^2 = \mathrm{O}(1)$. Thus, $|Q_n(b,c)| = \mathrm{O}_\mathsf{P}(n^{-2\eta})$ uniformly in b, c.

2. The term $Q_n(a,0,c)$. Write $Q_n(a,0,c) = n^{-1/2}\sum_{i=1}^n q_i(a,c)$, where

$$q_i(a,c) = \mathsf{E}_{i-1}\{1_{(\chi_i \le c+n^{-1/2}ac)} - 1_{(\chi_i\le c)}\} - n^{-1/2}ac\mathsf{h}(c).$$

Note $\chi_i = (\varepsilon_i - \varepsilon_{i+1})/(\sqrt{2}\sigma)$ has density h and is independent of $\mathcal{F}_{i-1}$. As f is differentiable by Assumption 3.1, so is h. Thus, the Mean-Value Theorem gives

$$q_i(a,c) = \int_c^{c+n^{-1/2}ac}\mathsf{h}(u)du - n^{-1/2}ac\mathsf{h}(c) = n^{-1}a^2c^2\dot{\mathsf{h}}(\tilde{c})/2,$$

where $|\tilde{c} - c| \le |n^{-1/2}ac|$. Since $|a| \le n^{1/4-\eta}B$, then $|n^{-1/2}a| \le 1/2$ for large n. The second inequality in Lemma A.2 then shows that $a^2c^2 \le 16a^2\tilde{c}^2$. We then have $q_i(a,c) \le 16n^{-1}a^2\tilde{c}^2\dot{\mathsf{h}}(\tilde{c})$.

Condition (i) shows that $\sup_{v\in\mathbb{R}}(1+v^2)\dot{\mathsf{f}}(v) < \infty$. Theorem 2.3(d) then implies that $\sup_{v\in\mathbb{R}} v^2\dot{\mathsf{h}}(v) < \infty$. Using that $|a| \le n^{1/4-\eta}B$ we get $q_i(a,c) = \mathrm{O}(n^{-1/2-2\eta})$ uniformly in a, c, i. It follows that $|Q_n(a,c)| = \mathrm{O}(n^{-2\eta})$ uniformly in a, c.

3. The term $R_n(a,b,c)$. Write $R_n(a,b,c) = n^{-1/2}\sum_{i=1}^n q_i(a,b,c)$, where

$$q_i(a,b,c) = \mathsf{h}(c_a)b'\mathsf{E}_{i-1}(\nabla x_{in}|\chi_i = c_a) - \mathsf{h}(c)b'\mathsf{E}_{i-1}(\nabla x_{in}|\chi_i = c)$$

with $c_a = c + n^{-1/2}ac$. Apply Theorem A.1(b). The setup is as in point 1, with $Y_i = \chi_i$ and $X_i = \nabla x_{in}$, while $\mathsf{m}_i(y,x)$ denotes the joint conditional density of Y_i, X_i given $\mathcal{F}_{i-1}$, and $C_\mathsf{m} = \max_{1\le i\le n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}(1+|y|)\,|\dot{\mathsf{m}}_i(y|x)| < \infty$, using conditions (ii, iii). We bound $|q_i(a,b,c)| \le |n^{-1/2}ab|C_\mathsf{m}\mathsf{E}_{i-1}|\nabla x_{in}|$. Since $|a|,|b| \le n^{1/4-\eta}B$ and $\sum_{i=1}^n \mathsf{E}_{i-1}|\nabla x_{in}| = \mathrm{O}(1)$ by condition (iii), we get $|R_n(a,b,c)| = \mathrm{O}(n^{-2\eta})$ uniformly in a, b and c. □

A.6 The empirical distribution function

We combine Theorems A.6 and A.8 to expand the empirical distribution function.

Theorem A.9. *Suppose Assumption 3.1 holds and that*
(i) the marginal density f satisfies: $\sup_{v\in\mathbb{R}}\mathsf{f}(v) < \infty$, $\sup_{v\in\mathbb{R}}(1+v^2)|\dot{\mathsf{f}}(v)| < \infty$;
(ii) the conditional density $\mathsf{m}_i(y|x)$ of χ_i given ∇x_i and $\mathcal{F}_{i-1}$ exists, it is differentiable in y and satisfies $\max_{1\le i\le n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}(1+|y|)\,|\dot{\mathsf{m}}_i(y|x)| < \infty$;
(iii) the regressors x_i satisfy $\mathsf{E}\sum_{i=1}^n|\nabla x_{in}|^2 = \mathrm{O}(1)$;
Let $\xi_n = n^{-1/2}\sum_{i=1}^n\mathsf{E}_{i-1}(\nabla x_{in}|\chi_i = c)$. Then, for all $B > 0$, $0 < \eta < 1/4$, $c \in \mathbb{R}$, and uniformly in $|a|,|b| \le n^{1/4-\eta}B$, we have

$$\sqrt{n}\{\widehat{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(0,0,c)\} = \mathbb{F}_n(0,0,c) + \mathsf{h}(c)\{ac + b'\xi_n(c)\} + \mathrm{o}_\mathsf{P}(1)$$

Proof of Theorem A.9. The conditions of Theorems A.6 and A.8 are directly listed in the present conditions apart from condition *(ii)* of Theorem A.6. The latter follows from the present condition *(iii)* by Jensen's inequality in that $n^{-1/2}\mathsf{E}\sum_{i=1}^{n}|\nabla x_{in}| \leq (\mathsf{E}\sum_{i=1}^{n}|\nabla x_{in}|^2)^{1/2} = \mathrm{O}(1)$. Rewrite $R_n = n^{1/2}\{\widehat{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(0,0,c)\}$ as

$$\begin{aligned} R_n &= n^{1/2}[\{\widehat{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(a,b,c)\} - \{\widehat{\mathsf{F}}_n(0,0,c) - \overline{\mathsf{F}}_n(0,0,c)\}] \\ &\quad +[n^{1/2}\{\overline{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(0,0,c)\} - \mathsf{h}(c)\{ac + b'\xi(c)\}] \\ &\quad +n^{1/2}\{\widehat{\mathsf{F}}_n(0,0,c) - \overline{\mathsf{F}}_n(0,0,c)\} + \mathsf{h}(c)\{ac + b'\xi(c)\}. \end{aligned}$$

By Theorems A.6 and A.8, the terms in square brackets is $\mathrm{o}_{\mathsf{P}}(1)$ uniformly in $|a|, |b|$. □

A.7 Results for stylized SIS

Lemma A.10. *Suppose Assumption 3.1, 3.2(ia, iiic). Let $\eta > 0$ be given. Let σ_n be a sequence of random variables so that $n_j^{1/2}(\sigma_n^2 - \sigma^2) = \mathrm{O}_{\mathsf{P}}(n_j^{1/4-\eta})$. Let V_n, Σ_n^{-1} be sequences of random vectors and square matrices which are $\mathrm{O}_{\mathsf{P}}(1)$. Let M_n be a sequence of deterministic square matrices satisfying $M_n = \mathrm{O}(n_j^{1/4-\eta})$ for some $\eta > 0$. All those vectors and square matrices have the same dimension as x_i. Let $w_{in}^2 = 1 + (\nabla x_{in})' M_n \Sigma_n^{-1} M_n' \nabla x_{in}$ and*

$$D_i = 1_{(|\nabla\varepsilon_i - V_n'\Sigma_n^{-1}M_n'\nabla x_{in}| > \sqrt{2}\sigma_n w_{in} c)} - 1_{(|\nabla\varepsilon_i - V_n'\Sigma_n^{-1}M_n'\nabla x_{in}| > \sqrt{2}\sigma_n c)}.$$

Then $\sum_{i\in I_j} D_i = \mathrm{o}_{\mathsf{P}}(n_j^{1/2})$.

Proof of Lemma A.10. By definition $w_{in}^2 \geq 1$, so that $1 \leq w_{in} \leq w_{in}^2$. Thus,

$$0 \leq D_i = 1_{(\sqrt{2}\sigma_n c < |\nabla\varepsilon_i - V_n'\Sigma_n^{-1}M_n'\nabla x_{in}| \leq \sqrt{2}\sigma_n w_{in} c)} \leq 1_{(\sqrt{2}\sigma_n c < |\nabla\varepsilon_i - V_n'\Sigma_n^{-1}M_n'\nabla x_{in}| \leq \sqrt{2}\sigma_n w_{in}^2 c)}.$$

Further, use the spectral norm as matrix norm. As this is sub-multiplicative, we can bound $w_{in}^2 - 1 \leq |\nabla x_{in}|^2 \|M_n\|^2 \|\Sigma_n^{-1}\|$ and get

$$0 \leq D_i \leq 1_{(\sqrt{2}\sigma_n c < |\nabla\varepsilon_i - V_n'\Sigma_n^{-1}M_n'\nabla x_{in}| \leq \sqrt{2}\sigma_n c + \sqrt{2}\sigma_n |\nabla x_{in}|^2 \|M_n\|^2 \|\Sigma_n^{-1}\| c)}.$$

Introduce the empirical distribution function and pseudo-compensator

$$\widehat{\mathsf{H}}_n(a, b_1, b_2, c) = n_j^{-1}\sum_{i\in I_j} 1_{(\chi_i \leq c + n_j^{-1/2}ac + b_1' z_{1i} + b_2 z_{2i} c)}, \tag{A.13}$$

$$\overline{\mathsf{H}}_n(a, b_1, b_2, c) = n_j^{-1}\sum_{i\in I_j} \mathsf{E}_{i-1} 1_{(\chi_i \leq c + n_j^{-1/2}ac + b_1' z_{1i} + b_2 z_{2i} c)}. \tag{A.14}$$

where $\chi_i = \nabla\varepsilon_i/(\sqrt{2}\sigma)$ as in (19), $z_i = (z_{1i}', z_{2i})' = \{(\nabla x_{in})', n_j^{1/2}|\nabla x_{in}|^2\}'$ is $\mathcal{F}_i$-adapted, and $\widehat{a} = n_j^{1/2}(\sigma_n/\sigma - 1)$, $\widehat{b}_1' = V_n'\Sigma_n^{-1}M_n'/\sqrt{2}$ and $\widehat{b}_2 = (\sigma_n/\sigma)n_j^{-1/2}\|M_n\|^2\|\Sigma_n^{-1}\|$. Then, noting that χ_i has a continuous distribution, we get with probability one

$$n_j^{-1}\sum_{i\in I_j} D_i \leq \widehat{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, \widehat{b}_2 c, c) - \widehat{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, 0, c) - \widehat{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, -\widehat{b}_2 c, -c) + \widehat{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, 0, -c).$$

The empirical distribution function $\widehat{\mathsf{H}}_n$ has the same structure as $\widehat{\mathsf{F}}_n$ defined in (A.2), albeit with $b'\nabla x_{in}$ replaced by $(b_1', b_2 c)z_i$. We note that $\widehat{a} = \mathrm{O}_\mathsf{P}(n_j^{1/4-\eta})$, $\widehat{b}_1 = \mathrm{O}_\mathsf{P}(n_j^{1/4-\eta})$, $\widehat{b}_2 = \mathrm{O}_\mathsf{P}(n_j^{-2\eta})$, so that all are $\mathrm{O}_\mathsf{P}(n_j^{1/4-\eta})$. Thus, we can on a set with large probability find a large $B > 0$ so that $|\widehat{a}|, |\widehat{b}_1| \leq Bn_j^{1/4-\eta}$ and $|\widehat{b}_2| \leq Bn_j^{-2\eta} \leq Bn_j^{1/4-\eta}$ uniformly in n_j. We can then apply the expansion in Theorem A.6 to replace $\widehat{\mathsf{H}}_n$ by $\overline{\mathsf{H}}_n$, while substituting $\widehat{a}, \widehat{b}_1, \widehat{b}_2$ for a, b_1, b_2 and using Assumption 3.1, 3.2($ia, iiic$). Thus,

$$0 \leq n_j^{-1} \sum_{i \in I_j} D_i \leq \overline{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, \widehat{b}_2 c, c) - \overline{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, 0, c) \\ - \overline{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, -\widehat{b}_2 c, -c) + \overline{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, 0, -c) + \mathrm{o}_\mathsf{P}(n_j^{-1/2}). \quad \text{(A.15)}$$

Finally, we show that $\overline{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, \widehat{b}_2 c, c) - \overline{\mathsf{H}}_n(\widehat{a}, \widehat{b}_1, 0, c) = \mathrm{o}_\mathsf{P}(n_j^{-1/2})$ for fixed $c \in \mathbb{R}$. For this, it suffices to show that

$$\sup_{|a|,|b_1| \leq Bn_j^{1/4-\eta}} \; \sup_{0 \leq b_2 \leq Bn_j^{-2\eta}} |\overline{\mathsf{H}}_n(a, b_1, b_2 c, c) - \overline{\mathsf{H}}_n(a, b_1, 0, c)| = \mathrm{o}_\mathsf{P}(n_j^{-1/2}). \quad \text{(A.16)}$$

Now, $\overline{\mathsf{H}}_n(a, b_1, b_2 c, c) - \overline{\mathsf{H}}_n(a, b_1, 0, c) = n_j^{-1} \sum_{i \in I_j} \mathsf{E}_{i-1} h_{in}(a, b_1, b_2, c)$ where

$$h_{in}(a, b_1, b_2, c) = 1_{\{\chi_i \leq c + n_j^{-1/2} ac + (b_1', b_2 c) z_i\}} - 1_{\{\chi_i \leq c + n_j^{-1/2} ac + (b_1', 0) z_i\}}$$

has the same sign as c. We get $\mathsf{E}_{i-1} h_{in}(a, b_1, b_2, c) = \mathsf{E}_{i-1}\{\mathsf{E}_i h_{in}(a, b_1, b_2, c)\}$ by iterated expectations. Recall that $\chi_i = (\varepsilon_i - \varepsilon_{i+1})/(\sqrt{2}\sigma)$, so that all elements of h_{in} but ε_{i+1} are $\mathcal{F}_i$-measurable while ε_{i+1} is independent thereof. Thus, we can write the E_i-expectation as an integral and use the Mean Value Theorem to get, for an intermediate point c^*,

$$\mathsf{E}_i h_{in}(a, b_1, b_2, c) = \int_{\varepsilon_i/\sigma - \sqrt{2}\{c + n_j^{-1/2} ac + (b_1', b_2 c) z_i\}}^{\varepsilon_i/\sigma - \sqrt{2}\{c + n_j^{-1/2} ac + (b_1', 0) z_i\}} \mathsf{f}(u) du = \sqrt{2}(0, b_2)' z_i c \mathsf{f}(c^*).$$

Now, $(0, b_2)' z_i = b_2 n_j^{1/2} |\nabla x_{in}|^2$, where $0 \leq b_2 \leq Bn_j^{-2\eta}$ by the construction (A.16). Further, c is fixed, while $\mathsf{f}(c^*) \leq \max_{v \in \mathbb{R}} \mathsf{f}(v)$ which is finite by Assumption 3.2(ia). Thus, we find $|\mathsf{E}_i h_{in}(a, b_1, b_2, c)| \leq C n_j^{1/2-2\eta} |\nabla x_{in}|^2$ for some constant $C > 0$, uniformly in a, b_1, b_2.

We note that h_{in} and hence the $\overline{\mathsf{H}}_n$-differences has the same sign as c so that

$$\mathcal{H}_n = \mathsf{E}|\overline{\mathsf{H}}_n(a, b_1, b_2, c) - \overline{\mathsf{H}}_n(a, b_1, 0, c)| = |\mathsf{E}\{\overline{\mathsf{H}}_n(a, b_1, b_2, c) - \overline{\mathsf{H}}_n(a, b_1, 0, c)\}|.$$

Writing out in terms of the h_{in} functions and using iterated expectations, we get

$$\mathcal{H}_n = |\mathsf{E} n_j^{-1} \sum_{i \in I_j} \mathsf{E}_{i-1} \mathsf{E}_i h_{in}(a, b_1, b_2, c)| = |\mathsf{E} n_j^{-1} \sum_{i \in I_j} \mathsf{E}_i h_{in}(a, b_1, b_2, c)|.$$

Thus, uniformly in a, b_1, b_2

$$\mathcal{H}_n \leq \mathsf{E} n_j^{-1} \sum_{i \in I_j} C n_j^{1/2-\eta} |\nabla x_{in}|^2 = \mathrm{O}(n_j^{-1/2-\eta})$$

by Assumption 3.2($iiia$), so that (A.16) follows. The desired result follows. □

Proof of Theorem 3.3. **1. The OLS estimator on the first sample.** Normalizing the OLS estimator $\hat{\beta}_1$ in (12) gives $N_1^{-1}(\hat{\beta}_1 - \beta) = \widehat{\Sigma}_{1n}^{-1}\widehat{V}_{1n}$ when using the following notation from (23), (24)

$$\widehat{\Sigma}_{1n}^{-1} = \sum_{i \in I_1} N_1'(x_i - \bar{x}_1)(x_i - \bar{x}_1)' N_1, \quad \widehat{V}_{1n} = \sum_{i \in I_1} N_1'(x_i - \bar{x}_1)(\varepsilon_i - \mathsf{E}\varepsilon_i),$$

while $\bar{x}_1 = n_1^{-1} \sum_{i \in I_1} x_i$, where the expectation can be subtracted from ε_i since the regressors are demeaned. By Assumption 3.2($iiia, b$), $N_1^{-1}(\hat{\beta}_1 - \beta) = \hat{\Sigma}_{1n}^{-1}\hat{V}_{1n} = \mathrm{O}_{\mathsf{P}}(1)$. Similarly, the normalized estimator for the residual variance in (23) is

$$n_1^{1/2}(\hat{\sigma}_1^2 - \sigma^2) = n_1^{-1/2} \sum_{i \in I_1} \{(\varepsilon_i - \bar{\varepsilon}_1)^2 - \sigma^2\} - n_1^{-1/2} \widehat{V}_{1n}' \widehat{\Sigma}_{1n}^{-1} \widehat{V}_{1n},$$

where $\bar{\varepsilon}_1 = n_1^{-1} \sum_{i \in I_1} \varepsilon_i$. By Assumption 3.1, the innovations ε_i are i.i.d. The first term converges in distribution by the Central Limit Theorem. The second vanishes as $\widehat{\Sigma}_{1n}$, $\widehat{V}_{1n}$ converge in distribution by Assumption 3.2($iiia, b$) while the factor $n_1^{-1/2}$ vanishes. Therefore, the estimators are converging and bounded with $\mathrm{O}_{\mathsf{P}}(1)$.

2. Apply Lemma A.10 with $j = 2$ and $\omega_{in} = \omega_{1,i}$. Since $(n_2/n_1)^{1/2} = \mathrm{o}(n_2^{1/4-\eta})$ by Assumption 3.2(iv), we get $n_2^{1/2}(\sigma_n^2 - \sigma) = (n_2/n_1)^{1/2} n_1^{1/2}(\hat{\sigma}_1^2 - \sigma^2) = \mathrm{o}_{\mathsf{P}}(n_2^{1/4-\eta})$. Note, that $\widehat{V}_{1n}$ and $\Sigma_n^{-1} = \widehat{\Sigma}_{1n}^{-1}$, are both $\mathrm{O}_{\mathsf{P}}(1)$. Let $M_n = N_2^{-1} N_1 = \mathrm{o}(n_2^{1/4-\eta})$ by Assumption 3.2(iv). Lemma A.10 also requires Assumption 3.2($ia, iiic$). □

Proof of Theorem 3.4. It was shown in step 1 of the proof of Theorem 3.3 that $N_1^{-1}(\hat{\beta}_1 - \beta), n_1^{1/2}(\hat{\sigma}_1^2 - \sigma^2) = \mathrm{O}_{\mathsf{P}}(1)$. This uses Assumptions 3.1, 3.2($iiia, b$). We rewrite the gauge and apply Theorem A.9.

1. Rewriting expression for the gauge. The gauge is defined in (16). Due to Theorem 3.3 with Assumptions 3.1, 3.2(ia, iii, iv) we can set $w_{i,1} = 1$ and we ignore the resulting remainder term. Forward difference equation (7) to get $\nabla y_i = \beta' \nabla x_i + \nabla \varepsilon_i$. With that, rewrite the gauge as

$$\hat{\gamma}_n = \frac{1}{n_2^\circ} \sum_{i \in I_2^\circ} 1_{(|\nabla y_i - \hat{\beta}_1' \nabla x_i| \geq \sqrt{2}\hat{\sigma}_1 c)} = \frac{1}{n_2^\circ} \sum_{i \in I_2^\circ} 1_{\{|\nabla \varepsilon_i - (\hat{\beta}_1 - \beta)' \nabla x_i| \geq \sqrt{2}\hat{\sigma}_1 c\}}.$$

We normalize $\hat{\beta}_1 - \beta$ by N_1 and x_i by N_2. At the same time we divide through by $\sqrt{2}\sigma$. Recall the notation $\chi_i = \varepsilon_i / (\sqrt{2}\sigma)$ and $x_{in} = N_2' x_i$ for $i \in I_2$. Define

$$\hat{a}_1 = (n_2/n_1)^{1/2} n_1^{1/2}(\hat{\sigma}_1/\sigma - 1), \quad \hat{b}_1 = (N_2^{-1} N_1) N_1^{-1}(\hat{\beta}_1 - \beta)/(\sqrt{2}\sigma), \tag{A.17}$$

so that $\hat{a}_1, \hat{b}_1 = \mathrm{O}_{\mathsf{P}}(n_2^{1/4-\eta})$ by the convergence results in point 1 of the proof of Theorem 3.3 and Assumption 3.2(iv). We then get the two-sided empirical distribution function

$$\hat{\gamma}_n = \frac{1}{n_2^\circ} \sum_{i \in I_2^\circ} 1_{\{|\chi_i - \hat{b}_1' \nabla x_{in}| \geq c + n_2^{-1/2} \hat{a}_1 c\}}. \tag{A.18}$$

2. Apply empirical process result. Theorem A.9 expands

$$n^{-1/2}\sum_{i=1}^{n}\{1_{(\chi_i\leq c+n^{-1/2}ac+b'\nabla x_{in})}-\mathsf{E}1_{(\chi_i\leq c)}\} = n^{-1/2}\sum_{i=1}^{n}\{1_{(\chi_i\leq c)}-\mathsf{E}1_{(\chi_i\leq c)}\} + \mathsf{h}(c)\{ac+b'\xi_n(c)\}+\mathrm{o}_{\mathsf{P}}(1),$$

uniformly in $|a|,|b|\leq n^{1/4-\eta}B$ for all $\eta, B>0$. As remarked above, this expansion will be used for observations in the second sub-sample. Thus, the conditions of Theorem A.9 are satisfied by Assumption 3.2 $(i, ii, iiic)$. Since $1_{(|\chi_i|\geq c)}=1-1_{(\chi_i<c)}+1_{(\chi_i\leq -c)}$ and noting that χ_i is continuously distributed we can form a two sided version

$$n^{-1/2}\sum_{i=1}^{n}\{1_{(|\chi_i|\geq c+n^{-1/2}ac+b'\nabla x_{in})}-\mathsf{E}1_{(|\chi_i|\geq c)}\} = n^{-1/2}\sum_{i=1}^{n}\{1_{(|\chi_i|\geq c)}-\mathsf{E}1_{(|\chi_i|\geq c)}\} + \mathcal{B}_n(a,b,c)+\mathrm{o}_{\mathsf{P}}(1). \quad (A.19)$$

with the bias term $\mathcal{B}_n(a,b,c)=-\mathsf{h}(c)\{ac+b'\xi_n(c)\}+\mathsf{h}(-c)\{-ac+b'\xi_n(-c)\}$. Theorem 2.3 shows that h is symmetric: $\mathsf{h}(c)=\mathsf{h}(-c)$. Thus, the bias term satisfies

$$\mathcal{B}_n(a,b,c)=-2c\mathsf{h}(c)a-\mathsf{h}(c)b'\{\xi_n(c)-\xi_n(-c)\}.$$

We now apply the expansion (A.19) to the expression for the gauge in (A.18) with two adjustments. First, the gauge in (A.18) depends on estimators $\hat{a}_1,\hat{b}_1$. These are $\mathrm{O}_{\mathsf{P}}(n_2^{1/4-\eta})$ as remarked above. Thus, on a set with large probability a large $B>0$ exists so that $|\hat{a}_1|,|\hat{b}_1|<Bn_2^{1/4-\eta}$ uniformly in n. Since the expansion in (A.19) is uniform in $|a_1|,|b_1|<Bn^{1/4-\eta}$, we can apply the expansion in (A.19) while substituting $\hat{a}_1,\hat{b}_1$ for a_1,b_1. Second, we will need to change the index of the observations, as the expansion in (A.18) is concerned with indices, $i\in I_2^\circ$ while the expansion in (A.19) has indices $i=1,\ldots n$.

Thus, defining $\gamma=\mathsf{E}1_{(|\chi_i|\geq c)}$ and $\xi_{2n}=n_2^{-1/2}\sum_{i\in I_2}\mathsf{E}_{i-1}(\nabla x_{in}|\chi_i=c)$, while noting $n_2^\circ/n_2\to 1$, we get the desired expansion (25).

3. Consistency. The terms on the right hand side of expansion (25) are $\mathrm{o}_{\mathsf{P}}(n_2^{1/2})$ under the stated conditions. This gives the convergence in probability. As the gauge is bounded by unit, this extends to convergence in mean (Billingsley, 1968, p. 32). □

A.8 Results for split-half SIS

Proof of Theorem 4.2. Define gauges for each of the sub-samples as

$$\hat{\gamma}_{1n}=\frac{1}{n_1^\circ}\sum_{i\in I_1^\circ}1_{(|\nabla y_i-\hat{\beta}_2'\nabla x_i|\geq\sqrt{2}\hat{\sigma}_2\omega_{2,i}c)}, \qquad \hat{\gamma}_{2n}=\frac{1}{n_2^\circ}\sum_{i\in I_2^\circ}1_{(|\nabla y_i-\hat{\beta}_1'\nabla x_i|\geq\sqrt{2}\hat{\sigma}_1\omega_{1,i}c)},$$

noting that $n^\circ\hat{\gamma}_n^{split}=n^\circ\hat{\gamma}_{1n}+n^\circ\hat{\gamma}_{2n}$. As in the proof of Theorem 3.3, we can apply Lemma A.10 and set $\omega_{j,i}=1$ and ignore the resulting remainder terms.

Apply Theorem 3.4 to each of $\hat{\gamma}_{1n}$, $\hat{\gamma}_{2n}$ noting that its derivation does not depend on the ordering of two sub-samples. This requires Assumptions 3.1, 4.1. We get expansions, with $(j,k)=(1,2)$ or $(j,k)=(2,1)$,

$$n_j^{1/2}(\hat\gamma_{jn}-\gamma) = n_j^{-1/2}\sum_{i\in I_j^\circ}\{1_{(|\chi_i|\geq c)}-\mathsf{E}1_{(|\chi_i|\geq c)}\} - c\mathsf{h}(c)(n_j/n_k)^{1/2}n_k^{-1/2}\sum_{i\in I_k}(\varepsilon_i^2/\sigma^2-1)$$
$$-\mathsf{h}(c)\{\xi_{jn}(c)-\xi_{jn}(-c)\}'N_j^{-1}N_k\widehat\Sigma_{kn}^{-1}\widehat V_{kn}/(\sqrt{2}\sigma)+\mathsf{o_P}(1).$$

Use that $n_j^\circ/n_j\to 1$ and $n_1^\circ+n_2^\circ=n^\circ$ and insert the Theorem 3.4 sub-sample expansions into $(n^\circ)^{1/2}(\hat\gamma_n^{split}-\gamma)=(n_1^\circ/n^\circ)^{1/2}(n_1^\circ)^{1/2}(\hat\gamma_{1n}-\gamma)+(n_2^\circ/n^\circ)^{1/2}(n_2^\circ)^{1/2}(\hat\gamma_{2n}-\gamma)$. Asymptotically, we can replace n_j°, I_j° with n_j, I_j. □

The proof of Theorem 4.4 uses the following non-stationary mixingale result

Lemma A.11. *(McLeish, 1977, Theorem 2.4) Let X_{ni} for $i,n=1,2,\dots$ be a double array of zero mean random variables. Let $k_n(t)$ be a sequence of nonrandom integer valued, nondecreasing, right continuous functions on $[0,\infty)$. Suppose a double array of constants $\sigma_{ni}^2>0$ exists such that for each $T<\infty$:*
(a) $\sup_{s<t<T}\limsup_{n\to\infty}\sum_{k_n(s)}^{k_n(t)}\sigma_{ni}^2/(t-s)<\infty$;
(b) $\{X_{ni}^2/\sigma_{ni}^2; n=1,2,\dots,i\leq k_n(T)\}$ is a uniformly integrable set;
(c) $\max_{i\leq k_n(T)}\sigma_{ni}\to 0$ as $n\to\infty$.
(d) $\mathsf{E}|\mathsf{E}\{(\sum_{i=k_n(t)}^{k_n(u)}X_{ni})^2\mid\mathcal{F}_{n,k_n(s)}\}-(u-t)|\to 0$ as $n\to\infty$ for each $s<t<u$.
Further, X_{ni} is a mixingale with respect to σ-fields $\mathcal{F}_{ni}$ that are nondecreasing in i and a vanishing sequence of constants $\psi_n>0$ so that for all $n,i,k+1\geq 1$ then
(e) $\mathsf{E}\{\mathsf{E}(X_{ni}\mid\mathcal{F}_{n,i-k})\}^2\leq\psi_k^2\sigma_{ni}^2$;
(f) $\mathsf{E}\{X_{ni}-\mathsf{E}(X_{ni}\mid\mathcal{F}_{n,i+k})\}^2\leq\psi_{k+1}^2\sigma_{ni}^2$;
(g) $\sum_{k=1}^\infty(\sum_{n=0}^k\psi_n^{-2})^{-1/2}<\infty$, which is satisfied when $\sum_{k=1}^\infty\psi_k<\infty$,
Then, $W_n(t)=\sum_{i=1}^{k_n(t)}X_{ni}$ converges weakly to a standard Wiener process in the Stone (1963) topology on the space of right continuous function with left limits, $D[0,\infty)$.

Proof of Theorem 4.4. We will rewrite the Theorem 4.2 expansion as

$$\sqrt{n^\circ}(\hat\gamma_n^{split}-\gamma)=\frac{1}{\sqrt{n^\circ}}\sum_{j=1}^2\sum_{i\in I_j^\circ}(d_{jn}'s_i-d_{3jn}'v_{in})+\mathsf{o_P}(1), \tag{A.20}$$

where s_i is a stationary mixingale, d_{jn} and its third component d_{3jn} are deterministic with different levels for the two sample periods, and v_{in} is a residual term. We will apply Lemma A.11 to the sum of $X_{ni}=d_{jn}'s_i/\sqrt{n^\circ}$ in (A.20) with filtration $\mathcal{F}_{ni}=\mathcal{F}_i$ as defined in Assumption 3.1. We then argue that $\sum_{j=1}^2\sum_{i\in I_j^\circ}d_{3jn}v_{in}/\sqrt{n^\circ}$ vanishes.

1. Mixingale expansion and notation. The expansion in Theorem 4.2 is

$$\sqrt{n}(\hat\gamma_n^{split}-\gamma)=n^{-1/2}\sum_{i=1}^{n-1}\{1_{(|\chi_i|\geq c)}-\gamma\}$$
$$-c\mathsf{h}(c)n^{-1/2}\sum_{i=1}^n\Big\{n_2n_1{}^{-1}1_{(i\in I_1)}+n_1n_2{}^{-1}1_{(i\in I_2)}\Big\}(\varepsilon_i^2\sigma^{-2}-1)$$
$$-\mathsf{h}(c)(\sqrt{2}\sigma)^{-1}\Big[n_1^{1/2}n^{-1/2}\{\xi_{1n}(c)-\xi_{1n}(-c)\}'N_1^{-1}N_2\widehat\Sigma_{2n}^{-1}\widehat V_{2n}$$
$$+n_2^{1/2}n^{-1/2}\{\xi_{2n}(c)-\xi_{2n}(-c)\}'N_2^{-1}N_1\widehat\Sigma_{1n}^{-1}\widehat V_{1n}\Big]+\mathsf{o_P}(1),$$

for fixed $c \in \mathbb{R}$. By Assumptions 4.1(iii, a, b), $\widehat{\Sigma}_{jn}^{-1} \to \Sigma_j$ and $\widehat{V}_{jn} \to V_j$ in distribution.

We define a stationary mixingale component

$$s_i = \begin{pmatrix} s_{1i} \\ s_{2i} \\ s_{3i} \end{pmatrix} = \begin{pmatrix} 1_{(|\chi_i| \geq c)} - \gamma \\ \varepsilon_i^2/\sigma^2 - 1 \\ \Sigma_x^{-1}(x_i - \mu_x)(\varepsilon_i/\sigma - \varkappa_1) \end{pmatrix}.$$

By Assumption 4.3(ii) the pairs x_i, ε_i are stationary with $\mu_x = \mathsf{E}x_i$ and $\Sigma_x = \mathsf{Var}(x_i)$ so that $\Sigma_j = \Sigma_x$. Here s_i is $\mathcal{F}_{i+1}$-adapted, while $\mathsf{E}_{i-1}s_i = 0$, since x_i is $\mathcal{F}_{i-1}$-adapted.

Define the deterministic component. The vector takes on two distinct values with $j = 1, 2$ for the observations within the two sub-samples I_j:

$$d_{jn} = \begin{pmatrix} d_{1jn} \\ d_{2jn} \\ d_{3jn} \end{pmatrix} = \begin{pmatrix} 1 \\ -c\mathsf{h}(c)n_k n_j^{-1} \\ -\mathsf{h}(c)\xi_c n_k n_j^{-1} 2^{-1/2} \end{pmatrix},$$

where (k, j) for $i \in I_1$ is $(2, 1)$ and for $i \in I_2$ is $(1, 2)$. For ξ_c, we used that as the pairs x_i, ε_i are stationary then $\mathsf{E}_{i-1}(\nabla x_i \mid \chi_i = c)$ is deterministic and constant in i, and thus

$$\xi_c = \xi_{jn}(c) - \xi_{jn}(-c) = \mathsf{E}_0(\nabla x_1 \mid \chi_1 = c) - \mathsf{E}_0(\nabla x_1 \mid \chi_1 = -c), \text{ for } j = 1, 2.$$

We note that d_{jn} is finite since n_k/n_j has a positive and finite limit, while Assumption 4.3(i) has that

$|c|\mathsf{f}(c)$ is bounded, hence $|c|\mathsf{h}(c)$ is bounded by Theorem 2.3(c).

Define the residual term for $i \in I_j$ as

$$v_{in} = (\Sigma_x^{-1} - \widehat{\Sigma}_{jn}^{-1})(x_i - \mu_x)(\varepsilon_i/\sigma - \varkappa_1) - \widehat{\Sigma}_{jn}^{-1}(\mu_x - \bar{x}_j)(\varepsilon_i/\sigma - \varkappa_1). \tag{A.21}$$

Finally, note that

$N_j = n_j^{-1/2} I_{\dim \mathrm{x}}$. Thus, the Theorem 4.2 expansion has the form (A.20).

2. Conditional autocovariance matrices for s_i. We have $\mathsf{E}_{i-1}s_i = 0$ since s_{1i}, s_{2i} are independent of $\mathcal{F}_{i-1}$ with zero mean, while s_{3i} has a factor with that property. Thus, for $\ell \geq 1$, we have $\mathsf{E}_{i-\ell}s_i = 0$. The term, $\mathsf{E}_{i-1}s_{1i}^2 = \gamma(1-\gamma)$ is a Bernoulli variance, while $\mathsf{E}_{i-1}s_{1i}s_{2i} = \varsigma_2$ and $\mathsf{E}_{i-1}s_{2i}^2 = \varkappa_4 - 1$ by definitions of second and fourth moments. Similarly, $\mathsf{E}_{i-1}s_{1i}s_{1,i+1} = \varsigma_0 - \gamma^2$ and $\mathsf{E}_{i-1}s_{1i}s_{2,i+1} = \varsigma_2$. Since s_{2i}, s_{3i} are $\mathcal{F}_i$-adapted, and $\mathsf{E}_i s_{i+1} = 0$ we get for $\ell = 2, 3$ that $\mathsf{E}_{i-1}s_{\ell i}s'_{i+1} = \mathsf{E}_{i-1}s_{\ell i}\mathsf{E}_i s'_{i+1} = 0$. Since s_i is $\mathcal{F}_{i+1}$-adapted, then it is also $\mathcal{F}_{i+m-1}$-adapted for $m \geq 2$. Therefore, for $m \geq 2$, we get $\mathsf{E}_{i-1}s_i s'_{i+m} = \mathsf{E}_{i-1}s_i \mathsf{E}_{i+m-1}s'_{i+m} = 0$. Since x_i is $\mathcal{F}_{i-1}$-adapted while χ_i and ε_i are independent of $\mathcal{F}_{i-1}$, we get

$$\mathsf{E}_{i-1}s_{3i}s_{1i} = \Sigma_x^{-1}(x_i - \mu_x)\mathsf{E}1_{(|\chi_i| \geq c)}(\varepsilon_i/\sigma - \varkappa_1) = \Sigma_x^{-1}(x_i - \mu_x)\varsigma_1,$$

where $\varsigma_1 = \mathsf{E}1_{(|\chi_i| \geq c)}(\varepsilon_i/\sigma - \varkappa_1)$. Similarly

$$\mathsf{E}_{i-1}s_{3i}s_{2i} = \Sigma_x^{-1}(x_i - \mu_x)(\varkappa_3 - \varkappa_1),$$
$$\mathsf{E}_{i-1}s_{3i}s'_{3i} = \Sigma_x^{-1}(x_i - \mu_x)(x_i - \mu_x)'\Sigma_x^{-1}(1 - \varkappa_1^2).$$

Decompose $x_{i+1} - \mu_x = x_i - \mu_x - x_i + x_{i+1} = x_i - \mu_x - \nabla x_i$. Write

$$\begin{aligned}\mathsf{E}_{i-1} s_{3,i+1} s_{1i} &= \Sigma_x^{-1}(x_i - \mu_x)\mathsf{E}1_{(|\chi_i| \geq c)}(\varepsilon_i/\sigma - \varkappa_1) - \Sigma_x^{-1}\mathsf{E}_{i-1}\nabla x_i(1_{(|\chi_i| \geq c)} - \gamma)(\varepsilon_i/\sigma - \varkappa_1) \\ &= \Sigma_x^{-1}(x_i - \mu_x)\varsigma_1 - \Sigma_x^{-1}\varsigma_{1xi},\end{aligned}$$

with $\varsigma_{1xi} = \mathsf{E}_{i-1}\{\nabla x_i(1_{(|\chi_i| \geq c)} - \gamma)(\varepsilon_i/\sigma - \varkappa_1)\}$. The conditional auto product moments are $\mathsf{E}_{i-1} s_i s_{i+j-1}' = 0$ for $j \geq 2$ and

$$\mathsf{E}_{i-1} s_i s_i' = \begin{pmatrix} \gamma(1-\gamma) & \varsigma_2 & \varsigma_1(x_i - \mu_x)'\Sigma_x^{-1} \\ * & \varkappa_4 - 1 & (\varkappa_3 - \varkappa_1)(x_i - \mu_x)'\Sigma_x^{-1} \\ * & * & (1-\varkappa_1^2)\Sigma_x^{-1}(x_i - \mu_x)(x_i - \mu_x)'\Sigma_x^{-1} \end{pmatrix}, \tag{A.22}$$

$$\mathsf{E}_{i-1} s_i s_{i+1}' = \begin{pmatrix} \varsigma_0 - \gamma^2 & \varsigma_2 & \{\varsigma_{1xi}' - \varsigma_1(x_i - \mu_x)'\}\Sigma_x^{-1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{A.23}$$

3. Unconditional autocovariance matrices for s_i. Note that $\mathsf{E}x_i = \mu_x$ and $\mathsf{Var}(x_i) = \Sigma_x$. Apply iterated expectations noting that $\mathsf{E}_{i-1}s_i = 0$ to get $\mathsf{Var}(s_i) = \mathsf{E}\mathsf{E}_{i-1}s_i s_i'$ so that

$$\Omega_0 = \mathsf{Var}(s_i) = \begin{pmatrix} \gamma(1-\gamma) & \varsigma_2 & 0 \\ \varsigma_2 & \varkappa_4 - 1 & 0 \\ 0 & 0 & \Sigma_x^{-1}(1-\varkappa_1^2) \end{pmatrix}, \tag{A.24}$$

$$\Omega_1 = \mathsf{Cov}(s_i, s_{i+1}) = \begin{pmatrix} \varsigma_0 - \gamma^2 & \varsigma_2 & \varsigma_{1x}'\Sigma_x^{-1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{A.25}$$

as in (33), whereas $\mathsf{Cov}(s_i, s_{i+j}) = 0$ for $j \geq 2$. We note that these autocovariances are finite when $\mathsf{E}x_i^2 < \infty$ and $\mathsf{E}\varepsilon_i^4 < \infty$ as assumed in Assumption 4.3(i, ii).

4. Two-level long-run variance. Let $e_1 = (1, 0, 0)'$ and define, for $j = 1, 2$,

$$\omega_{jn}^2 = \mathsf{Var}(d_{jn}' s_i) + 2\mathsf{Cov}(d_{jn}' s_i, d_{jn}' s_{i+1}) = d_{jn}'\Omega_0 d_{jn} + 2e_1'\Omega_1 d_{jn}. \tag{A.26}$$

For large n then $n_k/n_j \to \lambda_k/\lambda_j > 0$, so that $d_{jn} \to d_j$ and $\omega_{jn}^2 \to \omega_j^2$, where

$$d_j = \begin{pmatrix} 1 \\ -c\mathsf{h}(c)(\lambda_k/\lambda_j) \\ -\mathsf{h}(c)\xi_c(\lambda_k/\lambda_j)/\sqrt{2} \end{pmatrix}, \qquad \omega_j^2 = d_j'\Omega_0 d_j + 2e_1'\Omega_1 d_j,$$

as in (34). By Assumption 4.3(iii) we have that $\omega_1^2, \omega_2^2 > 0$. Following McLeish, consider $\mathsf{E}X_{ni}^2 + \sum_{i \neq j}\mathsf{E}X_{ni}X_{nj}$. For i which are not near n_1 this equals ω_{jn}^2/n for $i \in I_j$ by the above derivations. As $\omega_{jn} \to \omega_j$, it is convenient to define $\sigma_{ni}^2 = \omega_j^2/n$ for $i \in I_j$.

5. Time distortion function. Cumulate to get $\sum_{i=1}^n \sigma_{ni}^2 = (n_1/n)\omega_1^2 + (n_2/n)\omega_2^2$. Asymptotically, this is equivalent to $T = \lambda_1\omega_1^2 + \lambda_2\omega_2^2$. Define the time distortion

$$k_n(t) = \begin{cases} \text{int}\,(tn/\omega_1^2) & \text{for } t \leq \lambda_1\omega_1^2, \\ \lambda_1 n + \text{int}\,(t - \lambda_1\omega_1^2)n/\omega_2^2 & \text{for } \lambda_1\omega_1^2 < t \leq T. \end{cases}$$

Note that $k_n(t)$ increases in steps of 1 from 0 to $k_n(T) = n$. It maps the proportion of the cumulated variance to the original observations. Following McLeish, let $W_n(t) =$

$\sum_{i=1}^{k_n(t)} X_{ni}$, so that $\sum_{i=1}^{k_n(T)} X_{ni} = \sum_{i=1}^{n} X_{ni}$. The long-run variance cumulates linearly for the time distorted process W_n. We get $\sum_{i=k_n(s)}^{k_n(t)} \sigma_{ni}^2/(t-s) = 1 + \mathrm{o}(1)$ for $s < t < T$, where the remainder arises from a rounding error at the break point.

6. Checking conditions of Lemma A.11. We choose mixingale coefficients $\psi_k^2 = C$ for $k = 0, 1, 2$ and some constant $C > 0$ described below, while $\psi_k = 1/k^2$ for $k > 2$.

(a) By the definition of σ_{ni}^2 we have $\sum_{i=k_n(s)}^{k_n(t)} \sigma_{ni}^2/(t-s) \to 1$ for $s < t < 1$. Thus, $\sup_{s<t<T} \limsup_{n\to\infty} \sum_{i=k_n(s)}^{k_n(t)} \sigma_{ni}^2/(t-s) < \infty$.

(b) It suffices to show that $\mathsf{E}|X_{ni}|^{2+}/\sigma_{ni}^{2+}$ is bounded uniformly in n, i (Billingsley, 1968, p. 32). We have $\mathsf{E}|X_{ni}/\sigma_{ni}|^{2+} \leq |d_{jn}|^{2+}\mathsf{E}|s_i|^{2+}/\omega_j^{2+}$ for $i \in I_j$. Here $|d_{jn}|$ converges in n for $j = 1, 2$ while $\omega_j > 0$ by Assumption 4.3(iii) for $j = 1, 2$. Finally, s_i is stationary with 2+ moments since ε_i^2 and x_i have 2+ moments by Assumption 4.3(ii).

(c) Since $n\sigma_{ni}^2 = \omega_{j_i}^2$ takes two values only, then $\max_{i\leq n} \sigma_{ni}^2 \to 0$ for $n \to \infty$.

(d) Let $S_{stu} = \mathsf{E}_{k_n(s)}(\sum_{i=k_n(t)}^{k_n(u)} X_{ni})^2$. We check that $\mathsf{E}|S_{stu} - (u-t)| \to 0$ for each $s < t < u$. Since $\mathsf{E}_{i-1} s_i s_{i+j-1}' = 0$ for $j \geq 2$ we have

$$S_{stu} = \mathsf{E}_{k_n(s)}\Big\{ \sum_{i=k_n(t)}^{k_n(u)} X_{ni}^2 + 2 \sum_{i=k_n(t)}^{k_n(u)-1} X_{ni} X_{n,i+1} \Big\}.$$

Recalling the expressions for $\mathsf{E}_{i-1} X_{ni} X_{n,i+j}$ and $\mathsf{E} X_{ni} X_{n,i+j}$ in items 4, 5, we get that

$$S_{stu} = \mathsf{E}\Big\{ \sum_{i=k_n(t)}^{k_n(u)} X_{ni}^2 + 2 \sum_{i=k_n(t)}^{k_n(u)-1} X_{ni} X_{n,i+1} \Big\} + \text{remainder}.$$

The detailed comparison of (A.22), (A.23) and (A.24), (A.25) shows that the remainder is a linear function of components like $\mathsf{E}_{k-m} n^{-1} \sum_{i=k+1}^{k+n} (y_i - \mathsf{E} y_i)$, where y_i represents either of the three z_i sequences in Assumption 4.3(iv), which are x_i, $(x_i - \mu_x)(x_i - \mu_x)'$ and $\nabla x_i (1_{(|\chi_i| \geq c)} - \gamma)(\varepsilon_i/\sigma - \varkappa_1)$, for any k, m, n. These components vanish in mean by that assumption.

Now, as remarked in item 4, we have $\mathsf{E}(X_{ni}^2 + 2X_{ni} X_{n,i+1}) = \omega_{jn}/n$ for $i \in I_j$, where $\omega_{jn} \to \omega_j = n\sigma_{ni}^2$ This convergence is uniform over n, i using condition (b) and (Billingsley, 1968, Theorem 5.4). Thus, the desired statement $\mathsf{E}|S_{stu} - (u-t)| \to 0$ follows as for condition (a).

(e) Let $\phi_{n,-k} = \mathsf{E}(\mathsf{E}_{i-k} X_{ni})^2$. We will define a positive, vanishing sequence ψ_k for $k \geq 0$, so that $\phi_{n,-k} \leq \psi_k^2 \sigma_{ni}^2$ for all $k \geq 0$.

For $k > 0$, we find $\mathsf{E}_{i-k} s_i = \mathsf{E}_{i-k} \mathsf{E}_{i-1} s_i = 0$ so that $\phi_{n,-k} = 0$. Any $\psi_k > 0$ suffices.

For $k = 0$, Jensen's inequality gives $\phi_{n,-0} = \mathsf{E}(\mathsf{E}_i X_{ni})^2 \leq \mathsf{E}\mathsf{E}_i X_{ni}^2 = \mathsf{E} X_{ni}^2$. Following the analysis for condition (b), we have $\mathsf{E} X_{ni}^2 \leq C\sigma_{ni}^2$ for some $C > 0$ and all n, i. We can choose $\psi_0^2 > C$.

(f) Let $\phi_{n,+k} = \mathsf{E}\{X_{ni} - \mathsf{E}_{i+k}(X_{ni})\}^2$. The sequence ψ_k must also satisfy that $\phi_{n,+k} \leq \psi_{k+1}^2 \sigma_{ni}^2$ for all $k \geq 0$.

For $k > 1$, then s_i is $\mathcal{F}_{i+k}$-adapted, so that $\phi_{n,+k} = 0$. Any choice of $\psi_k > 0$ suffices.

For $k = 0, 1$, bound $\phi_{n,+k} = \mathsf{E}\mathsf{E}_{i+k}\{X_{ni} - \mathsf{E}_{i+k}(X_{ni})\}^2 \leq \mathsf{E}\mathsf{E}_{i+k} X_{ni}^2 = \mathsf{E} X_{ni}^2$. As in condition ($b$), we have $\mathsf{E} X_{ni}^2 \leq C\sigma_{ni}^2$ for some $C > 0$ and all n, i. We must choose $\psi_1^2, \psi_2^2 \geq C$.

(g) This holds since $\sum_{k=0}^{\infty}\psi_k = 3C + \sum_{k=3}^{\infty}(1/k^2) < \infty$.

7. Applying Lemma A.11. We get that $\sum_{i=1}^{k_n(t)} X_{ni}$ converges in distribution to a standard Brownian motion. In particular $\sum_{i=1}^{k_n(T)} X_{ni}$ is asympotically $\mathsf{N}(0,B)$, where

$$\begin{aligned} B = \lambda_1\omega_1^2 + \lambda_2\omega_2^2 &= \lambda_1(d_1'\Omega_0 d_1 + 2e_1'\Omega_1 d_1) + \lambda_2(d_2'\Omega_0 d_2 + 2e_1'\Omega_1 d_2) \\ &= \lambda_1 d_1'\Omega_0 d_1 + \lambda_2 d_2'\Omega_0 d_2 + 2e_1'\Omega_1(\lambda_1 d_1 + \lambda_2 d_2). \end{aligned}$$

Inserting the expressions for Ω_0, Ω_1, d_j in (33), (34), see also (A.24), (A.25), (A.26), gives

$$\begin{aligned} B &= (\lambda_1+\lambda_2)\gamma(1-\gamma) - 2c\mathsf{h}(c)(\lambda_1+\lambda_2)\varsigma_2 \\ &\quad + \{c\mathsf{h}(c)\}^2(\lambda_1^2/\lambda_2 + \lambda_2^2/\lambda_1)(\varkappa_4 - 1) + \{\mathsf{h}(c)\}^2(1-\varkappa_1^2)(\lambda_1^2/\lambda_2 + \lambda_2^2/\lambda_1)\xi_c'\Sigma_x^{-1}\xi_c/2 \\ &\quad + 2(\lambda_1+\lambda_2)(\varsigma_0 - \gamma^2) - 2c\mathsf{h}(c)(\lambda_1+\lambda_2)\varsigma_2 - \sqrt{2}\mathsf{h}(c)(\lambda_1+\lambda_2)\varsigma_{1x}'\Sigma_x^{-1}\xi_c. \end{aligned}$$

Noting that $\lambda_1 + \lambda_2 = 1$ this reduces to the desired expression in (36), which is

$$\begin{aligned} B &= \gamma(1-\gamma) + 2(\varsigma_0 - \gamma^2) - 4c\mathsf{h}(c)\varsigma_2 - \sqrt{2}\mathsf{h}(c)\varsigma_{1x}'\Sigma_x^{-1}\xi_c \\ &\quad + (\lambda_1^2/\lambda_2 + \lambda_2^2/\lambda_1)\{\mathsf{h}(c)\}^2\{c^2(\varkappa_4 - 1) + (1-\varkappa_1^2)\xi_c'\Sigma_x^{-1}\xi_c/2\}. \end{aligned}$$

8. Remainder term. We argue that $n^{-1/2}\sum_{j=1}^{2}\sum_{i\in I_j} d_{3jn}v_{in}$ vanishes, where

$$v_{in} = (\Sigma_x^{-1} - \widehat{\Sigma}_{jn}^{-1})(x_i - \mu_x)(\varepsilon_i/\sigma - \varkappa_1) - \widehat{\Sigma}_{jn}^{-1}(\mu_x - \bar{x}_j)(\varepsilon_i/\sigma - \varkappa_1), \text{ for } i \in I_j$$

see (A.21). We have $\widehat{\Sigma}_{jn} \to \Sigma_x$ and $\bar{x}_j \to \mu_x$ in probability by Assumptions 3.2$(iiia)$, 4.1(a), 4.3(v). Further, $n^{-1/2}\sum_{i=1}^{n}(x_i - \mu_x)(\varepsilon_i/\sigma - \varkappa_1) = n^{-1/2}\sum_{i=1}^{n}\Sigma_x s_{3i}$ and is asymptotically normal using the above mixingale considerations. Finally, $n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i/\sigma - \varkappa_1)$ converges by a standard Central Limit Theorem. □

Proof of Theorem 4.5. The third term in the expansion of Theorem 4.2 vanishes since ξ_{jn} vanishes. We can then proceed exactly as in the proof of Theorem 4.4, but dropping any consideration to the third term in the expansion and in the assumptions. □

A.9 Explicit formulas for stationary case

Example 4.1. *Derivation of* ς_0 *in (35).* Write ς_0 as $\mathsf{E}\{1 - 1_{(|\chi_i|<c)}\}\{1 - 1_{(|\chi_{i+1}|<c)}\} = 1 - 2(1-\gamma) + \mathcal{I}_c$, where $\mathcal{I}_c = \mathsf{E}1_{(|\chi_i|<c)}1_{(|\chi_{i+1}|<c)}$. Conditional on ε_{i+1} the two indicators are i.i.d. with expectation $\Phi(\varepsilon_{i+1} + c\sqrt{2}) - \Phi(\varepsilon_{i+1} - c\sqrt{2})$. Thus,

$$\mathcal{I}_c = \int_{\mathbb{R}} \varphi(x)\{\Phi(x + c\sqrt{2}) - \Phi(x - c\sqrt{2})\}^2 dx.$$

Following Owen (1980, 2.2; 2.8) let $T(h,a) = \int_h^\infty \varphi(x)\int_0^{ax}\varphi(y)dydx$. By Owen (1980, 20,010.3) then $2\int_{\mathbb{R}}\varphi(x)\Phi(x+u\sqrt{2})\Phi(x+v\sqrt{2})dx = \Phi(u)+\Phi(v) - 2T\{u,(2v/u-1)/\sqrt{3}\} - 2T\{v,(2u/v-1)/\sqrt{3}\} - 1_{(uv<0)}$ for $u, v \neq 0$, so that

$$\begin{aligned} \mathcal{I}_c = \{\Phi(c) - 2T(c, 1/\sqrt{3})\} + \{\Phi(-c) - 2T(-c, 1/\sqrt{3})\} \\ - 2\{\frac{1}{2}\Phi(c) + \frac{1}{2}\Phi(-c) - T(c, -\sqrt{3}) - T(-c, -\sqrt{3}) - 1/2\}. \end{aligned}$$

By Owen (1980, 2.5; 2.6) then $T(-h,a) = T(h,a)$ and $T(h,-a) = -T(h,a)$. Thus,

$$\mathcal{I}_c = 1 - 4\{T(c, 1/\sqrt{3}) + T(c, \sqrt{3})\},$$

so that $\varsigma_0 = 1-2(1-\gamma)+1-4\{T(c,1/\sqrt{3})+T(c,\sqrt{3})\} = 2\gamma-4\{T(c,1/\sqrt{3})+T(c,\sqrt{3})\}$.

Derivation of ς_2 *in (35).* Recall $\varsigma_2 = \mathsf{E}\{1_{(|\chi_i|\geq c)}(\varepsilon_i^2/\sigma^2-1)\}$. Since $\mathsf{E}\{(\varepsilon_i^2/\sigma^2-1)\} = 0$, then $\varsigma_2 = -\mathcal{I}_c$ where

$$\mathcal{I}_c = \mathsf{E}\{1_{(|\chi_i|<c)}(\varepsilon_i^2/\sigma^2 - 1)\} = \int_{\mathbb{R}} \varphi(s) \int_{s-c\sqrt{2}}^{s+c\sqrt{2}} (t^2-1)\varphi(t)dtds.$$

As $\varphi(t)$ has derivates $\dot{\varphi}(t) = -t\varphi(t)$ and $\ddot{\varphi}(t) = (t^2-1)\varphi(t)$ then

$$\mathcal{I}_c = -\int_{\mathbb{R}} \varphi(s)\{(s+c\sqrt{2})\varphi(s+c\sqrt{2}) - (s-c\sqrt{2})\varphi(s-c\sqrt{2})\}ds$$

Since $\varphi(x)\varphi(x+c\sqrt{2}) = \varphi(c)\varphi(x\sqrt{2}+c)$ and $\varphi(x)\varphi(x-c\sqrt{2}) = \varphi(c)\varphi(x\sqrt{2}-c)$, we get

$$\mathcal{I}_c = -\varphi(c)\int_{\mathbb{R}}\{(s+c\sqrt{2})\varphi(s\sqrt{2}+c) - (s-c\sqrt{2})\varphi(s\sqrt{2}-c)\}ds.$$

Substituting $t = s\sqrt{2}+c$ and $u = s\sqrt{2}-c$, we get $\varsigma_2 = -c\varphi(c)$ since

$$\mathcal{I}_c = -\frac{1}{2}\varphi(c)\Big[\int_{\mathbb{R}}(t+c)\varphi(t)dt - \int_{\mathbb{R}}(u-c)\varphi(u)du\Big] = -c\varphi(c).$$

Example 4.2. Let $y_i = \mu + \alpha y_{i-1} + \varepsilon_i$ where ε_i/σ is standard normal, $|\alpha| < 1$ and the stationary distribution is normal with mean $\mu_y = \mu/(1-\alpha)$ and variance $\sigma_y^2 = \sigma^2/(1-\alpha^2)$. Note $\nabla x_i = y_{i-1} - y_i = (1-\alpha)y_{i-1} - \varepsilon_i$.

Proof that $\sigma_{\nabla\chi} = -\sigma/\sqrt{2}$. Recall that $\sigma_{\nabla\chi} = \mathsf{Cov}(\nabla x_1, \chi_1 \mid \mathcal{F}_0)$. Insert the expressions for x_1, χ_1 to get $\sigma_{\nabla\chi} = \mathsf{Cov}\{(1-\alpha)y_0 - \varepsilon_1, (\varepsilon_1-\varepsilon_2)/(2^{1/2}\sigma)\} = -\sigma/\sqrt{2}$.

We show that $\varsigma_1 = \mathsf{E}\{(1_{(|\chi_i|\geq c)} - \gamma)(\varepsilon_i/\sigma - \varkappa_1)\} = 0$ for a symmetric density. We get $\mathsf{E}1_{(\varepsilon_i-\varepsilon_{i+1}\leq -q)}\varepsilon_i = \mathsf{E}1_{(-\varepsilon_i+\varepsilon_{i+1}\leq -q)}(-\varepsilon_i)$ by the symmetry and independence of ε_i, ε_{i+1}. Change sign in the right indicator and combine the two expectations to get $\varsigma_1 = 0$.

Proof that $\varsigma_{1x} = -\sigma\varsigma_2$. Let $\varsigma_{1x} = \mathsf{E}\{\nabla x_i(1_{(|\chi_i|\geq c)} - \gamma)(\varepsilon_i/\sigma - \varkappa_1)\}$. We show $\varsigma_{1x} = -\sigma\varsigma_2$. Write $\nabla x_i = (1-\alpha)y_{i-1} - \varepsilon_i$ as $\varsigma_{1x} = (1-\alpha)\varsigma_{1x1} - \varsigma_{1x2}$. Since y_{i-1} is $\mathcal{F}_{i-1}$-measurable while ε_i, χ_i are independent of $\mathcal{F}_{i-1}$, we get $\varsigma_{1x1} = (\mathsf{E}y_{i-1})\varsigma_1 = 0$ by the above result. Further, $\varsigma_{1x2} = \mathsf{E}\{\varepsilon_i(1_{(|\chi_i|\geq c)} - \gamma)(\varepsilon_i/\sigma - \varkappa_1)\}$ satisfies

$$\begin{aligned}\varsigma_{1x2} = \sigma\mathsf{E}\{1_{(|\chi_i|\geq c)}(\varepsilon_i^2/\sigma^2 - 1)\} - \sigma\gamma\mathsf{E}\{(\varepsilon_i^2/\sigma^2-1)\} \\ + \sigma\mathsf{E}\{(1_{(|\chi_i|\geq c)} - \gamma)(1-\varkappa_1^2) - \sigma\varkappa_1\mathsf{E}\{(1_{(|\chi_i|\geq c)} - \gamma)(\varepsilon_i/\sigma - \varkappa_1)\}\end{aligned}$$

The first term equals $\sigma\varsigma_2$, the next two terms are zero by (27), (21), and the last term is 0 as $\varsigma_1 = 0$. Thus, $\varsigma_{1x} = -\sigma\varsigma_2$.

We check condition (iv). Write $y_i - \mu_y = \sum_{j=0}^{i-k+m-1}\alpha^j\varepsilon_{i-j} + \alpha^{i-k+m}(y_{k-m} - \mu_y)$.

For the case $z_i = y_i$ note that $z_{ikm} \equiv \mathsf{E}_{k-m}(y_i - \mathsf{E}y_i) = \alpha^{i-k+m}(y_{k-m} - \mu_y)$, so that $s_{km} \equiv n^{-1}\sum_{i=k+1}^{k+n} z_{ikm} = n^{-1}\alpha^{m+1}(y_{k-m} - \mu_y)(1-\alpha^n)/(1-\alpha)$, which converges to zero in mean as $\min(k,m,n) \to \infty$.

For the case $z_i = (y_i - \mu_y)^2$ note that $z_{ikm} \equiv \mathsf{E}_{k-m}\{(y_i - \mu_y)^2 - \mathsf{E}(y_i - \mu_y)^2\} = \alpha^{2(i-k+m)}\{(y_{k-m} - \mu_y)^2 - \sigma_y^2\}$, so that $s_{km} \equiv n^{-1}\sum_{i=k+1}^{k+n} z_{ikm} = n^{-1}\alpha^{2(m+1)}\{(y_{k-m} - \mu_y)^2 - \sigma_y^2\}(1-\alpha^{2n})/(1-\alpha^2)$, which converges to zero in mean as $\min(k, m, n) \to \infty$.

For the case $z_i = y_i^2$ note that $y_i^2 - \mathsf{E}y_i^2 = 2\mu_y(y_i - \mu_y) + \{(y_i - \mu_y)^2 - \sigma_y^2\}$ and use the previous two cases.

For the case $z_i = \nabla y_{i-1} 1_{(|\chi_i| \geq c)}$ write $\nabla y_{i-1} = y_{i-1} - y_i = (1-\alpha)y_{i-1} - \varepsilon_i$ so that $z_i = z_{1i} - z_{1i}$ where $z_{1i} = (1-\alpha)y_{i-1}1_{(|\chi_i|\geq c)}$ and $z_{2i} = \varepsilon_i 1_{(|\chi_i|\geq c)}$. Since z_{1i} is a product of independent terms, we get $z_{1ikm} \equiv \mathsf{E}_{k-m}(z_{1i} - \mathsf{E}z_{1i}) = (1-\alpha)\gamma\mathsf{E}_{k-m}(y_i - \mathsf{E}y_i)$ and the above results can be used. Since z_{2i} is independent of $\rangle - \infty$, we get $z_{2ikm} \equiv \mathsf{E}_{k-m}(z_{1i} - \mathsf{E}z_{2i}) = 0$.

A.10 Proof of Poisson results

We write $a_n \sim b_n$ if $a_n/b_n \to 1$. We state a special case of Chen (1975, Theorem 4.3).

Lemma A.12 (Chen, 1975)**.** *Suppose ε_i are i.i.d. so that $\chi_i = \nabla\varepsilon_i/(\sqrt{2}\sigma)$ satisfies $\mathsf{P}(|\chi_i| > c_n) = \lambda/n$. Suppose Assumption 5.1(i, d). Then $\sum_{i=1}^n 1_{(|\chi_i|>c_n)} \overset{\mathsf{D}}{\to} \mathsf{Poisson}(\lambda)$.*

We check the condition for normal variables

Lemma A.13. *If ε_i/σ are i.i.d. standard normal then $n\{\mathsf{E}1_{(|\chi_i|>c_n)}1_{(|\chi_{i+1}|>c_n)}\} \to 0$.*

Proof of Lemma A.13. Since ε_i/σ are i.i.d. standard normal, then

$$\begin{pmatrix} \chi_i \\ \chi_{i+1} \end{pmatrix} = \frac{1}{\sqrt{2}\sigma}\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} \varepsilon_i \\ \varepsilon_{i+1} \\ \varepsilon_{i+2} \end{pmatrix} \overset{\mathsf{D}}{=} \mathsf{N}\{0, \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}\}.$$

We can bound the covariance matrix, Ω_1, in terms of positive definite ordering, that is, for any 2-vector $v \neq 0$, then by

$$v'\Omega_1^{-1}v = v'\begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}^{-1} v > v'\begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}^{-1} v = v'\Omega_2^{-1}v.$$

Thus, we find

$$\begin{aligned} \mathcal{P}_1 = \mathsf{E}1_{(|\chi_i|>c_n)}1_{(|\chi_{i+1}|>c_n)} &= \iint_{|v_1|,|v_2|>c_n} \frac{1}{2\pi(\det\Omega_1)^{1/2}} \exp\Big(-\frac{1}{2}v'\Omega_1^{-1}v\Big)dv_1dv_2 \\ &< \Big(\frac{\det\Omega_2}{\det\Omega_1}\Big)^{1/2} \iint_{|v_1|,|v_2|>c_n} \frac{1}{2\pi(\det\Omega_2)^{1/2}} \exp\Big(-\frac{1}{2}v'\Omega_2^{-1}v\Big)dv_1dv_2 = \mathcal{P}_2. \quad \text{(A.27)} \end{aligned}$$

Substituting $z'z = v'\Omega_2^{-1}v$, that is $z_j = v_j/2$ and $dz_j = dv_j/2$, so that

$$\mathcal{P}_1 < \mathcal{P}_2 = \Big(\frac{\det\Omega_2}{\det\Omega_1}\Big)^{1/2}\Big\{\int_{|z|>c_n/2}\varphi(z)dz\Big\}^2 = \sqrt{\frac{64}{3}}\{1-\Phi(c_n/2)\}^2.$$

By Mill's ratio, it holds that $x\{1-\Phi(x)\} \sim \varphi(x)$ for $x \to \infty$ (Sampford, 1953), so that $\log x \sim \log\varphi(x) - \log\{1-\Phi(x)\}$. Apply for $x = c_n = \Phi^{-1}\{1-\lambda/(2n)\}$ while recalling the expression for the normal density and noting $1-\Phi(c_n)=\lambda/(2n)$, to get $2\log c_n \sim -\log(2\pi) - c_n^2 - 2\log\{\lambda/(2n)\}$. This implies $c_n^2 \sim -2\log\{\lambda/(2n)\} \sim 2\log n$ noting that $\log c_n = \mathrm{o}(c_n)$ and $2\pi = O(1)$. We then expand the normal density as

$$\varphi(c_n/2) \sim (2\pi)^{-1/2}\exp\{-(2\log n)/4\} = \mathrm{O}(n^{-1/2}).$$

Insert in Mill's ratio, that is $\{1-\Phi(c_n/2)\} \sim \varphi(c_n/2)/(c_n/2)$, to get $1-\Phi(c_n/2) = \mathrm{O}\{(n\log n)^{-1/2}\}$. Insert this in the bound (A.27) to get $n\mathcal{P}_1 = \mathrm{o}(1)$. □

We extract the following result from Johansen & Nielsen (2016b), see item 3 in the proof of their Theorem 8 as well as their Remark 2.

Lemma A.14 (Johansen & Nielsen, 2016b)**.** *Consider a continuous random variable χ with distribution function H and density h. Given $\lambda > 0$ choose c_n so that $\mathsf{P}(|\chi| > c_n) = \lambda/n$. Suppose, as $n \to \infty$,*
(a) $\mathsf{E}|\chi|^r < \infty$ *for some $r > 4$;*
(b) $\mathsf{h}(c_n)/[c_n\{1-\mathsf{H}(c_n)\}] = \mathrm{O}(1)$*;*
(c) $\mathsf{h}(c_n - n^{-1/4}A)/\mathsf{h}(c_n) = \mathrm{O}(1)$ *for all $A > 0$.*
Then, for all $A > 0$, as $n \to \infty$,

$$n\mathsf{E}1_{(c_n - n^{-1/4}A \le |\chi| \le c_n + n^{-1/4}A)} \to 0.$$

The conditions (a)-(c) are satisfied if χ is normal.

We need a modification of Lemma 1.11 in Johansen & Nielsen (2009).

Lemma A.15. *If $|a|+|b| < \zeta$ and $c > \zeta > 0$ then $|1_{(|\chi-b|>c+a)} - 1_{(|\chi|>c)}| \le 1_{(c-\zeta\le|\chi|\le c+\zeta)}$.*

Proof of Lemma A.15. Let $\mathcal{D} = |1_{(|\chi-b|>c+a)} - 1_{(|\chi|>c)}|$. Using that $1_{(\chi>c)} = 1 - 1_{(\chi\le c)}$, we get $\mathcal{D} = |1_{(|\chi-b|\le c+a)} - 1_{(|\chi|\le c)}|$. Write out as

$$\mathcal{D} = |1_{(-c-a+b\le\chi\le c+a+b)} - 1_{(-c\le\chi\le c)}| = |1_{(\chi\le c+a+b)} - 1_{(\chi\le c)} - 1_{(\chi<-c-a+b)} + 1_{(\chi<-c)}|,$$

which can be bounded around the focal points c and $-c$ by

$$\mathcal{D} \le 1_{(c-|a|-|b|\le\chi\le c+|a|+|b|)} + 1_{(-c-|a|-|b|\le\chi\le -c+|a|+|b|)} = 1_{(c-|a|-|b|\le|\chi|\le c+|a|+|b|)}.$$

Using the assumption $|a|+|b| \le \zeta$, the desired result follows. □

We combine Chen's Poisson limit in Lemma A.12 with Lemmas A.14, A.15.

Lemma A.16. *Suppose the conditions of Lemmas A.14, A.15 hold. Let a_i, b_i be sequences so that $\max_{i\le n}|a_i| + \max_{i\le n}|b_i| = \mathrm{O}_\mathsf{P}(n^{-1/4})$. Then*

$$\sum_{i=1}^n 1_{(|\chi_i - b_i| > c_n + a_i)} = \sum_{i=1}^n 1_{(|\chi_i| > c_n)} + \mathrm{o}_\mathsf{P}(1) \overset{\mathsf{D}}{\to} \mathsf{Poisson}(\lambda).$$

Proof of Lemma A.16. Let $\gamma_n = \sum_{i=1}^n 1_{(|\chi_i - b_i| > \sigma c_n + a_i)}$. Add and subtract $\sum_{i=1}^n 1_{(|\chi_i| > c_n)}$ and let n-vectors a, b represent a_i and b_i for $i \leq n$ to get

$$\gamma_n = \sum_{i=1}^n 1_{(|\chi_i| > \sigma c_n)} + \mathcal{R}_n(a,b) \quad \text{where} \quad \mathcal{R}_n(a,b) = \sum_{i=1}^n 1_{(|\chi_i - b_i| > \sigma c_n + a_i)} - 1_{(|\chi_i| > \sigma c_n)}.$$

The first term is asymptotically $\mathsf{Poisson}(\lambda)$ distributed by Lemma A.12. We show that the second term vanishes.

Since $\max_{i \leq n} |a_i| + \max_{i \leq n} |b_i| = \mathrm{O}_\mathsf{P}(n^{-1/4})$, we can, for any $\epsilon > 0$ and sufficiently large n, construct a sequence of sets $\mathcal{S}_n$ with $\mathsf{P}(\mathcal{S}_n^c) \leq \epsilon$, so that $|a_i| + |b_i| \leq An^{-1/4}$ for $i \leq n$ on $\mathcal{S}_n$. We find

$$\mathsf{P}\{|\mathcal{R}_n(a,b)| > \epsilon\} = \mathsf{P}[\{|\mathcal{R}_n(a,b)| > \epsilon\} \cap \mathcal{S}_n\} \mathsf{P}(\mathcal{S}_n) + \mathsf{P}[\{|\mathcal{R}_n(a,b)| > \epsilon\} \cap \mathcal{S}_n^c] \mathsf{P}(\mathcal{S}_n^c).$$

Bounding $\mathsf{P}(\mathcal{S}_n)$ and $\mathsf{P}[\{|\mathcal{R}_n(a,b)| > \epsilon\} \cap \mathcal{S}_n^c]$ by unity and the last probability by ϵ gives

$$\mathsf{P}\{|\mathcal{R}_n(a,b)| > \epsilon\} \leq \mathcal{P}_n + \epsilon \qquad \text{where} \qquad \mathcal{P}_n = \mathsf{P}[\{|\mathcal{R}_n(a,b)| > \epsilon\} \cap \mathcal{S}_n].$$

It suffices to show that $\mathcal{P}_n$ is small. We rewrite $\mathcal{P}_n$. On the set $\mathcal{S}_n$, we apply first the triangle inequality and then Lemma A.15 to get

$$|\mathcal{R}_n(a,b)| \leq \sum_{i=1}^n |1_{(|\chi_i - b_i| > \sigma c_n + a_i)} - 1_{(|\chi_i| > \sigma c_n)}| \leq \sum_{i=1}^n |1_{(c_n - An^{-1/4} \leq |\chi_i| \leq c_n + An^{-1/4})}| = \mathcal{R}_n^*.$$

Thus, $\mathcal{P}_n \leq \mathsf{P}\{(\mathcal{R}_n^* > \epsilon) \cap \mathcal{S}_n\} \leq \mathsf{P}(\mathcal{R}_n^* > \epsilon) = \mathcal{P}_n^*$. It suffices to show that $\mathcal{P}_n^*$ vanishes. Using the Markov inequality and then Lemma A.14 gives

$$\mathcal{P}_n^* \leq \frac{1}{\epsilon} \mathsf{E} \mathcal{R}_n^* = \frac{1}{\epsilon} n \mathsf{E} |1_{(c_n - An^{-1/4} \leq |\chi_1| \leq c_n + An^{-1/4})}| \to 0$$

for any (fixed) $\epsilon > 0$. Thus $\mathcal{R}_n^*$ vanishes and hence $\mathcal{R}_n$ vanishes. □

We assess the order of magnitude of the initial estimators and weights.

Lemma A.17. *Suppose Assumption 5.1(ia, ii, iii). Then* $N_1^{-1}(\hat{\beta}_1 - \beta)$, $n_1^{1/2}(\hat{\sigma}_1^2 - \sigma^2)$, $n^{1/4} \max_{i \in I_2^\circ} |\nabla x_{in}|$ *and* $n^{1/2} \max_{i \in I_2^\circ} |w_{j,i}^2 - 1|$ *are all* $\mathrm{O}_\mathsf{P}(1)$.

Proof of Lemma A.17. (a) We have $N_1^{-1}(\hat{\beta}_1 - \beta) = \widehat{\Sigma}_{1n}^{-1} \widehat{V}_{1n}$ when using the notation in (23), (24). Note that these expressions are invariant to the expectation of ε_i due to the demeaning. Using Assumption 5.1(ii, a, b), we find that $N_1^{-1}(\hat{\beta}_1 - \beta)$ is then $\mathrm{O}_\mathsf{P}(1)$.

(b) Let $\bar{\varepsilon}_1 = n_1^{-1} \sum_{i \in I_1} \varepsilon_i$ and write

$$n_1^{1/2}(\hat{\sigma}_1^2 - \sigma^2) = n_1^{-1/2} \sum_{i \in I_1} \{(\varepsilon_i - \bar{\varepsilon}_1)^2 - \sigma^2\} - n_1^{-1/2} \widehat{V}_{1n}' \widehat{\Sigma}_{1n}^{-1} \widehat{V}_{1n}.$$

By Assumption 5.1(i, a), then ε_i are i.i.d. with second moment, so that the first term converges in distribution by the Central Limit Theorem. The second term vanishes since $\widehat{\Sigma}_{1n}$, $\widehat{V}_{1n}$ are $\mathrm{O}_\mathsf{P}(1)$ by Assumption 5.1(ii, a, b) while the factor $n_1^{-1/2}$ vanishes.

(c) We have $\mathcal{P} = \mathsf{P}(\max_{i\in I_2^\circ} |\nabla x_{in}| > Cn^{-1/4}) = \mathsf{P}\cup_{i\in I_2^\circ} (|\nabla x_{in}| > Cn^{-1/4})$ for any $C > 0$. Boole's and Markov's inequalities give $\mathcal{P} \leq \sum_{i\in I_2^\circ} \mathsf{P}(|\nabla x_{in}| > Cn^{-1/4}) \leq C^{-4}n\mathsf{E}\sum_{i\in I_2^\circ} |\nabla x_{in}|^4$, which is small for large C due to Assumption 5.1(ii, c).

(d) Using the definition of the weights in (14) we find

$$w_{1,i}^2 - 1 = (N_2'\nabla x_i)' N_2^{-1} N_1 (2\widehat{\Sigma}_1)^{-1} (N_2^{-1}N_1)'(N_2'\nabla x_i).$$

Here, $N_2'\nabla x_i = \nabla x_{in}$ is $\mathrm{O}_\mathsf{P}(n^{-1/4})$ by part (c), while $\widehat{\Sigma}_1^{-1}$ and $N_2^{-1}N_1$ are $\mathrm{O}_\mathsf{P}(1)$ by Assumption 5.1(iia, iii). □

Proof of Theorem 5.2. (a) The stylized gauge has the expansion

$$\hat{\Gamma}_n^{stylized} = \sum_{i\in I_2^\circ} 1_{(|\nabla y_i - \hat{\beta}_1'\nabla x_i| > \sqrt{2}\hat{\sigma}_1 w_{1,i} c_n)} = \sum_{i\in I_2^\circ} 1_{(|\chi_i - b_i| > c_n + a_i)},$$

where $\chi_i = \nabla\varepsilon_i/(\sqrt{2}\sigma)$, while $a_i = (\hat{\sigma}_1 w_{1,i}/\sigma - 1)c_n$ and $b_i = (\hat{\beta}_1 - \beta)'\nabla x_i/(\sqrt{2}\sigma)$. We show that $\max_{i\in I_2^\circ} |a_i| + |b_i| = \mathrm{O}_\mathsf{P}(n^{-1/4})$ with a view to applying Lemma A.16

We bound a_i, b_i. Bound a_i by

$$\begin{aligned} |a_i| &= \big[\{1 + (\hat{\sigma}_1^2 - \sigma^2)/\sigma^2\}^{1/2}\{1 + (w_{1,i}^2 - 1)\}^{1/2} - 1\big]c_n. \\ &\leq \big\{(1 + |\hat{\sigma}_1^2 - \sigma^2|/\sigma^2)^{1/2}(1 + \max_{i\in I_2^\circ} |w_{1,i}^2 - 1|)^{1/2} - 1\big\}c_n. \end{aligned}$$

Here, $|\hat{\sigma}_1^2 - \sigma^2|$ and $\max_{i\in I_2^\circ} |w_{1i}^2 - 1|$ are $\mathrm{O}_\mathsf{P}(n^{-1/2})$ by Lemma A.17 using Assumption 5.1(ia, ii, iii). Thus, using the Taylor expansions $(1+x)^{1/2} = 1 + x/2 + \cdots = 1 + \mathrm{O}(x)$ and $(1+x)^2 - 1 = 1 + 2x + x^2 - 1 = \mathrm{O}(x)$ for small x, we find that $\max_{i\in I_2^\circ} |a_i| = \mathrm{O}_\mathsf{P}(n^{-1/2})c_n$. The cut-off c_n is $\mathrm{o}_\mathsf{P}(n^{1/4})$, see Johansen & Nielsen (2016b, Remark 1) using Assumption 5.1(ia). Thus, we find $\max_{i\in I_2^\circ} |a_i| = \mathrm{o}_\mathsf{P}(n^{-1/4})$.

Finally, write $\sqrt{2}\sigma b_i = \{N_1^{-1}(\hat{\beta}_1 - \beta)\}' N_1'\nabla x_i = (\widehat{\Sigma}_{1n}^{-1}\widehat{V}_{1n})'\nabla x_{in}$ using (23), (24). Here, $\widehat{\Sigma}_{1n}^{-1}$, $\widehat{V}_{1n}$ are $\mathrm{O}_\mathsf{P}(1)$ by Assumption 5.1(iia, iib), and $\max_{i\in I_2^\circ} |\nabla x_{in}| = \mathrm{O}_\mathsf{P}(n^{-1/4})$ by Lemma A.17 using Assumption 5.1(ia, ii, iii). Thus, $\max_{i\in I_2^\circ} |b_i| = \mathrm{O}_\mathsf{P}(n^{-1/4})$.

Having established that $\max_{i\in I_2^\circ} |a_i| + |b_i| = \mathrm{O}_\mathsf{P}(n^{-1/4})$, we can now apply Lemma A.16 using Assumption 5.1(i) to conclude that the stylized gauge satisfies $\hat{\Gamma}_n^{stylized} = \sum_{i\in I_2^\circ} 1_{(|\chi_i| > c_n)} + \mathrm{o}_\mathsf{P}(1)$, which is asymptotically $\mathsf{Poisson}$. Note, that $\mathsf{P}(|\chi_i| > c_n) = \lambda/n = (\lambda/n_2)(n_2/n)$, where $n_2/n \to \psi$. Hence, the Poisson parameter is $\lambda\psi$.

(b) For the split gauge, we can analyze the stylized gauges separately for each sub-sample as above and combine the expansions. □

A.11 Proof of power results

We prove the local power results for the Andrews test. The F-test statistic for a break at time t given in (A.28) has the expression

$$Z_t^2 = (n-2)\frac{S_{yt}^2/S_{tt}}{S_{yy} - S_{yt}^2/S_{tt}} = (n-2)\frac{S_{yt}^2/(S_{yy}S_{tt})}{1 - S_{yt}^2/(S_{yy}S_{tt})}, \qquad \text{(A.28)}$$

where the product moment statistics in (A.29) have the form

$$S_{tt} = \frac{1}{n}\sum_{i=1}^{n}\left\{1_{(i\le t)} - \frac{t}{n}\right\}^2, \quad S_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2, \quad S_{yt} = \frac{1}{n}\sum_{i\le t}(y_i - \overline{y}). \qquad (A.29)$$

We note that the test statistic is location-scale invariant.

Mode of convergence. We will be interested in sequences of continuous break functions that may have a single discontinuity in the limit. Such functions are members of the $D[0,1]$ space of discontinuous function that are right continuous on [0,1), with left limits on (0,1], and left continuous at 1. Skorokhod (1956) suggested five metrics of which we focus on: U, J_1 and M_1. Broadly speaking, the uniform metric U applies when the limit is continuous; the one-jump metric J_1 applies when the elements of the sequence have isolated jumps; the metric M_1 controls upcrossings and applies when the sequence members have isolated jumps or isolated smooth level shifts with discontinuous limits (Skorokhod, 1956, 2.2.11). Billingsley (1968) refers to J_1 as 'the' Skorokhod metric, whereas Whitt (2002) prefers to use the M_1 metric. Both metrics result in a separable and topologically complete metric space (Whitt, 2002, Theorem 12.8.1). Skorokhod (1956, p. 267) argues that U-convergence implies J_1-convergence, which in turn implies M_1-convergence. Equally, U-weak convergence implies J_1-weak convergence, which implies M_1-weak convergence, noting that for weak convergence under a metric m, the involved probability measures must be measurable under the Borel σ-field generated by the m-topology. For a limiting Brownian bridge formed from convergence of a random walk in a time series, we have U-weak convergence (Billingsley, 1968, Chapter 18). We can check M_1 convergence of elements $x_n(u) \to x(u)$ by showing that the number of upcrossings of x_n over a, b strips converges over all intervals $[u_1, u_2]$, where u_1, u_2 are continuity points of $x(u)$ and for almost every $a < b$ (Skorokhod, 1956, 2.2.11). Running suprema are U, J_1, M_1 continuous mappings into D (Whitt, 2002, Lemma 13.4.1, Theorem 13.4.1). Consequently, k-dimensional coordinate projections are U, J_1, M_1 continuous mappings into $\mathbb{R}^k$. Addition is U, J_1 and M_1-continuous for continuous limits. Addition is J_1 and M_1-continuous for limits with no common jumps. Further, addition is also M_1-continuous for limits with common jumps with the same sign (Whitt, 2002, Example 3.3.1, Theorem 12.7.3). Thus, addition of a U-convergent process and a J_1 or an M_1 convergent process is continuous.

A single central break. The data generating process is given by $y_i = \mu + \sigma\delta 1_{(i\le\tau)} + \varepsilon_i$, where ε_i is i.i.d. $\mathsf{N}(0,\sigma^2)$ while $0 < \tau < n$. In the following, note that τ is the break in the data generating process and t is the position of the break in the test statistic. Due to the location-scale invariance, it suffices to consider $\mu = 0$ and $\sigma = 1$. Take average to get $\overline{y} = \delta\tau/n + \overline{\varepsilon}$ and residuals $y_i - \overline{y} = \delta\{1_{(i\le\tau)} - \tau/n\} + \varepsilon_i - \overline{\varepsilon}$. We find

$$\sqrt{n}S_{yt} = \frac{1}{\sqrt{n}}\sum_{i\le t}(\varepsilon_i - \overline{\varepsilon}) + x_n(u) \qquad (A.30)$$

for the deterministic function

$$x_n(u) = \delta\sqrt{n}\Big\{\Big(1 - \frac{\tau}{n}\Big)\frac{t}{n}1_{(t\le\tau)} + \frac{\tau}{n}\Big(1 - \frac{t}{n}\Big)1_{(t>\tau)}\Big\}. \qquad (A.31)$$

We embed S_{yt} as a process in $D[0,1]$ through $t = \lfloor un \rfloor$ for $0 \le u \le 1$. The first component in (A.30) U-converges to a standard Brownian bridge. To see this note that $n^{-1/2}\sum_{i\le\lfloor un\rfloor}\varepsilon_i$ will J_1-converge to a Brownian motion (Billingsley, 1968, Theorem 16.1). Due to the continuity of the Brownian motion it will also U-converge (Skorokhod, 1956, Theorem 2.6.2). The convergence is also weak, that is Borel measurable Billingsley (1968, Section 18). The Brownian bridge convergence emerges with the mapping $x(t) \mapsto x(t) - tx(1)$, which is U-continuous due to U-continuity of multiplication, addition and the coordinate mapping. We can also expand

$$S_{yy} = \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2 + 2\delta\frac{1}{n}\sum_{i\le\tau}(\varepsilon_i - \overline{\varepsilon}) + \delta^2\frac{\tau}{n}\Big(1 - \frac{\tau}{n}\Big). \tag{A.32}$$

The first term converges to one in probability by the Law of Large Numbers. We also have U-convergence in $u = t/n$ for $0 \le u \le 1$ of

$$S_{tt} = \frac{t}{n}\Big(1 - \frac{t}{n}\Big) \to u(1-u).$$

The local power result (49) arises when $\delta\sqrt{n} = \phi$ and $\tau/n = \lambda$ for fixed ϕ, λ. We embed x_n from (A.31) in $D[0,1]$ through $u = t/n$. Connecting grid points linearly gives the continuous process

$$x_n(u) = \phi\big\{(1-\lambda)u1_{(u\le\lambda)} + \lambda(1-u)1_{(u>\lambda)}\big\} = \phi s_u^\lambda.$$

This is constant in n and therefore U-convergent. Further, $S_{yy} \to 1$ when δ vanishes since $n^{-1/2}\sum_{i\le\tau}(\varepsilon_i - \overline{\varepsilon})$ converges by the Central Limit Theorem. Using that addition, multiplication and supremum are U-continuous, we get by continuous mapping

$$\frac{\sqrt{n}S_{yt}}{\sqrt{S_{yy}S_{tt}}}1_{(\underline{\lambda}\le t/n\le\overline{\lambda})} \overset{\mathsf{D}}{\to} \frac{\mathbb{B}_u + \phi s_u^\lambda}{\sqrt{u(1-u)}}1_{(\underline{\lambda}\le u\le\overline{\lambda})}$$

as U-convergence on $D[0,1]$. Form the t-statistic and take supremum to get (49).

The local power result (51) arises as $\tau/n \to 1$. Thus, let $\delta\sqrt{n}(1-\tau/n) = \psi$ where ψ is fixed while $\tau/n \to 1$. Note that $\tau \le n-1$ implies $\delta/\sqrt{n} \le \psi$ with equality for $\tau = n-1$. We get

$$x_n(u) = \psi\Big\{u1_{(u\le\tau/n)} + \frac{\tau/n}{1-\tau/n}(1-u)1_{(u>\tau/n)}\Big\} \to x(u) = \psi u$$

as U-convergence on $D[0,\overline{\lambda}]$ noting $\tau/n > \overline{\lambda}$ for large n. Consider S_{yy} in (A.32). We have that $n^{-1/2}\sum_{i\le\tau}(\varepsilon_i - \overline{\varepsilon}) = -n^{-1/2}\sum_{i>\tau}(\varepsilon_i - \overline{\varepsilon})$, which vanishes as $\tau/n \to 1$. Further, with $\delta/\sqrt{n} \to \eta$, we must have $0 \le |\eta| \le |\psi|$ and $\eta\psi \ge 0$. Using that $\delta\sqrt{n}(1-\tau/n) = \psi$ and $\tau/n \to 1$, we get

$$S_{yy} = 1 + \mathsf{o}_\mathsf{P}(1) + \Big(\frac{\delta}{\sqrt{n}}\Big)\delta\sqrt{n}(1-\tau/n)\Big(\frac{\tau}{n}\Big) \to 1 + \eta\psi. \tag{A.33}$$

Combine as before to get, as U-convergence on $D[0,\overline{\lambda}]$ and on $D[0,1]$,

$$\frac{\sqrt{n}S_{yt}}{\sqrt{S_{yy}S_{tt}}}1_{(\underline{\lambda}\le t/n\le\overline{\lambda})} \overset{\mathsf{D}}{\to} \frac{\mathbb{B}_u + \psi u}{\sqrt{u(1-u)(1+\eta\psi)}}1_{(\underline{\lambda}\le u\le\overline{\lambda})}.$$

Form the t-statistic and take supremum to get (51).

Two central breaks. The data generating process is given by $y_i = \mu + \sigma\delta_1 1_{(i\le\tau_1)} + \sigma\delta_2 1_{(i\le\tau_2)} + \varepsilon_i$, where ε_i is i.i.d. $\mathsf{N}(0,\sigma^2)$ where $0<\tau_1<\tau_2<n$. Due to location-scale invariance, we can set $\mu=0$ and $\sigma=1$. Proceed as before to get

$$\sqrt{n}S_{yt} = \frac{1}{n}\sum_{i\le t}(\varepsilon_i - \overline{\varepsilon}) + x_n(u),$$

where the deterministic part is now

$$x_n\Big(\frac{t}{n}\Big) = \delta_1\sqrt{n}\Big\{\Big(1-\frac{\tau_1}{n}\Big)\frac{t}{n}1_{(t\le\tau_1)} + \frac{\tau_1}{n}\Big(1-\frac{t}{n}\Big)1_{(t>\tau_1)}\Big\}$$
$$+\,\delta_2\sqrt{n}\Big\{\Big(1-\frac{\tau_2}{n}\Big)\frac{t}{n}1_{(t\le\tau_2)} + \frac{\tau_2}{n}\Big(1-\frac{t}{n}\Big)1_{(t>\tau_2)}\Big\}.$$

Local power arises as $\tau_j/n \to \lambda_j$ for $0<\lambda_1<\lambda_2<1$ and $\delta_j = \xi_j/\sqrt{n}$ for fixed ξ_j.

The local power result (52) arises when the breaks are close and offsetting each other. Thus, let $\tau_1/n=\lambda$ and $\delta_2(\tau_2-\tau_1)/\sqrt{n}=\psi$ while $(\delta_1+\delta_2)\sqrt{n}=\xi$ for fixed λ_1, ψ, ξ, while $(\tau_2-\tau_1)/n\to 0$. As for (51), we note that $\tau_2>\tau_1$ implies $\delta_2/\sqrt{n}\le\psi$ with equality when $\tau_2-\tau_1=1$. Finally, we let $u=t/n$.

We analyze x_n. Add and subtract δ_2 to δ_1 and to get $x_n = x_{n,1}+x_{n,2}$ where

$$x_{n,1}(u) = (\delta_1+\delta_2)\sqrt{n}\{(1-\lambda)u1_{(u\le\lambda)} + \lambda(1-u)1_{(u>\lambda)}\} = s_u^\lambda = x_1(u), \tag{A.34}$$
$$x_{n,2}(u) = \delta_2\sqrt{n}\Big\{\Big(1-\frac{\tau_2}{n}\Big)u1_{(t\le\tau_2)} + \frac{\tau_2}{n}(1-u)1_{(t>\tau_2)}$$
$$-\Big(1-\frac{\tau_1}{n}\Big)u1_{(t\le\tau_1)} - \frac{\tau_1}{n}(1-u)1_{(t>\tau_1)}\Big\}.$$

Here $x_{n,1}$ is continuous and constant in n, so U-converges to x_1 say. We rewrite $x_{n,2}$ further as

$$x_{n,2}(u) = \delta_2\frac{\tau_2-\tau_1}{\sqrt{n}}\Big\{1_{(t>\tau_1)} - \frac{t}{n} - \frac{\tau_2-t}{\tau_2-\tau_1}1_{(\tau_1<t\le\tau_2)}\Big\}$$
$$= \psi\Big[1_{(u>\lambda)} - u - \frac{\tau_2-un}{\tau_2-\tau_1}1_{\{\lambda<u\le\lambda+(\tau_2-\tau_1)/n\}}\Big],$$

which is continuous. We argue that as $(\tau_2-\tau_1)/n$ shrinks, $x_{n,2}$ has a discontinous M_1-limit on $D[0,1]$ given by

$$x_2(u) = \psi\{1_{(u\ge\lambda)} - u\}. \tag{A.35}$$

We apply the Skorokhod (1956, 2.2.11) criterion for M_1-convergence. It suffices to consider convergence of $z_{n,2}(u) = x_{n,2}(u)+\psi u$ to $z_2 = x_2+\psi u$, noting that if we have M_1-convergence of $z_{n,2}$ then addition with the U-convergent function $-\psi u$ is continuous. The function $z_{n,2}$ is the x'-example of Skorokhod (1956, p. 266) of a function that is M_1-converging but not J_1-converging. To establish M_1-convergence, consider a,b-upcrossings over intervals $[u_1,u_2]$ for z_2-continuity points, so that $u_1,u_2\ne\lambda$. The functions $z_{n,2}(u)$ and $z_2(u)$ are continuous and identical for $u<\lambda$ and $u>\tau_2/n$. So we have convergence when $u_1<u_2<\lambda$ and $\lambda<u_1<u_2$. Thus, consider a,b-upcrossings

for $u_1 < \lambda < u_2$. For $0 < a < b < 1$ there is one upcrossing for $z_{n,2}$ and z_2 for large n. If $a < 0$ or $b > 1$ there are zero upcrossing for $z_{n,2}$ and z_2. In both cases, the number of upcrossings converges. Thus, $z_{n,2}$ and hence $x_{n,2}$ will M_1-converge. Next, we show that S_{yy} satisfies a result resembling (A.33). Write

$$y_i - \overline{y} = \varepsilon_i - \overline{\varepsilon} + \big(\delta_1 + \delta_2\big)\Big\{1_{(i \le \tau_1)} - \frac{\tau_1}{n}\Big\} + \delta_2\Big\{1_{(\tau_1 < i \le \tau_2)} - \frac{\tau_1 - \tau_1}{n}\Big\}.$$

From this we find

$$\begin{aligned} S_{yy} = &\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2 + 2\big(\delta_1 + \delta_2\big)\frac{1}{n}\sum_{i \le \tau_1}(\varepsilon_i - \overline{\varepsilon}) + 2\delta_2\frac{1}{n}\sum_{\tau_1 < i \le \tau_2}(\varepsilon_i - \overline{\varepsilon}) \\ &+ \big(\delta_1 + \delta_2\big)^2\Big\{\frac{\tau_1}{n}\Big(1 - \frac{\tau_1}{n}\Big)\Big\} + \delta_2^2\frac{\tau_2 - \tau_1}{n}\Big(1 - \frac{\tau_2 - \tau_1}{n}\Big) - 2\big(\delta_1 + \delta_2\big)\delta_2\frac{(\tau_2 - \tau_1)\tau_1}{n^2}. \end{aligned}$$

We note that $n^{-1}\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2$ and $n^{-1/2}\sum_{i \le \tau_1}(\varepsilon_i - \overline{\varepsilon})$ converge while $n^{-1/2}\sum_{\tau_1 < i \le \tau_2}(\varepsilon_i - \overline{\varepsilon})$ vanishes when $(\tau_2 - \tau_1)/n$ vanishes. Use also that $\tau_1/n = \lambda$, $\delta_1 + \delta_2 = \xi/\sqrt{n}$ and $\delta_2(\tau_2 - \tau_1) = \psi\sqrt{n}$. Finally, let $\delta_2/\sqrt{n} \to \eta$ where $0 \le |\eta| \le \psi$ and $\eta\psi \ge 0$. We get that $S_{yy} \to 1 + \eta\psi$. Put all together to get as M_1 convergence on $D[0,1]$ that

$$\frac{\sqrt{n}S_{yt}}{\sqrt{S_{yy}S_{tt}}}1_{(\underline{\lambda} \le t/n \le \overline{\lambda})} \overset{\mathsf{D}}{\to} \frac{\mathbb{B}_u + \xi s_u^\lambda + \psi\{1_{(u \ge \lambda)} - u\}}{\sqrt{u(1-u)(1+\eta\psi)}}1_{(\underline{\lambda} \le u \le \overline{\lambda})}.$$

Form the t-statistic and take supremum to get (52).

References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*, 821–856.

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277–297.

Bai, J. & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*, 47–78.

Becker, W., Paruolo, P., & Saltelli, A. (2021). Variable selection in regression models using global sensitivity analysis. *Journal of Time Series Econometrics*, *13*, 187–233.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

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

Berenguer-Rico, V. & Nielsen, B. (2023). Normality testing after outlier removal. Technical report.

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

Bernstein, D. H. & Martinez, A. B. (2021). Jointly modeling male and female labor participation and unemployment. *Econometrics*, *9*(4), 46.

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

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, *3*, 1941–1928.

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.

Castle, J. L. & Hendry, D. F. (2014). Model selection in under-specified equations facing breaks. *Journal of Econometrics*, *178*, 286–293.

Castle, J. L., Hendry, D. F., & Martinez, A. B. (2023). The historical role of energy in UK inflation and productivity with implications for price inflation. *Energy Economics*, *126*, article 106947.

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

Chadha, J. S. (2022). Productivity in the UK: Evidence review. Report of the UK productivity commission, National Institute of Economic and Social Research.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Annals of Probability*, *3*, 534–545.

Chuffart, T. & Hooper, E. (2019). An investigation of oil prices impact on sovereign credit default swaps in russia and venezuela. *Energy Economics*, *80*, 904–916.

Clements, M. P. & Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Doornik, J. A. (2009). Autometrics. In Castle & Shephard (2009), (pp. 88–121).

Doornik, J. A., Castle, J. L., & Hendry, D. F. (2022). Short-term forecasting of the coronavirus pandemic. *International Journal of Forecasting*, *38*(2), 453–466.

Doornik, J. A. & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, *70*, 927–939.

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

Dudoit, S. & van der Laan, M. J. (2010). *Multiple Testing Procedures with Applications to Genomics*. New York: Springer.

EViews (2020). Using indicator saturation to detect outliers and structural shifts. https://blog.eviews.com/2020/12/using-indicator-saturation-to-detect.html.

Freedman, D. A. (1975). On tail probabilities for martingales. *Annals of Probability*, *3*, 100–118.

Giraitis, L., Koul, H. L., & Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. London: Imperial College Press.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, *46*, 1293–1301.

Hendry, D. F. & Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*. London: MIT Press.

Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in fully saturated regression. *Computational Statistics*, *23*, 317–335. Erratum *ibid.* pp. 337–339.

Hendry, D. F. & Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, *115*, 32–61.

Hendry, D. F. & Mizon, G. E. (2011). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, *3*, 1941–1928.

Hendry, D. F. & Nielsen, B. (2007). *Econometric Modeling: A Likelihood Approach*. Princeton, NJ: Princeton University Press.

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

Hidalgo, J. & Seo, M. H. (2013). Testing for structural stability in the whole sample. *Journal of Econometrics*, *175*, 84–93.

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. (2019). A simple robust procedure in instrumental variables regression. Mimeo, University of Oxford.

Jiao, X. & Nielsen, B. (2017). Asymptotic analysis of iterated 1-step Huber-skip M-estimators with varying cut-offs. In J. Antoch, J. Jurečková, M. Maciak, & M. Pešta (Eds.), *Analytic Methods in Statistics*, volume 193 of *Springer Proceedings in Mathematics & Statistics* (pp. 23–55). New York: Springer.

Jiao, X. & Pretis, F. (2022). Testing the presence of outliers in regression models. *Oxford Bulletin of Economics and Statistics*, *84*, 1452–1484.

Johansen, S. & Nielsen, B. (2009). Saturation by indicators in regression models. In Castle & Shephard (2009), (pp. 1–36).

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. Corrigendum (2019) 25, 3201.

Johansen, S. & Nielsen, B. (2016b). Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics*, *43*, 321–81.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1993). *Continuous univariate distributions* (2nd ed.)., volume 2. New York: John Wiley & Sons.

Koch, N., Naumann, L., Pretis, F., Ritter, N., & Schwarz, M. (2022). Attributing agnostically detected large reductions in road co2 emissions to policy mixes. *Nature Energy*, *7*(9), 844–853.

Koul, H. L. & Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics*, *22*, 540–562.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*, 391–413.

Lancaster, T. (2002). Orthogonal parameters and panel data. *The Review of Economic Studies*, *69*, 647–666.

Marczak, M. & Proietti, T. (2016). Outlier detection in structural time series models: The indicator satuation approach. *International Journal of Forecasting*, *32*, 180–202.

Marriott, F. & Pope, J. (1954). Bias in the estimation of autocorrelations. *Biometrika*, *41*, 390–402.

McLeish, D. L. (1977). On the invariance principle for nonstationary mixingales. *Annals of Probability*, *5*, 616–621.

Miller, R. G. (1981). *Simultaneous Statistical Inference* (2 ed.). New York: Springer.

Nielsen, B. (2006). Order determination in general vector autoregressions. In H. C. Ho, C. K. Ing, & T. L. Lai (Eds.), *Time series and related topics: In memory of Ching-Zong Wei*, volume 52 of *Lecture Notes–Monograph Series* (pp. 93–112). Beachwood, OH: Institute of Mathematical Statistics.

Nielsen, B. & Whitby, A. (2015). A joint Chow test for structural instability. *econometrics*, *3*, 156–186.

O'Callaghan, B., Yau, N., & Hepburn, C. (2022). How stimulating is a green stimulus? the economic attributes of green fiscal spending. *Annual review of Environment and Resources*, *47*, 697–723.

Owen, D. B. (1980). A table of normal integrals. *Communications in Statistics. Simulation and Computation*, *9*, 389–419.

Pellini, E. (2021). Estimating income and price elasticities of residential electricity demand with autometrics. *Energy Economics*, *101*, 105411.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, *57*, 1361–1401. Erratum in volume 61, 248–249.

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

Pretis, F., Reade, J., & Sucarrat, G. (2018). Automated general-to-specific (gets) regression modeling and indicator saturation for outliers and structural breaks. *Journal of Statistical Software*, *86*, 1–44.

Pretis, F., Schwarz, M., Tang, K., Haustein, K., & Allen, M. R. (2018). Uncertain impacts on economic growth when stabilizing global temperatures at 1.5 C or 2 C warming. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2119), 20160460.

Raggad, B. (2018). Carbon dioxide emissions, economic growth, energy use, and urbanization in Saudi Arabia: evidence from the ARDL approach and impulse saturation break tests. *Environmental Science and Pollution Research*, *25*, 14882–14898.

Sampford, M. R. (1953). Some inequalities on Mill's ratio and related functions. *Annals of Mathematical Statistics*, *24*, 130–132.

Sheppard, K. (2018). *MFE MATLAB function reference financial econometrics.* www.kevinsheppard.com.

Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theory of Probability and its Applications*, *1*, 261–290.

Stone, C. (1963). Weak convergence of stochastic processes defined on semi-infinite time intervals. *Proceedings of the American Mathematical Society*, *14*, 694–696.

Sucarrat, G. (2020). User-specified general-to-specific and indicator saturation methods. *R Journal*, *12*, 251–265.

Whitt, W. (2002). *Stochastic-Process Limits: An introduction to stochastic-process limits and their application to queues.* New York: Springer.

Yao, Y. C. & Davis, R. A. (1986). The asymptotic behavior of the likelihood ratio test statistic for testing a shift in mean in a sequence of independent normal variables. *Sankhya A*, *48*, 339–353.