# The Asymptotic Theory of the Gauge of the Step-indicator Saturation Estimator and its Application to Study the Effect of Deindustrialisation on Real Rates




Matthias Qian
St Edmund Hall
University of Oxford

A thesis submitted for the degree of
*Doctor of Philosophy*

Trinity 2018

To my parents, Wei and Deyu.

# Acknowledgements

When I chose to embark on the journey of a DPhil four years ago, it was Professor Bent Nielsen who chose to help me make it a success. His patient guidance has steadily steered me toward the completion of this document. With contagious enthusiasm he taught me how to develop asymptotic theory for statistical algorithms. The mathematics that Bent uses is intricate and it requires a detail oriented mindset. I have a natural tendency to think fast and to gloss over complicated details. Bent has educated me to think more cautiously. The experience has been a location shift in the amount of slow thinking, as compared to fast thinking, that I engage in. Although I still catch myself in logical jumps and inconsistencies, I now more often than not follow Bent's example and take a pen and pencil to think slowly about a problem. Indeed, there are few better opportunities to practise slow thinking than to write chaining arguments for empirical processes. While Bent is a theorist, he has pointed me to many useful empirical techniques that I have applied in my empirical work. For example, in my third chapter, I use the catalyst method that he has developed with Dr. Vassili Bazinas, as an alternative method to study causal relationships. Whenever I tested for co-integration, Bent has offered indispensable help in setting up the right econometric specification. Beyond being the ultimate academic mentor, Bent has also been a kind friend. In my many DPhil crisis that I have lived through, he has always found a way to lift me out of my misery. I am forever indebted.

An exceptional initiative during my DPhil from which I benefited greatly was the probability and measure reading group that Prof. James Duffy has led. Life's tragedy is that we get old too soon and

wise too late. I longed to have learned the insights of the measure-theoretic foundation of probability, before embarking on this thesis project. And it was James' initiative that helped to fill the many gaps in my understanding of probability.

Thank you to the Oxford Man Institute under both the leadership of Professor Terry Lyons and Professor Stephen Roberts. It has been a terrific institutional bridge between academic research and the quantitative investment community. They have been very generous in providing high quality office space and a wonderful interdisciplinary community of scholars who are interested in finance.

A special thanks to Dr. Fiona Spencely from Lady Margaret Hall who has welcomed me to teach at this beautiful college. The senior common room at LHM has been an extraordinarily stimulating environment for enlightening discussions. All students have been great fun to teach, and it was a particularly memorable honour to teach Nobel laureate Yousafzai Malala. Many of my best memories at Oxford have been at LMH.

Thanks to the many generations of scholars before us, who have shaped the most wonderful place on earth, the University of Oxford. The St. Sepulchre's Cemetery that I overlook from Eagle House and the graves in the garden of St.Edmund Hall on the way to their library have forcefully reminded me that we expand on the contributions of many generations. No single generation can build a university as great as Oxford is.

My graduate studies have been supported by a number of organisations. Thank you, to the Studienstiftung des deutschen Volkes, a wonderful organisation full of critical thinkers that has been a great inspiration to me. Over the past year, I was fortunate to be awarded the David Walton Fellowship of the Department of Economics in Oxford. I have never met David personally, because he suddenly and unexpectedly passed away in June 2006. But through the generosity of his wife Nicola, I have not only received funding for the fourth year

of my DPhil, but also had the opportunity to work at Macro Research at Goldman Sachs; an experience that I have greatly enjoyed.

My friends in Oxford have made this place much better than any place that I have been before. Thank you Houye, Jiabao, Xiaowu, Yixuan, Ang, Hongri, Jakob, Christoph, Favour, James, Charlie, Jonas, Zoe, Xiyu, Tara, Jan, Lena, Filip and Bowen if you ever read this, I hope you remember our shared moments in Oxford.

My parents have been a constant source of support and love. It was their brave decision to migrate from China to Germany during their early adulthood, that endowed me with the fusion of the German rational mindset and the Chinese work ethic.

A very special thank you to Xuejie for her devoted encouragement and love that made the time during my DPhil the best time of my life. Yes, the DPhil has been exceptional hard work, but Xuejie balanced my life in her own caring and cosy way that made it enjoyable. We met on the very first day I arrived in Oxford and started to date soon after. We can be called a St Hugh's couple and a Hello Queen couple, depending on the perspective of what defines a relationship. Xuejie has shaped my time in Oxford, like no one else. This is partly, because we met each other on nearly every single day that we have stayed in Oxford - it's over two thousand days and counting - but also because we successfully navigated the highs and lows of Oxford life. My hope is that we will also successfully navigate the life beyond our graduate studies.

# Abstract

This thesis is about Step-indicator Saturation, an algorithm that identifies and models location shifts which are changes in the intercept of time series regression models. It contributes both to the asymptotic theory as well as applications of the algorithm. On theory, we develop asymptotic results about the empirical gauge of Step-indicator Saturation which is the frequency of false positives. The ability to control the empirical gauge is essential to have confidence that the location shifts identified by Step-indicator Saturation are not statistical artefacts. On applications, we use Step-indicator Saturation to study the empirical relationship between deindustrialisation and the real interest rate.

**Chapter 1.** We set out the mathematical framework for the asymptotic theory of Step-indicator Saturation. We establish empirical process results for differenced residuals, which introduce a specific time-series dependence between the indicator functions. We develop techniques to handle the time-series dependence within the traditional chaining argument that is used to show uniform convergence of empirical processes.

**Chapter 2.** We derive asymptotic results of the empirical gauge of a stylized version of Step-indicator Saturation that uses the split-half algorithm. We prove that the empirical gauge converges in probability to the population gauge which depends on the distribution of the error terms. Moreover, asymptotic distribution results of the empirical gauge are derived. We show that the asymptotic distribution the empirical gauge of Step-indicator Saturation is larger than for Impulse-indicator Saturation.

**Chapter 3.** We present two new ways to use Step-indicator Saturation in an empirical study. We first show that location shifts can hide an empirical relationship between two macroeconomic variables, in our case the relationship between deindustrialisation and the real interest rate. Step-indicator Saturation identifies two step-indicators that suffice to uncover a highly significant empirical relationship. Second, we apply the catalyst method to study the causal transmission of terms of trade shocks which first hit the manufacturing sector before they transmit into the real interest rate. Step-indicator Saturation is used to construct the catalyst from the location shifts in the Britsh terms of trade. We estimate that deindustrialisation can explain a 419 basis point reduction of the real interest rate from 1990 to 2017.

# Contents

**Introduction** **1**

**1 Empirical process results of temporal differences of residuals** **11**
1.1 Introduction and Model . . . . . 11
1.2 Assumptions . . . . . 16
1.3 Results . . . . . 18
1.3.1 Results on two-sided processes . . . . . 18
1.3.2 Results on one-sided processes . . . . . 21
1.3.3 Other general interest results . . . . . 22
1.4 Proofs . . . . . 25
1.4.1 A auxiliary metric and some inequalities . . . . . 25
1.4.2 Proof of Theorem 1.2.1 . . . . . 26
1.4.3 Proofs of general interest results . . . . . 27
1.4.4 Proofs of the empirical process results . . . . . 30
1.4.5 Proofs of the results of compensator terms . . . . . 44
1.4.6 Proof of Theorem 1.3.3 . . . . . 48
1.5 Conclusion . . . . . 50

**2 Gauging the Step-indicator Saturation Algorithm** **52**
2.1 Introduction . . . . . 52
2.2 Algorithm . . . . . 57
2.2.1 The taxonomy of split-half algorithms . . . . . 57
2.2.2 The split-half algorithm with step-indicators . . . . . 60
2.3 The gauge . . . . . 63
2.4 Assumptions . . . . . 65

2.5 Asymptotic results . . . 68
2.6 Simulation results . . . 72
2.6.1 Simulations on the convergence . . . 72
2.6.2 Simulations on the distribution . . . 74
2.7 Proofs . . . 77
2.7.1 A collection of preliminary theorems . . . 77
2.7.2 The empirical process representation of the gauge . . . 79
2.7.3 Proof of mean square convergence of gauge . . . 81
2.7.4 Proof of convergence in distribution . . . 82
2.8 Conclusion . . . 85

**3 Deindustrialisation: A secular force towards lower real rates 87**
3.1 Introduction . . . 87
3.2 Preliminary evidence . . . 90
3.2.1 Our measure of the deindustrialisation . . . 91
3.2.2 Our measure of the real rate . . . 92
3.2.3 Reduced form regressions . . . 94
3.3 Theory . . . 96
3.3.1 The setting of the 1-2-3 model . . . 98
3.3.2 The algebra of the 1-2-3 model . . . 100
3.3.3 A graphical illustration of the 1-2-3 model . . . 104
3.4 Empirical strategy . . . 107
3.4.1 The structure of causal transmission . . . 108
3.4.2 Our catalyst . . . 110
3.4.2.1 Catalysts from step-indicators . . . 110
3.4.2.2 The construction of the terms of trade catalyst . . . 111
3.5 Empirical Results . . . 114
3.6 Conclusion . . . 117

**Bibliography 120**

# List of Figures


1 An illustration of a location shift. . . . . 3
2 The bias introduced by location shifts in time-series . . . . . 4
3 Location shifts in British macroeconomic data . . . . . 6

2.1 An illustration of a step-indicator. . . . . 53
2.2 Split-half Step-indicator Saturation under the null . . . . . 61
2.3 Split-half Step-indicator Saturation under the alternative . . . . . 62
2.4 $\vartheta_1$ and $\vartheta_2$ for a standard normal $\mathsf{f}$ . . . . . 66
2.5 Simulations on the small sample bias of the gauge . . . . . 73
2.6 Simulations for the asymptotic variance with varying $c$ . . . . . 74
2.7 Decomposition of the asymptotic variance of the gauge . . . . . 75
2.8 Comparison of the asymptotic variance of the gauge . . . . . 76
2.9 The bias of Step-indicator Saturation in autoregressive models . . . 86

3.1 Comparison of deindustrialisation in Britain with other countries . 88
3.2 The real interest rate and deindustrialisation . . . . . 92
3.3 Scatter-plot of the real interest rate and deindustrialisation . . . . . 97
3.4 The effect of the China shock on output and consumption . . . . . 105
3.5 The effect of the China shock on the domestic capital market . . . 106
3.6 The British terms of trade . . . . . 112

# Introduction

In this thesis we study the Step-indicator Saturation algorithm that detects and models location shifts and that helps to alleviate the pernicious effect of unmodelled shifts. The Great Recession in 2008, the existential crisis of the Euro in 2012 and the Brexit and Trump presidency in 2016, have reminded economists about the abrupt changes to economic systems. These sudden changes in the level of macroeconomic variables go beyond models that assume stationarity or unit-root behaviour. They require an econometric time series model that can capture changes in the intercept, changes that we call location shifts.

The idea of Step-indicator Saturation is to add a saturating set of step-inductors to a regression equation and to remove the unnecessary variables in an iterative procedure. Because the regression equation has more variables than regressors, we have to add the regressors in blocks. Different algorithms exist to select relevant variables over blocks, with Autometrics being the most sophisticated algorithm that is used widely by practitioners. On the other end of the spectrum, the algorithm we study is the split-half algorithm, which simply splits the set of regressors into two subsets. We develop asymptotic theory that justifies setting the tuning parameter to target an average frequency of false positives.

Addressing the asymptotic theory of Step-indicator Saturation is important, because Step-indicator Saturation is being widely applied in the empirical literature. The Autometrics implementation of Step-indicator Saturation has been used by central banks: Ericsson (2017) study the bias of the US government forecast of the federal debt; Ericsson (2016) elicit GDP forecasts from FOMC's minutes around the financial crisis and evaluate them using Step-indicator Saturation. In climate econometrics, Pretis et al. (2015) use the procedure to study explanations for the slowdown in the rise of global mean surface temperatures. Ahumada and

Cornejo (2016) use the procedure to evaluate their forecasts on the price of corn, soybeans and wheat. While we have not yet developed the toolkit to study the asymptotic theory of Autometrics, the study of the split-half algorithm with Step-indicator Saturation is the first step towards a more comprehensive asymptotic theory of algorithms that search for relevant variables in blocks of regressors.

We conclude this thesis with new empirical applications of Step-indicator Saturation. First, we find location shifts in the conjectured co-integrating relationship between deindustrialisation and the real interest rate. Second, we apply the catalyst method to study the causal transmission of terms of trade shocks through the manufacturing sector to real rates. The terms of trade shocks are constructed as a collection of step-indicators that are identified with Step-indicator Saturation.

## Background: Location shifts and Step-indicator Saturation

In this section we define location shifts and show that they are ubiquitous in long-run macroeconomic data. We then outline the use of Step-indicator Saturation to model these shifts. We conclude with a definition of the empirical gauge as the frequency of false positives.

In Figure 1 we have the example for the simplest case of a single location shift. The data generating process for this example is, with $\mu_0 = 10$ and $\mu_1 = 0$

$$y_i = \mu_0 1_{(1 \leq i \leq 75)} + \mu_1 1_{(76 \leq i \leq 100)} + u_i, \quad i = 1, ..., 100,$$

where $u_i$ are i.i.d. realisations of a standard normal density. This single location shift is easy to spot, partly because of the large change in the mean with $\mu_0 - \mu_1 = 10$ relative to the unit standard deviation of the error terms $u_i$. More generally, we define location shifts as non-zero coefficients $\mu_\ell$ in the data generating process,

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

where the regressors $x_i$ contain an intercept and $\varepsilon_i$ have common reference density $\mathsf{f}$ and are independent from $\mathcal{F}_{i-1} = \sigma(x_1, ..., x_i, \varepsilon_1, ..., \varepsilon_{i-1})$. In fact, $\mathsf{f}$ can be very general and does not need to be symmetric. $L$ is the total number of breakpoints and the $\tau_\ell$'s refer to specific breakpoints which determine the timing of the location shifts.

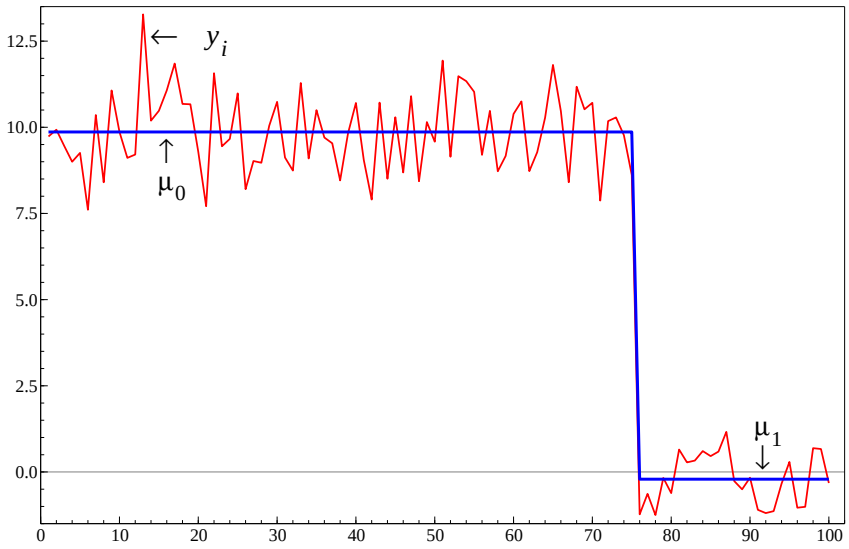



Figure 1: An illustration of a location shift.

There are many reasons for such location shifts in economic time series. They include among others technological progress, research and development, new legislation, institutional changes, regime shifts, financial innovation, shifting demography, evolving social and political mores, as well as conflicts and other major catastrophes.

Unmodelled location shifts give rise to three distinct statistical problems. First, model selection procedures are affected that start from a general unrestricted model and target a parsimonious and encompassing empirical model. Castle and Hendry (2014) document the effect of unmodelled location shifts on selection probabilities of the variables, their lags and non-linear functions. Second, parameter estimation is distorted. Hendry and Mizon (2011) show that an unmodelled location shift caused by the food stamp program of the Great Depression leads to implausible regression coefficients in their model for the US food expenditure. Third, as described in Clements and Hendry (1998) and Hendry (2000), unanticipated location shifts at or near the forecast origin lead to forecast failure.

We illustrate the distorted parameter estimation in a stylized example. Consider model (1) with $x_i = (1, y_{i-1})'$, $\beta = (\mu_0, \alpha_1)'$ and $\mathsf{f}$ as the standard normal density: this is the first order autoregressive model with an intercept. Let there

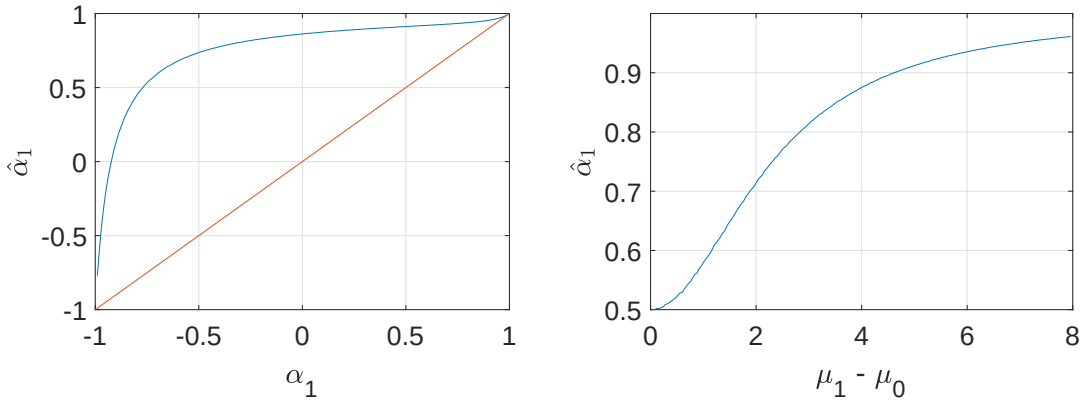



Figure 2: Simulation results on the bias introduced by a single location shift in a first-order autoregressive model. On the left plot we vary $\alpha_1$ and on the right plot we vary the magnitude of the location shift in the data generating process.

be a single location shift in the middle of the sample at $n/2$, let $\mu_0 = 1$ and set $n = 1,000,000$. We perform two experiments. In the first experiment, we hold constant the magnitude of the location shift, with $\mu_1 - \mu_0 = 5$. We vary $\alpha_1$ within the stationary parameter space. In the second experiment, we hold constant $\alpha_1 = 0.5$ and vary the magnitude of the location shift $\mu_1 - \mu_0$. The results are illustrated in Figure 2. We find in both experiments that the coefficient is biased towards unity. With both a large $\alpha_1$ and $\mu_1 - \mu_0$, the unit-root hypothesis becomes harder to reject. The bias does not vanish asymptotically.

In this doctoral thesis, we study Step-indicator Saturation which is a statistical procedure to model location shifts. It can be applied when the location, duration, magnitude and sign of multiple shift are unknown. It uses an extension of general-to-specific modelling to select significant step indicators from a saturating set to capture location shifts. The complete set of step-indicators is $\mathcal{S} = \{1_{(i \leq j)}, j = 1, ..., n-1\}$, where $1_{(i \leq j)} = 1$ for observations up to $j$, and zero elsewhere. As whole-sample vectors, step-indicators take the form $\iota_1' = (1, 0, 0, ..., 0), \iota_2' = (1, 1, 0, ..., 0), ..., \iota_{n-1}' = (1, 1, 1, ..., 1, 0)$. When we add the saturating set of step-indicators to a regression equation, say the equation (1) with $\mu_\ell = 0$ for all $\ell$, we have

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

which is infeasible to estimate as it contains more regressors as observations. There exist a host of algorithms to handle regression with more regressors than observations which are based on the idea to select regressors by adding them in blocks. On the two extremes there is the split-half algorithm and Autometrics. The split-half algorithm is the simplest way to implement the elimination of unnecessary step-indicators as it splits the saturating set of step-indicators into just two blocks. In contrast, Autometrics searches through expanding and contracting blocks of regressors and it is robust to a wide range of contamination with location shifts. Due to its robustness, it is the primary implementation of the algorithm that practitioners use. Doornik (2009) contains a more detailed discussion of Autometrics. Another implementation of Step-indicator Saturation which has been popular with practitioners is implemented in the R package *gets* as discussed in Pretis, Reade and Sucarrat (2016). They implement Step-indicator Saturation with a fixed block size, where indicators retained in each block are re-combined and selected over to yield the terminal models.

In Figure 3 we apply Step-indicator Saturation using Autometrics at a 1% selection probability, a metric we will define swiftly, to six British long-run macroeconomic time series from 1891 to 2016. They are GDP growth rates to measure economic growth, GDP deflator growth rates to measure changes in the prices, real consumption earnings growth to measure the growth in real incomes, the unemployment rate, the house price growth and the debt-to-GDP ratio. The data is from the 'A millennium of macroeconomic data'-dataset by the Bank of England (see Thomas and Dimsdale (2017)). Step-indicator Saturation detects a plethora of location shifts in each of the presented variables. Easily observable is the impact of the two world wars, which both resulted in a severe post-war recession and soaring war time inflation. We observe that World War 2 has disrupted the post-depression period of high unemployment and that it has ballooned the national debt-to-GDP ratio, only to be followed be a decades long deleveraging process. We also observe the impact of the different recessions that have hit the British economy and we see that the average GDP growth following each post-war recession has been lower than before. The frequent location shifts in these variables could be the driver for the failure to reject the random walk hypothesis for many macroeconomic variables, as documented in Nelson and Plosser (1982).

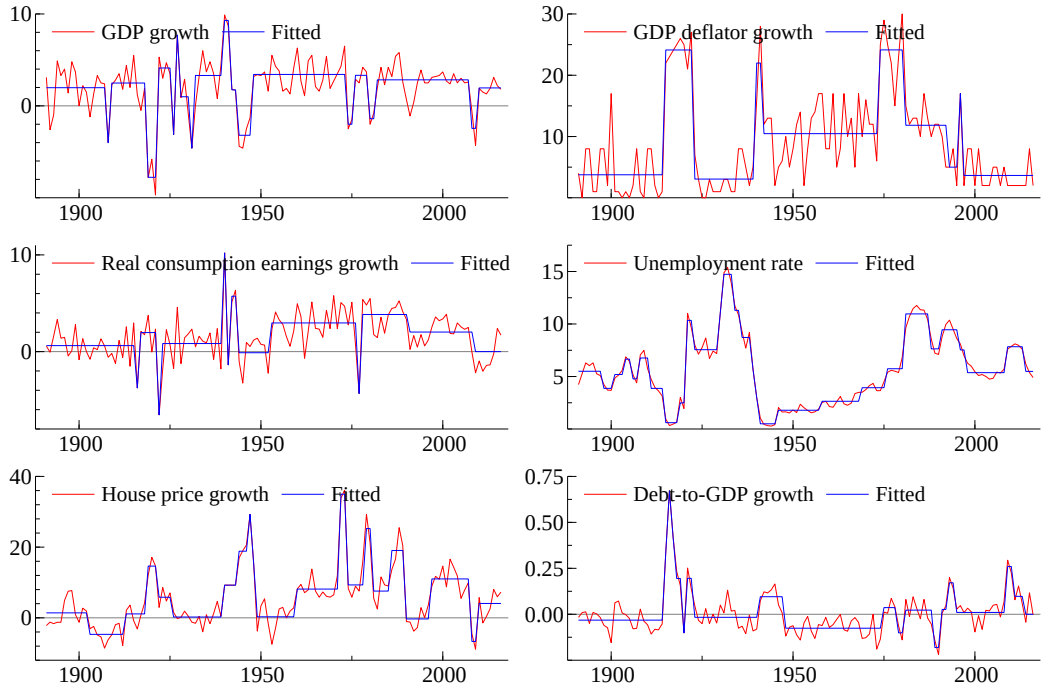



Figure 3: An illustration of location shifts in British macroeconomic time series from 1891 to 2016. Location shifts are detected using Step-indicator Saturation and Autometrics.

The key question that arises with Step-indicator Saturation is whether the location shifts that it finds are indeed location shifts in the data generating process. Only if the rate of false positives is sufficiently small, we can be confident that the identified breaks are true positives. That is why for Step-indicator Saturation, the only tuning parameter of the procedure is linked with the frequency of false positives. In fact, we aim to set the tuning parameter to match our desired frequency of false positives. The central result of this doctoral thesis will be to give an asymptotic justification to pick this tuning parameter.

We formalise the idea of the frequency of false positives by introducing the empirical gauge. When deciding on breakpoints, let $v_i = 0$ denote an observation classified as 'normal' and $v_i = 1$ an observation classified as an 'location shift'. Then, under the null of no location shift in the data generating process, the

empirical gauge is defined as the fraction of breakpoints found

$$\hat{\gamma}_n = n^{-1} \sum_{i=1}^{n} v_i. \tag{2}$$

The prefix 'empirical' is used to distinguish the empirical gauge from the population gauge which is a function of the distribution of the error terms $\varepsilon_i$. We will show that the empirical gauge converges in probability to the population gauge and we pick the tuning parameter to target our desired population gauge.

## Our contribution in this thesis

This thesis is conceptually split into two parts. In the first part, which contains the first two chapters, we study the asymptotic properties of the empirical gauge of Step-indicator Saturation. In the second part, we showcase a novel application of the algorithm in an empirical setting.

The mathematical foundation of our asymptotic results rests on a series of new empirical process results that we develop in the first chapter. While the traditional literature, i.e. Johansen and Nielsen (2009, 2016*a*); Berenguer-Rico and Nielsen (2017); Jiao and Nielsen (2017), concerns the empirical processes for residuals, we study empirical processes for differences of residuals where the empirical process includes

$$1_{(|\nabla \hat{\varepsilon}_i| \leq \sqrt{2}\hat{\sigma} c)} = 1_{(|\nabla \varepsilon_i - \nabla x_{in}' \hat{b})| \leq \sqrt{2}\sigma c + \sqrt{2} n^{-1/2} \hat{a} c)},$$

where $\nabla$ denotes the forward difference such that $\nabla \varepsilon_i = \varepsilon_i - \varepsilon_{i+1}$, $\hat{a} = n^{1/2}(\hat{\sigma} - \sigma)$ is the estimation error in the scale parameter, $\hat{b} = N^{-1}(\hat{\beta} - \beta)$ is the estimation error in the location parameter with some normalisation matrix $N$ and where $x_{in} = N x_i$. The differenced residuals introduce a specific dependency structure that manifests itself in a correlation between the $\nabla \varepsilon_i$'s and the $\nabla x_i$'s. Our main contribution is to develop mathematical techniques to handle the dependency structure that is introduced by differenced residuals. The central result will be that

$$\mathcal{G}_n(a, b, c) = \mathbb{G}_n(a, b, c) - \mathbb{G}_n(0, 0, c)$$

vanishes uniformly in $a, b$ where $\mathbb{G}_n(a, b, c) = n^{1/2}\{\widehat{\mathsf{G}}_n(a, b, c) - \overline{\mathsf{G}}_n(a, b, c)\}$ is the empirical process that consists of $\widehat{\mathsf{G}}_n(a, b, c) = n^{-1} \sum_{i=1}^{n} 1_{(|\nabla \varepsilon_i - \nabla x_{in}' b| \leq \sqrt{2}\sigma c + \sqrt{2} n^{-1/2} a c)}$

as the empirical distribution function and a compensator term with $\mathsf{E}_{i-1}(\cdot) = \mathsf{E}(\cdot|\mathcal{F}_{i-1})$ which is $\overline{\mathsf{G}}_n(a,b,c) = n^{-1}\sum_{i=1}^{n}\mathsf{E}_{i-1}1_{(|\nabla\varepsilon_i - \nabla x_{in}'b| \leq \sqrt{2}\sigma c + \sqrt{2}n^{-1/2}ac)}$.

After we have developed the mathematical tool-kit in the first chapter, we apply it in the second chapter to study the asymptotic properties of the empirical gauge of Step-indicator Saturation as defined in (2). We study the simplest version of Step-indicator Saturation which is based on split-half selection. The insight in the second chapter is that we can rewrite the empirical gauge of Step-indicator Saturation with split-half selection in an empirical process representation. This enables us to apply the empirical process results from the first chapter to study the asymptotic properties of the empirical gauge. Our first result is to show that the empirical gauge converges in probability to the population gauge. This result provides an asymptotic justification to pick the tuning parameter to match our desired population gauge, as we know that the empirical gauge will converge to the population gauge. When we set the tuning parameter to match, say the 1% selection probability, then asymptotically, we will have 1% false positives. With a sample size of $n = 100$ this is equivalent to one false positive on average. The second result concerns the asymptotic distribution of the empirical gauge which gives us an understanding about the variation of the empirical gauge around the population gauge. This is the first asymptotic theory that was developed for the Step-indicator Saturation estimator.

We then turn to the second part of the thesis with aims to showcase an application of Step-indicator Saturation in an empirical setting. We do so in the third chapter, by investigating the reasons for the low real interest rate environment of the past decade. Few economists would have believed during the Great Recession of 2008 that British interest rates would stay in negative territory for over ten years to come and that the Bank of England would purchase close to a third of government debt. Upon hearing this news, most economists would have confidently predicted that such monetary policy would command inflation to unprecedented heights. However, inflation has mostly remained well below its target of two percent and only the exchange rate devaluation following the Brexit referendum has lifted it briefly to three percent.

Academic economists have revived the idea of the secular stagnation to explain this phenomena. First put forward in Hansen (1939), it posits that the industrial

economies suffer from an imbalance resulting from an increasing propensity to save and a decreasing propensity to invest. As a consequence of the imbalance between savings and investments the long-term real interest rates are dragged down.

We put forward a related hypothesis which states that the deindustrialisation of Britain has lowered the demand for capital and thus resulted in lower real rates. This is in the context of globalisation and the shift of the manufacturing supply chain to low labour cost countries in Asia. We formalise the idea with the 1-2-3 trade model. In this model, is are terms of trade shocks that explain why deindustrialisation is associated with lower real rates.

The central contribution of the third chapter is to find a causal effect of the fall in the share of manufacturing output on real rates. The causality is established with the application of the 'catalyst method' by Bazinas and Nielsen (2015). In general, the deduction of a causal ordering between two observed variables from their joint distribution is difficult. However, if we can find a third variable, which represents large exogenous shocks we name catalysts, we can determine the order in which this shock transmits between the two variables of interest. A set of testable conditions can then be used to establish the causal transmission of the catalyst between the two endogenous variables. It pins down a unique direction of the influence from one variable to another.

Based on our theoretical framework, which suggests that terms of trade shocks can explain why deindustrialisation is associated with lower real rates, we use them as our catalyst. Our empirical results establish the causal transmission of our terms of trade catalyst to the real rate through the manufacturing sector. The causal transmission suggests a unique causal ordering in which the terms of trade shocks cause deindustrialisation which in turn result in a fall of the real rate. On average a one percent point decline of the share of manufacturing output resulted in a 79 basis point decline in the real rate. From 1990 to 2017, this accounts for a 419 basis point decline of the real rate.

In the third chapter, we use Step-indicator Saturation in two ways. First, we use it to uncover the empirical relationship between deindustrialisation and the real rate. In a univariate regression model of the real rate, our measure for deindustrialisation is insignificant. It is only the inclusion of two step-indicators that we identify with Step-indicator Saturation that uncover the highly significant

relationship. Second, the terms of trade catalyst is constructed from a collection of step-indicators that we identify with Step-indicator Saturation. Starting from the theoretical model that argues that terms of trade shocks can explain why deindustrialisation is coupled with lower real interest rates, we run Step-indicator Saturation to find the location shifts in the British terms of trade variable. Out of these location shifts, we select the subset of structural breaks that best fit the conditions for causal transmission.

## Probability theory conventions:

- $X_n \xrightarrow{\mathsf{P}} X$ denotes the convergence in probability of the random variable $X_n$ to the (possibly constant) random variable $X$ as $n \to \infty$.
- $X_n \xrightarrow{\mathsf{D}} L$ denotes the (weak) convergence of the distribution of the random variables $X_n$ to $L$ if $L$ is a distribution, respectively to the distribution of $L$ if $L$ is a random variable.
- $X \overset{\mathsf{D}}{=} L$ means that the random variable $X$ has the distribution $L$.
- For a sequence of random variables $X_n$ and a deterministic sequence $a_n$ the stochastic order notation is defined as follows (see Brockwell and Davis (1991)):
  $X_n = \mathrm{o}_{\mathsf{P}}(a_n)$ if and only if $X_n/a_n \xrightarrow{\mathsf{P}} 0$;
  $X_n = \mathrm{O}_{\mathsf{P}}(a_n)$ if and only if $X_n/a_n$ is *bounded in probability*, i.e. for all $\epsilon > 0$ there exists a $B(\epsilon) > 0$ and $n_0$ such that $\mathsf{P}\{|X_n/a_n| > B(\epsilon)\} < \epsilon$ for all $n > n_0$.
- $X_n \sim a_n$ if and only if $X_n = \mathrm{O}_{\mathsf{P}}(a_n)$ and $a_n = \mathrm{O}_{\mathsf{P}}(X_n)$.

# Chapter 1

# Empirical process results of temporal differences of residuals

Abstract. We investigate a class of empirical process results of temporal differences of residuals. These results are based on a large literature that has considered empirical processes of residuals. In contrast to residuals, temporal differences of residuals introduce more intricate temporal dependence between the indicator functions. In this chapter we showcase a series of techniques that we have developed to tackle the dependence of indicator functions and use them to derive some properties of our empirical processes under study.

At the heart of our results sits an exponential martingale inequality, in our case the Freedman inequality. Based on this inequality, we develop an inequality to bound tail probabilities of certain families of martingales. These tail probabilities are then used in a chaining argument to prove the tightness of our empirical processes of temporal differences of residuals. The results can be used to develop asymptotic properties of statistical procedures such as the Step-indicator Saturation estimator or Q-Q and P-P plots.

## 1.1 Introduction and Model

We consider a sequence of random variables $\{x_i, y_i\}_{i=1}^{n+1}$ which are generated by a linear time-series regression model

$$y_i = x_i'\beta + \varepsilon_i, \tag{1.1}$$

where $y_i$ is a one-dimensional dependent variable, $x_i$ is a vector of $\dim x$ contemporaneous regressors which contain the intercept and $\beta$ is the vector of $\dim x$ regression coefficients. The error term $\varepsilon_i$ is independent of the filtration $\mathcal{F}_{i-1} = \sigma(x_1, \ldots, x_i, \varepsilon_1, \ldots, \varepsilon_{i-1})$ and identically distributed and we say that $\varepsilon_i/\sigma$ has density $\mathsf{f}$ and a first moment centred around the zero. In most cases, the density $\mathsf{f}$ will be the normal density. We allow the regressors to be from a wide range of different classes, namely stationary regressors, trend stationary regressors and unit-root regressors. Because we use martingale arguments for our proofs the regressors can have an arbitrary time-series correlation structure. A type of regressor we do not allow for are explosive regressors. We choose to have the sample size $n+1$ to allow the empirical processes defined in (1.4), which contains random variables $\hat{\varepsilon}_i$ and $\hat{\varepsilon}_{i+1}$ in its indicator functions, to sum up to $n$.

We introduce a shorthand notation to handle the $\sigma$ and $\sqrt{2}\sigma$ normalisations of the error terms $\varepsilon_i$, the differenced errors $\nabla\varepsilon_i = \varepsilon_i - \varepsilon_{i+1}$ and the regressors $x_i$. This notation will be used throughout this chapter. We denote

$$\tilde{\varepsilon}_i = (2\sigma^2)^{-1/2}\varepsilon_i, \quad \check{\varepsilon}_i = \sigma^{-1}\varepsilon_i, \quad \tilde{x}_i = (2\sigma^2)^{-1/2}x_i, \quad \check{x}_i = \sigma^{-1}x_i. \tag{1.2}$$

Further, define the estimation error in the scale and location parameters, as well as the normalised regressors

$$\hat{a} = n^{1/2}(\hat{\sigma} - \sigma), \quad \hat{b} = N^{-1}(\hat{\beta} - \beta), \quad x_{in} = N'x_i.$$

We choose the normalisation matrix $N$ depending on the stochastic properties of the regressors $x_i$, so that $\sum_{i=1}^n x_{in}x_{in}'$ converges, $n^{-1/2}\sum_{i=1}^n |x_{in}|$ is bounded, and $x_i'(\hat{\beta} - \beta) = x_{in}'b$. If $(y_i, x_i)$ is stationary, we pick $N = n^{-1/2}I_{\dim x}$ and thus $\hat{b} = n^{1/2}(\hat{\beta} - \beta)$ and $x_{in} = n^{-1/2}x_i$. If $x_i$ is a random walk, we select $N = n^{-1}I_{\dim x}$ and $\hat{b} = n(\hat{\beta} - \beta)$ and $x_{in} = n^{-1}x_i$. With trend stationary regressors, we choose $N = n^{-3/2}I_{\dim x}$ and $\hat{b} = n^{3/2}(\hat{\beta} - \beta)$ and $x_{in} = n^{-3/2}x_i$. With $x_i$ consisting of different types of regressors, we have for example

$$N = \begin{pmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{pmatrix} \text{ when } x_i = \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

We define the empirical process of differenced residuals, by considering its components first: the empirical distribution function and the (pseudo-) compensator

term. Note that we define the empirical process as a function of three parameters, the estimation error in the scale $a$, the estimation error in the location $b$ and a cut-off value $c$ which determines the threshold at which the indicator function turns from zero to one. The two-sided empirical distribution function is

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

where as before, $\nabla$ denotes the forward difference, such that $\nabla\varepsilon_i = \varepsilon_i - \varepsilon_{i+1}$ and $\nabla x_i = x_i - x_{i+1}$. We have $\sigma + n^{-1/2}a > 0$ by our choice of $a$ as $\hat{a} = n^{1/2}(\hat{\sigma} - \sigma)$. In this case $\sigma + n^{-1/2}a = \hat{\sigma} > 0$ and we only need to consider $c > 0$. The (pseudo-) compensator is

$$\overline{\mathsf{G}}_n(a,b,c) = n^{-1}\sum_{i=1}^{n} \mathsf{E}_{i-1} 1_{(|\nabla\varepsilon_i - \nabla x_{in}'b| \leq \sqrt{2}\sigma c + \sqrt{2}n^{-1/2}ac)}, \tag{1.3}$$

where $\mathsf{E}_{i-1}(\cdot) = \mathsf{E}(\cdot|\mathcal{F}_{i-1})$. The pseudo prefix derives from the fact that $1_{(\nabla\varepsilon_i \leq c)}$ is $\mathcal{F}_{i+1}$ measurable and not $\mathcal{F}_i$-measurable. The usual compensator term would have been defined using $\mathsf{E}_i 1_{(\cdot)}$. Combine both terms to obtain the empirical process that we will study in this chapter:

$$\mathbb{G}_n(a,b,c) = n^{1/2}\{\widehat{\mathsf{G}}_n(a,b,c) - \overline{\mathsf{G}}_n(a,b,c)\}. \tag{1.4}$$

Given the definition of the indicator function, the distinct feature of our study relates to the forward differenced residuals that we consider

$$1_{(|\nabla\hat{\varepsilon}_i| \leq \sqrt{2}\hat{\sigma}c)} = 1_{(|\nabla\varepsilon_i - \nabla x_{in}'\hat{b})| \leq \sqrt{2}\sigma c + \sqrt{2}n^{-1/2}\hat{a}c)}, \tag{1.5}$$

where $\hat{\varepsilon}_i = \varepsilon_i - x_{in}'b$. There is a vast literature on empirical processes for residuals, to name a few that are most closely related the work in this chapter, see Koul (2002); Koul and Ossiander (1994), Giraitis et al. (2012), Engler and Nielsen (2009), Johansen and Nielsen (2009, 2016*a*,*b*) and Jiao and Nielsen (2017). These papers consider indicator functions of the form

$$1_{(|\hat{\varepsilon}_i| \leq \hat{\sigma}c)} = 1_{(|\varepsilon_i - x_{in}'\hat{b}| \leq \sigma c + n^{-1/2}\hat{a}c)},$$

where the $\varepsilon_i$ in the indicator functions are independent from each other. To our knowledge, we are the first to study empirical processes of temporal differences of residuals.

The contribution of this chapter is to find mathematical techniques to handle the temporal dependence of the indicator functions in the empirical processes with differenced residuals. The dependence of adjacent indicator functions with differenced residuals has two sources. Given fixed estimation errors $a, b = 0$

$$\mathsf{cov}\{1_{(|\nabla\tilde{\varepsilon}_i|\leq c)}, 1_{(|\nabla\tilde{\varepsilon}_{i+1}|\leq c)}\} = \mathsf{E}\{1_{(|\nabla\tilde{\varepsilon}_i|\leq c)}1_{(|\nabla\tilde{\varepsilon}_{i+1}|\leq c)}\} - \mathsf{E}\{1_{(|\nabla\tilde{\varepsilon}_i|\leq c)}\}\mathsf{E}\{1_{(|\nabla\tilde{\varepsilon}_{i+1}|\leq c)}\},$$

where $\mathsf{E}\{1_{(|\nabla\tilde{\varepsilon}_i|\leq c)}1_{(|\nabla\tilde{\varepsilon}_{i+1}|\leq c)}\} = \int_{-\sqrt{2}c}^{\sqrt{2}c}\int_{-\sqrt{2}c}^{\sqrt{2}c}\int_{-\infty}^{\infty}\mathsf{f}(x+y+z)\mathsf{f}(x+z)\mathsf{f}(z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x$. While for indicator functions of the type $1_{(|\tilde{\varepsilon}_i|\leq c)}$, we get $\mathsf{cov}\{1_{(|\tilde{\varepsilon}_{i+1}|\leq c)}, 1_{(|\tilde{\varepsilon}_i|\leq c)}\} = 0$, with differenced residuals, the corresponding covariance is usually non-zero. However, the temporal dependence is limited to temporally adjacent indicator functions. For $j > 1$ or $j < -1$ we have

$$\mathsf{cov}\{1_{(|\nabla\tilde{\varepsilon}_i|\leq c)}, 1_{(|\nabla\tilde{\varepsilon}_{i+j}|\leq c)}\} = 0.$$

In this case, the indicator functions each contain a non-overlapping subset of error terms, which are therefore independent of each other.

With non-zero estimation error $b$, we have to consider the dependence between the regressors $\nabla\varepsilon_i$ and $\nabla x_i$ as well. In the existing literature considering indicators of residuals, this is not a problem, because we could assume the regressors $x_i$ to be $\mathcal{F}_{i-1}$-adapted. When we would consider the compensator term, for example, within the conditional expectation $\mathsf{E}_{i-1}1_{(|\varepsilon_i - x_i'b|\leq\sigma c+n^{-1/2}ac)}$, only $\varepsilon_i$ is not $\mathcal{F}_{i-1}$-adapted. With forward differenced error terms, however, in $\mathsf{E}_{i-1}1_{(|\nabla\tilde{\varepsilon}_i - \nabla\tilde{x}_i'b|\leq c+n^{-1/2}\sigma^{-1}ac)}$ we have three random variables $\varepsilon_i$, $\varepsilon_{i+1}$ and $x_{i+1}$ that are not $\mathcal{F}_{i-1}$-adapted. The new theorems that we prove in this chapter are about ways to handle these two types of dependence of the indicator functions.

We assume that the residuals are formed over a conformable subsample, where the location parameter $\beta$ and the scale parameter $\sigma$ can be consistently estimated. We do not explicitly model how these initial estimators $\hat{\beta}$ and $\hat{\sigma}$ are obtained. Our asymptotic results will hold uniformly over the set of conformable initial estimates. As the empirical process results that we study are often used in robust statistics, where we are interested to estimate regression coefficients in the presence of contamination, the question of how to obtain robust initial estimates for $\beta$ and $\sigma$ is an important open question that we do not address.

The empirical process results presented in this paper are used in the second chapter of the thesis. There, we derive asymptotic convergence and distribution results for the gauge of Step-indicator Saturation. The results can also be used to analyse Q-Q plots and P-P plots as in Engler and Nielsen (2009).

All our results relate to studying the impact of estimation errors $\hat{a}, \hat{b}$ on our empirical processes results of interest. For example, our central result will study the distance of two empirical processes as in $|\mathbb{G}_n(\hat{a}, \hat{b}, c) - \mathbb{G}_n(0, 0, c)|$. While we will show that this distance vanishes in probability, for the corresponding distance defined on our compensator terms $\sqrt{n}|\overline{\mathsf{G}}_n(\hat{a}, \hat{b}, c) - \overline{\mathsf{G}}_n(0, 0, c)|$ we will require a first-order bias term.

When we consider the impact of the estimation errors $\hat{a}, \hat{b}$, they remain random variables. We make the problem tractable, by transforming the estimation errors from random variables to constants and then showing uniform convergence of the distances of interest over $(a, b)$ varying in some compact set. Because of this transformation, the results of this chapter are uniform convergence results.

The framework for our proofs of the empirical process results rests on the foundation of Johansen and Nielsen (2016*a*), where the authors analyse the asymptotic properties of the Forward Search algorithm. As in their paper, we will engage in a chaining argument, to show the uniform convergence of empirical processes over the space of estimation errors in the location and scale parameters. The idea of the chaining argument is to partition the parameter space into smaller subsets, by covering them with balls around a collection of grid points. This partitioning separates the empirical process into discrete point terms, which have parameter values centred on the grid points, and oscillation terms, which have parameter values in-between the grid points. The chaining argument concludes when we are able to show that both the discrete point terms and the oscillation terms vanish.

Because we are analysing empirical processes for temporal differences of residuals instead of empirical processes for residuals, we faced two challenges. First, we needed to design a chaining argument that works for differenced residuals in the indicator functions of the empirical processes. Second, the dependent nature of the differenced residuals complicates the classic Taylor expansion needed to derive the bias term of the compensator.

We proceed as follow. In section 1.2 we state and discuss the assumptions. In section 1.3 we present our main empirical process results. In the final section 1.4 we prove our theorems, starting from the auxiliary theorems, before pulling the strings together for the proofs of the main results.

## 1.2 Assumptions

In this section we discuss the assumptions for the empirical process results that we will present. Before we state the assumptions, we introduce more notation about the various probability distributions that we will form assumptions over. We introduce the density for $\nabla\tilde{\varepsilon}_i$, recalling $\tilde{\varepsilon}_i = (2\sigma^2)^{-1/2}\varepsilon_i$ in (1.2), to have

$$\mathsf{h}(x) = \sqrt{2}\int_{-\infty}^{\infty} \mathsf{f}(y)\mathsf{f}(\sqrt{2}x + y)\,\mathrm{d}y. \tag{1.6}$$

It is a convolution of the density $\mathsf{f}$ for random variable $\tilde{\varepsilon}_i$. We highlight four properties of the density $\mathsf{h}$ in the following Theorem.

**Theorem 1.2.1.** *Assume $\varepsilon_i$ are i.i.d. and continuous with density $\mathsf{f}$. The density $\mathsf{h}$ then satisfies the following properties:*

(*a*) *Symmetry:* $\mathsf{h}(x) = \mathsf{h}(-x)$;

(*b*) *Suppose* $\mathsf{f}$ *has second moments. Then* $\mathsf{f} = \mathsf{h}$ *if and only if* $\mathsf{f}$ *is normal;*

(*c*) *If* $\sup_{v\in\mathbb{R}} |v|^k \mathsf{f}(v) < \infty$, *then* $\sup_{v\in\mathbb{R}} |v|^k \mathsf{h}(v) < \infty$ *for* $k \in \mathbb{N}_0$;

(*d*) *If* $\sup_{v\in\mathbb{R}} |v|^k |\dot{\mathsf{f}}(v)| < \infty$ *and* $\mathsf{E}|\varepsilon_i^k| < \infty$, *then* $\sup_{v\in\mathbb{R}} |v^k \dot{\mathsf{h}}(v)| < \infty$ *for* $k \in \mathbb{N}_0$.

Define the distribution function corresponding to $\mathsf{h}$ as $\mathsf{H}(c) = \mathsf{P}(\nabla\tilde{\varepsilon}_i \leq c)$. The absolute value of the scaled forward differences then has distribution function

$$\mathsf{G}(c) = \mathsf{P}(|\nabla\tilde{\varepsilon}_i| < c) = \int_{-c}^{c} \mathsf{h}(x)dx. \tag{1.7}$$

The conditional joint density $\mathsf{m}_i(y, x)$ given $\mathcal{F}_{i-1}$ is the density of the two random variables $\nabla\tilde{\varepsilon}_i$ and $\nabla\tilde{x}_i$. Let $\mathsf{k}_i(x)$ be the conditional marginal density of the regressors $\nabla\tilde{x}_i$ conditional on $\mathcal{F}_{i-1}$. Then the conditional density $\mathsf{m}_i(x|y)$ given

$\mathcal{F}_{i-1}$ and the derivative of the conditional density $\mathsf{m}_i(y|x)$ given $\mathcal{F}_{i-1}$ with respect to the differenced error term, if it exists, is

$$\mathsf{m}_i(c|y) = \frac{\mathsf{m}_i(y,c)}{\mathsf{h}(y)} \quad \text{and} \quad \dot{\mathsf{m}}_i(c|x) = \frac{1}{\mathsf{k}_i(x)} \left. \frac{\partial \mathsf{m}_i(y,x)}{\partial y} \right|_{y=c}. \tag{1.8}$$

Finally, define the third and fourth moment of the scaled innovations $\check{\varepsilon}_i$ as

$$\varkappa_3 = \int_{-\infty}^{\infty} u^3 \mathsf{f}(u) du, \quad \varkappa_4 = \int_{-\infty}^{\infty} u^4 \mathsf{f}(u) du. \tag{1.9}$$

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

**Assumption 1.2.2.** *Let $\mathcal{F}_i$ be a filtration so that $\varepsilon_{i-1}$ and $x_i$ are $\mathcal{F}_{i-1}$-adapted and $\check{\epsilon}_i$ is independent of $\mathcal{F}_{i-1}$ with distribution function $\mathsf{F}$ and positive density $\mathsf{f}$ on $\mathbb{R}$ with derivative be $\dot{\mathsf{f}}$.*

Most of the assumptions in this chapter are similar to Johansen and Nielsen (2016*a*). The most notable difference is the new assumption on the conditional density $\mathsf{m}_i(y|x)$ given $\mathcal{F}_{i-1}$, which was defined in (1.8). We use this assumption to control the dependence between the error terms $\nabla\varepsilon_i$ and the regressors $\nabla x_i$.

**Assumption 1.2.3.** *Set $0 \leq \kappa < \eta \leq 1/4$ and suppose*
*(i) the marginal density $\mathsf{f}$ satisfies*
*(a) moments: $\int_{-\infty}^{\infty} u\mathsf{f}(u)du = 0$ and $\int_{-\infty}^{\infty} |u^2|\mathsf{f}(u)du < \infty$;*
*(b) boundedness: $\sup_{v\in\mathbb{R}} \mathsf{f}(v)$, $\sup_{v\in\mathbb{R}} |v|\mathsf{f}(v)$, $\sup_{v\in\mathbb{R}} |\dot{\mathsf{f}}(v)|$, $\sup_{v\in\mathbb{R}} v^2|\dot{\mathsf{f}}(v)| < \infty$;*
*(ii) the conditional density $\mathsf{m}_i(y|x)$ given $\mathcal{F}_{i-1}$, see (1.8), exists and satisfies*
$\max_{1\leq i\leq n} \sup_{y\in\mathbb{R},x\in\mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)| < \infty$;
*(iii) the regressors $x_i$ satisfy $n^{-1}\mathsf{E}\sum_{i=1}^{n} |n^{1/2}N'\nabla x_i|^2 = \mathrm{O}(1)$.*

The Assumption 1.2.3(ii) simplifies, if the joint density of $\mathsf{m}_i$ given $\mathcal{F}_{i-1}$ follows the multivariate normal distribution. This would be satisfied, for example, for an autoregressive regression model with normally distributed error terms $\varepsilon_i$. Moreover, Assumption 1.2.3(i) is satisfied if the density $\mathsf{f}$ is the normal or t-density. Assumption 1.2.3(iii) is satisfied for stationary, random walk regressors and trend stationary regressors $x_i$, given sufficient moments exist. A more detailed discussion can be found in Johansen and Nielsen (2016*a*).

**Remark 1.2.1.** *If the conditional joint density* $\mathsf{m}_i$ *given* $\mathcal{F}_{i-1}$ *follows the multivariate normal distribution, 1.2.3(ib) would imply 1.2.3(ii). To see this, write the conditional density as,*

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

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

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

*With the normal density* $\phi$ *we have* $\sup_{v\in\mathbb{R}} |v\phi(v)| < \infty$ *and consequently it holds that* $\sup_{y\in\mathbb{R}, x\in\mathbb{R}^p} |\dot{\mathsf{f}}_{Y|X}(y|x)| < \infty$*.*

# 1.3 Results

We structured the presentation of our results into two subsections. In the first subsection, we present results for empirical processes that contain two-sided indicator functions, such as the indicator function displayed in (1.5). We then proceed to show in the second subsection a collection of empirical process results about one-sided indicator functions. In the proofs, most of the one-sided empirical process results will be used to derive the two-sided results.

## 1.3.1 Results on two-sided processes

The first theorem is also the central theorem of this chapter. It allows us to control perturbations of the empirical process $\mathbb{G}_n(a, b, c)$ in (1.4) that are due to the estimation error of the location and the scale parameters from a regression.

**Theorem 1.3.1.** *Suppose Assumption 1.2.2 and 1.2.3(ib,iii) hold. Let*

$$\mathcal{G}_n(a, b, c) = \mathbb{G}_n(a, b, c) - \mathbb{G}_n(0, 0, c). \tag{1.10}$$

*Then, for any* $B > 0$ *and* $c \in \mathbb{R}$*, we have* $\sup_{|a|,|b|\leq n^{1/4-\eta_B}} |\mathcal{G}_n(a, b, c)| = \mathsf{o_P}(1)$*.*

Recall from section 1.1 that $x_{in} = Nx_i$ refers to normalised regressors. Define a column vector related to the covariance of the differenced regressors $\nabla\tilde{x}_{in}$ and the differenced error terms $\nabla\tilde{\varepsilon}_{in}$

$$n^{-1}\sum_{i=1}^{n}\mathsf{E}_{i-1}(n^{1/2}\nabla\tilde{x}_{in}|\nabla\tilde{\varepsilon}_i = c) \xrightarrow{\mathsf{D}} \xi(c) \tag{1.11}$$

Consider next the difference of two $\xi(c)$ functions

$$\xi_c = \xi(c) - \xi(-c). \tag{1.12}$$

The term $\xi_c$ vanishes in two distinct cases. If the regressors are strictly exogenous, then $\mathsf{E}_{i-1}\left(\nabla\tilde{x}_{in}|\,\nabla\tilde{\varepsilon}_i = c\right) = \mathsf{E}_{i-1}\nabla\tilde{x}_{in}$, which does not depend on $c$ and we have $\xi_c = 0$ for all $c$. Alternatively, if the regressors are of the random walk type, then as $\mathsf{E}_{i-1}\left(\nabla\tilde{x}_i|\,\nabla\tilde{\varepsilon}_i = c\right)$ is of order O(1), the normalisation in $x_{in} = Nx_i$, which is $N = n^{-1}$, dominates the expression and $\xi(c)$ vanishes for all $c$. The same holds for trend stationary regressors, as in that case $N = n^{-3/2}$. Thus, only for stationary regressors $\xi_c$ does not vanish.

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

If $\nabla\tilde{x}_i$ and $\nabla\tilde{\varepsilon}_i$ are also jointly normally distributed and $\nabla\tilde{x}_i = v_i + w_i$ where $v_i$ is $\mathcal{F}_{i-1}$ measurable and $w_i$ is independent of $\mathcal{F}_{i-1}$ with $\mathsf{E}v_i = \mathsf{E}w_i = 0$ and

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

then $n^{-1}\sum_{i=1}^{n}(v_i + c\sigma_{w\varepsilon}) \xrightarrow{\mathsf{P}} c\sigma_{w\varepsilon}$. We then get for the term $\xi(c) - \xi(-c) = 2c\sigma_{w\varepsilon}$. For example, if $y_i = \alpha_1 y_{i-1} + \varepsilon_i$ and $-1 < \alpha_1 < 1$, we get $\sigma_{\omega\varepsilon} = -\alpha_1$ and thus $\xi(c) - \xi(-c) = -2c\alpha_1$.

The $\xi_c$ term emerges due to the dependence of $\nabla\varepsilon_i$ and $\nabla x_i$. It is used to express the first-order bias in the Taylor expansion of the quasi-compensator term which we will require in Theorem 1.3.2.

**Theorem 1.3.2.** *Suppose Assumptions 1.2.2 and 1.2.3(i,ii,iii) hold. Introduce the term*

$$\overline{\mathcal{G}}_n(a,b,c) = n^{1/2}\{\overline{G}_n(a,b,c) - \overline{G}_n(0,0,c)\} - \mathcal{B}(a,b,c), \tag{1.13}$$

*where the bias term, with $\xi_c$ defined as* (1.12)*, is*

$$\mathcal{B}(a,b,c) = \mathsf{h}(c)(2ac\sigma^{-1} + \xi_c' b). \tag{1.14}$$

*Then, for all $B > 0$ we have* $\sup_{c\in\mathbb{R}} \sup_{|a|,|b|\leq n^{1/4-\eta}B} \left|\overline{\mathcal{G}}_n(a,b,c)\right| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta})$.

The proof of these results relates to those of Koul and Ossiander (1994), Johansen and Nielsen (2016*a*), Jiao and Nielsen (2017) and Berenguer-Rico and Nielsen (2017). The main difference is that those papers have indicators for residuals whereas we consider indicators for differences of residuals. This gives rise to the more complicated dependence structure between the random variables in the indicator function.

We also consider an alternative empirical distribution function $\widehat{\mathsf{H}}_n$. Define

$$\widehat{\mathsf{H}}_n(a,b,c,\delta_n) = n^{-1}\sum_{i=1}^{n} 1_{\{|\nabla\tilde{\varepsilon}_i - |c|(1+n^{-1/2}\sigma^{-1}|a|) - \nabla\tilde{x}_{in}'b| \leq |c|n^{-1/2}\sigma^{-1}\delta_n\}}, \tag{1.15}$$

where $\delta_n$ is some sequence of real numbers. The $\widehat{\mathsf{H}}_n$ term emerges when we consider as the estimate of the scale parameter not $\hat{\sigma}^2$, but instead $\hat{\omega_i}^2 = \hat{\sigma}^2\{2 + \nabla x_i'(\sum_{i=1}^n x_i x_i')^{-1}\nabla x_i\}$ which follows the usual finite sample formula for the distribution of residuals for fixed regressors. We will derive this finite sample formula in section 2.2.1 and in particular in equation (2.1). We will in section 2.7.2 show how to handle indicator functions involving $\hat{\omega}_i$ using the following theorem, which holds uniformly in $a$ and $c$.

**Theorem 1.3.3.** *Suppose Assumptions 1.2.2 and 1.2.3(ia,ib,iii) hold. Then, for any $B > 0$ and $c \in \mathbb{R}$ it follows for some $\delta_n \to 0$ as $n \to \infty$ that*

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

### 1.3.2 Results on one-sided processes

Theorem 1.3.1 concerns the empirical process $\mathbb{G}_n$ which contains two-sided indicator functions such as $1_{(|\nabla\tilde{\varepsilon}_i|\leq c)}$. To prove Theorem 1.3.1, we first decompose the two-sided indicator function to two one-sided indicator functions of the form $1_{(\nabla\tilde{\varepsilon}_i\leq c)}$, noting that

$$1_{(|\nabla\tilde{\varepsilon}_i|\leq c)} = 1_{(\nabla\tilde{\varepsilon}_i\leq c)} - 1_{(\nabla\tilde{\varepsilon}_i<-c)}$$

Define the one-sided empirical distribution function with differenced error terms and estimation error in the scale $a$ and location $b$ as

$$\widehat{\mathsf{F}}_n(a,b,c) = n^{-1}\sum_{i=1}^{n} 1_{(\nabla\tilde{\varepsilon}_i\leq c+N_a ac+\nabla\tilde{x}'_{in}b)}, \tag{1.16}$$

which has the (pseudo-)compensator

$$\overline{\mathsf{F}}_n(a,b,c) = n^{-1}\sum_{i=1}^{n} \mathsf{E}_{i-1}1_{(\nabla\tilde{\varepsilon}_i\leq c+N_a ac+\nabla\tilde{x}'_{in}b)}, \tag{1.17}$$

and which together form the one-sided empirical process

$$\mathbb{F}_n(a,b,c) = n^{1/2}\{\widehat{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(a,b,c)\}. \tag{1.18}$$

The equivalent theorem to Theorem 1.3.1 that involves one-sided empirical processes, is stated in the next theorem.

**Theorem 1.3.4.** *Suppose Assumptions 1.2.2 and 1.2.3(ib) are satisfied. Then, for any $B > 0$*

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

**Theorem 1.3.5.** *Suppose Assumptions 1.2.2 and 1.2.3(ib,iii) are satisfied. Then, for any $B > 0$ and $c \in \mathbb{R}$*

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

We note here that we can control uniform convergence of $|\mathbb{F}_n(a,b,c)-\mathbb{F}_n(0,0,c)|$ over $a,b$ and $a,c$, but that we are still unable to show uniform convergence over all three $a,b,c$.

**Theorem 1.3.6.** *Suppose Assumptions 1.2.2 and 1.2.3 (i,ii,iii) hold. Define*

$$Q_n(a,b,c) = n^{1/2}\{\overline{\mathsf{F}}_n(a,b,c) - \overline{\mathsf{F}}_n(0,0,c)\} - \mathcal{B}_n^{\dagger}(a,b,c) \tag{1.19}$$

*where the bias term, with $\xi(c)$ defined as in 1.11, is*

$$\mathcal{B}_n^{\dagger}(a,b,c) = ac\sigma^{-1}\mathsf{h}(c) + \mathsf{h}(c)\xi'(c)b.$$

*Then, for all $B > 0$ and $0 < \eta < 1/4$*

$$\sup_{|a|,|b|\leq n^{1/4-\eta}B} \sup_{c\in\mathbb{R}} |Q_n(a,b,c)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}).$$

Note the difference between the bias term $\mathcal{B}_n^{\dagger}(a,b,c)$ and $\mathcal{B}_n(a,b,c)$. For the two-sided compensator term, the bias term involves $\xi_c = \xi(c) - \xi(-c)$, while for the one-sided compensator term we consider $\xi(c)$.

### 1.3.3 Other general interest results

We highlight two theorems that have general applicability beyond the asymptotic results presented in this paper. Both theorems are needed due to the dependence structure of the indicator function in the presence of differenced residuals. The first is Theorem 1.3.7 which is the bias correction for differences in probabilities involving two random variables. The second is Theorem 1.3.12 which is an iterative application of the exponential martingale inequality, as it is intended to be used within an application of Theorem 1.3.11.

We start with the derivation of the bias term for the compensators. The compensators requires a more intricate analysis as compared to a standard Taylor approximation in Johansen and Nielsen (2009), Johansen and Nielsen (2016*a*) and Jiao and Nielsen (2017), because the conditioning on the filtration $\mathcal{F}_{i-1}$ leaves both $\nabla\varepsilon_i$ and $x_{i+1,n}$ as random variables.

**Theorem 1.3.7.** *Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}$ be random with density $\mathsf{m}(y,x)$ with respect to the product of the Lebesgue measure and some measure $\upsilon$ on $\mathbb{R}^p$. Suppose the density decomposes into conditional and marginal densities as $\mathsf{m}(y,x) = \mathsf{m}(y|x)\mathsf{k}(x) = \mathsf{m}(x|y)\mathsf{m}_Y(y)$ and that it is differentiable in $y$. Let $\mathsf{E}|NX|^2 < \infty$ for a normalisation matrix $N$ and let $b \in \mathbb{R}^p$. Then*

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

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

$$\dot{\mathsf{m}}(c|x) = \frac{1}{\mathsf{k}(x)} \left. \frac{\partial \mathsf{m}(y,x)}{\partial y} \right|_{y=c}. \tag{1.20}$$

*Moreover, for some $a \in \mathbb{R}$ introduce $c^\dagger = c + ac$ and get*

$$\sup_{c\in\mathbb{R}} |\mathsf{m}_Y(c^\dagger)\mathsf{E}(X'Nb|Y=c^\dagger) - \mathsf{m}_Y(c)\mathsf{E}(X'Nb|Y=c)| \leq |ab| \sup_{y\in\mathbb{R},x\in\mathbb{R}^p} |\dot{\mathsf{m}}(y|x)|\, \mathsf{E}|NX|.$$

To prove Theorem 1.3.1 and 1.3.4 and 1.3.5 we will employ a chaining argument to show uniform convergence. The chaining argument transforms a problem of proving tightness of a class of empirical processes to a problem of bounding tail probabilities for the maximum of a certain family of martingales. We bound the latter tail probabilities using an exponential martingale inequality, the Freedman inequality.

**Theorem 1.3.8.** *(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 bounded and $\mathcal{F}_i$-adapted with $|z_i - \mathsf{E}_{i-1}z_i| \leq 1$. Let $V_n = \mathsf{V}(z_i - \mathsf{E}_{i-1}z_i|\mathcal{F}_{i-1})$ and let $T_n = \sum_{i=1}^n V_n$. For $a, b > 0$ we get,*

$$\mathsf{P}(M_n \geq a, T_n \leq b) \leq \exp\left\{-\frac{a^2}{2(a+b)}\right\}.$$

For our proofs we require to slightly loosen the bound $|z_{i\ell} - \mathsf{E}_{i-1}z_{i\ell}| \leq 1$ to some $d \in \mathbb{N}$ such that $|z_{i\ell} - \mathsf{E}_{i-1}z_{i\ell}| \leq d$. We get the following modified version of the Freedman Inequality.

**Theorem 1.3.9.** *We have the same conditions as in Theorem 1.3.8, except $|z_{i\ell} - \mathsf{E}_{i-1}z_{i\ell}| \leq d$ with $d \in \mathbb{N}$. For $a, b > 0$ we get,*

$$\mathsf{P}(M_n \geq a, T_n \leq b) \leq \exp\left\{-\frac{a^2/d^2}{2(a/d + b/d^2)}\right\} \leq \exp\left\{-\frac{a^2/d^2}{2(a+b)}\right\}.$$

We first consider a general result on a bound of a finite number of martingales. We show this result through an application of the Freedman inequality. Our theorem is related to Theorem 5.1 in Johansen and Nielsen (2016*a*). In contrast to their theorem, we do not need to iterate the martingale inequality, because we consider bounded martingale difference sequences. We also use Freedman's inequality instead of the inequality in Bercu and Touati (2008). Given the bounded martingale difference sequences, this simplifies the inequality that we obtain.

**Theorem 1.3.10.** *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 bounded and $\mathcal{F}_i$-adapted with $|z_{i\ell} - \mathsf{E}_{i-1}z_{i\ell}| \leq d$ and $d \in \mathbb{N}$. Then for all $\kappa_0, \kappa_1 > 0$, we have*

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

The next result follows from Theorem 1.3.10and it modifies Theorem 5.2 in Johansen and Nielsen (2016*a*). In this theorem, we consider the condition for the family of martingales to vanish, when the number of elements in the martingale family is increasing.

**Theorem 1.3.11.** *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 bounded and $\mathcal{F}_i$-adapted with $|z_{i\ell} - \mathsf{E}_{i-1}z_{i\ell}| \leq d$ and $d \in \mathbb{N}$. Suppose $\exists \varsigma \geq 0$, $\lambda > 0$ so that $\mathsf{E}\max_{1\leq\ell\leq L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{li}^2 = \mathrm{O}(n^{\varsigma})$ and $L = \mathrm{O}(n^{\lambda})$. Then, for all $2v > \varsigma$ and $\kappa > 0$ we get*

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

Because we have used Freedman's inequality, we only need to consider one condition $2v > \varsigma$, instead of the two conditions $2v > \varsigma$ and $\varsigma + \lambda < v2^{\bar{r}}$, where $\bar{r} \in \mathbb{N}$ is set such that $\mathsf{E}\max_{1\leq\ell\leq L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{i\ell}^{2\bar{r}} = \mathrm{O}(n^{\varsigma})$.

The next theorem is used to control the terms that arise in the chaining argument of the proof of Theorem 1.3.1 and 1.3.4. It is used when we apply the exponential martingale inequality of Theorem 1.3.10 on the $\mathsf{E}\max_{1\leq\ell\leq L}\sum_{i=1}^{n}\mathsf{E}_{i-1}z_{li}^2$ term in Theorem 1.3.11, in order to move backwards in the filtration.

**Theorem 1.3.12.** *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 bounded and $\mathcal{F}_i$-adapted with $|z_{i\ell} - \mathsf{E}_{i-1}z_{i\ell}| \leq d$ and $d \in \mathbb{N}$. Suppose $\exists \varsigma \geq 0$, $\lambda > 0$ so that $\mathsf{E}\max_{1\leq\ell\leq L}\sum_{l=1}^{L}\mathsf{E}_{i-1}z_{li}^2 = \mathrm{O}(n^{\varsigma})$ and $L = \mathrm{O}(n^{\lambda})$. Then, for all $2v > \varsigma$ we get*

$$\int_0^{\infty}\mathsf{P}\left\{\max_{1\leq\ell\leq L}|M_{\ell n}| > \kappa\right\}d\kappa = \mathrm{O}(n^{v}).$$

The proofs of the three theorems in this section are presented in section 1.4.3.

## 1.4 Proofs

In the section we collect all the proofs for our theorems. The section is structured as follows. In section 1.4.1, we introduce a metric on $c$ which will be used in the chaining argument over $c$. It also states a number of auxiliary theorems for the subsequent proofs. Then, in section 1.4.2, we state the proof of Theorem 2.1, which connects the properties of density $\mathsf{f}$ to $\mathsf{h}$. A key contribution of this chapter are in section 1.4.3, where we provide proofs of new general interest theorems in section 1.4.3. It will include two distinct applications of the exponential martingale inequality by Freedman (1975) which we will frequently use within the chaining arguments. The proofs of the main empirical process results are in section 1.4.4. Then, in section 1.4.5 we collect our proofs that derive the first-order bias term for the compensators. The final proof in section 1.4.6 is for Theorem 1.3.3, which will use most of the results from the previous sections.

### 1.4.1 A auxiliary metric and some inequalities

Introduce notation for the difference of two indicator functions

$$J_i(x, y) = 1_{(\nabla\tilde{\varepsilon}_i \leq y)} - 1_{(\nabla\tilde{\varepsilon}_i \leq x)}, \tag{1.21}$$

where the normalised differenced error terms have density $\mathsf{h}$, see (1.6).

We partition the unit interval into $K$ intervals of equal size. The corresponding quantiles of the distribution function $\mathsf{H}$, which was defined as the distribution function of $\nabla\tilde{\varepsilon}_i$, are $c_k = \mathsf{H}^{-1}(k/K)$ for $0 \leq k \leq K$, such that

$$|\mathsf{E}J_i(c_{k-1}, c_k)| = \mathsf{E}J_i(c_{k-1}, c_k) = \mathsf{H}(c_k) - \mathsf{H}(c_{k-1}) = K^{-1}, \tag{1.22}$$

where we can leave out the absolute values because $c_{k-1} < c_k$.

We state a preliminary inequality from Jiao and Nielsen (2017), which will be used repeatedly in the proofs. The purpose of this theorem is to handle terms involving $c\mathsf{f}(c^*)$, where $c \neq c^*$.

**Theorem 1.4.1.** *(Jiao and Nielsen (2017), Lemma 1.1) If* $|c^* - c| \leq |Ac + B|$ *and* $|A| \leq 1/2$*, then* $|c| \leq 2(|c^*| + |B|)$ *and* $(Ac + B)^2 \leq 16\{A^2(c^*)^2 + B^2\}$.

Let $N_a = n^{-1/2}\sigma^{-1}$. The next theorem is used to the compute uniform covering number of the distance $|\mathsf{H}(c+N_a ac) - \mathsf{H}(c)|$ in terms of grid points $c_k$ of the $c$-axis where $c \in \mathbb{R}$.

**Theorem 1.4.2.** *(Berenguer-Rico et al. (2018), Lemma A.5) Let Assumptions 1.2.2 and 1.2.3(ib) hold. Then a constant $C > 0$ exists so that for all $\eta > 0$*

$$\sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta}} |\mathsf{H}(c + N_a ac) - \mathsf{H}(c)| \leq Cn^{-1/4-\eta}.$$

*In particular, if we choose an index $k(a,c)$ and grip points $c_{k(a,c)-1}, c_{k(a,c)}$ so that $c_{k(a,c)-1} < c + N_a ac \leq c_{k(a,c)}$, then, with $K$ denoting the total number of grid points*

$$\sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta}} |k(a,c) - k(0,c)| \leq n^{1/4-\eta} CK.$$

### 1.4.2 Proof of Theorem 1.2.1

**Proof of Theorem 1.2.1**. *a) Symmetry.* As $\mathsf{h}(x)$ is the density of the difference of two independent, identically distributed random variables $\varepsilon_i$ and $\varepsilon_j$, symmetry is given, as

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

the ordering of $\varepsilon_i$ and $\varepsilon_j$ in the difference does not change the distribution of the difference.
*b) Normal distribution.* If $\mathsf{f}$ is the Gaussian density with mean 0 and variance $\sigma^2$, then $\mathsf{h}$ defined in (1.6) has the same Gaussian density. For the proof of the opposite direction, note first that from part (a) we know that $\mathsf{h} = \mathsf{f}$ is symmetric. Therefore it holds

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

By Pólya (1923), as we assumed a continuous density with bounded second moment, the normal density is the only density satisfying $\mathsf{f} = \mathsf{h}$.
*c) Bounded densities.* The boundedness of $\mathsf{f}$ and the inequality $|x-y|^k \leq C_k(|x|^k + |y|^k)$ for some $C_k > 0$ implies

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

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. This gives the further bound

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

*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 (c) to get

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

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

### 1.4.3 Proofs of general interest results

**Proof of Theorem 1.3.7**. Let $\mathcal{P} = \mathsf{P}(Y - X'Nb \leq c) - \mathsf{P}(Y \leq c)$. Write the probability as an expectation and then expand the expectation to an integral

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

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

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

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

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

The first term is

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

We have by the triangle inequality

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

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

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

For the second part of the theorem, we consider the term

$$\mathcal{Q} = \mathsf{m}_Y(c^\dagger)\mathsf{E}(X'Nb|Y = c^\dagger) - \mathsf{m}_Y(c)\mathsf{E}(X'Nb|Y = c).$$

Apply the mean value theorem to get

$$\mathcal{Q} = (c^\dagger - c) \, \frac{\partial}{\partial y} \int_{-\infty}^{\infty} x'Nb\mathsf{m}(y, x) d\upsilon(x) \bigg|_{y=c^*},$$

where $|c^* - c| \leq |c^\dagger - c| \leq |ac|$. By the Leibniz rule we get uniformly in $c$ that

$$\mathcal{Q} \leq |ac| \int_{-\infty}^{\infty} x'Nb \sup_{y\in\mathbb{R}, x\in\mathbb{R}^p} |\dot{\mathsf{m}}(y|x)| \, \mathsf{k}(x) d\upsilon(x).$$

Reorder the terms to get the desired result. ■

**Proof of Theorem 1.3.10**. Let $m_{i\ell} = z_{i\ell} - \mathsf{E}_{i-1} z_{i\ell}$. Let $A_\ell = \sum_{i=1}^n m_{i\ell}$ and $\mathcal{A} = (\max_{1\leq\ell\leq L} |A_\ell| > \kappa_0)$. Define $B_\ell = \sum_{i=1}^n \mathsf{E}_{i-1} m_{i\ell}^2$ and $\mathcal{B} = (\max_{1\leq\ell\leq L} B_l \leq \kappa_1)$. Then we bound

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

*Bounding* $\mathsf{P}(\mathcal{A} \cap \mathcal{B})$. Let $\mathcal{A}_\ell = (|A_\ell| > \kappa_0)$ 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 Theorem 1.3.9 with $|m_{i\ell}| \leq d$ to get

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

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

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

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

**Proof of Theorem 1.3.11**. For some $\kappa > 0$, apply Theorem 1.3.10 with $\kappa_0 = \kappa n^v$ and $\kappa_1 = \kappa^2 n^{2v}\{8\lambda \log n\}^{-1}$. Let $n$ be fixed and sufficiently large such that $\kappa_0 < \kappa_1$. Use the bound

$$\exp\left\{-\frac{\kappa_0^2/d^2}{2(\kappa_1+\kappa_0)}\right\} \leq \exp\left\{-\frac{\kappa_0^2/d^2}{4\kappa_1}\right\} = n^{-2\lambda}, \tag{1.23}$$

For our probability of interest, we get that

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

where we used $2v > \varsigma \geq 0$. ■

**Proof of Theorem 1.3.12**. Let $W = \max_{1\leq l\leq L} |\sum_{i=1}^n (z_{li} - \mathsf{E}_{i-1} z_{li})|$. Our object of interest is then

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

Use the substitution formula with $\kappa = n^v \tilde{\kappa}_0^2$ so $d\kappa = 2n^v \tilde{\kappa}_0 d\tilde{\kappa}_0$ to get $\mathcal{E}_W \leq n^v + 2n^v \mathcal{J}$, where $\mathcal{J} = \int_1^\infty \tilde{\kappa}_0 \mathsf{P}\left(W > \tilde{\kappa}_0^2 n^v\right) d\tilde{\kappa}_0$. Apply Theorem 1.3.10 to get $\mathcal{J} \leq \mathcal{J}_1 + \mathcal{J}_2$, where

$$\mathcal{J}_1 = \int_1^\infty \frac{\tilde{\kappa}_0}{\kappa_1} \mathsf{E} \max_{1\leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{li}^2 d\tilde{\kappa}_0, \quad \mathcal{J}_2 = 2L \int_1^\infty \tilde{\kappa}_0 \exp\left\{-\frac{\kappa_0^2/d^2}{2(\kappa_1+\kappa_0)}\right\} d\tilde{\kappa}_0,$$

with $\kappa_0 = \tilde{\kappa}_0^2 n^v$ and where we choose $\kappa_1 = \tilde{\kappa}_0^4 n^{2v}\{8d^2\lambda \log n + 12d^2 \log \tilde{\kappa}_0\}^{-1}$. We show that $\mathcal{J}_1$ and $\mathcal{J}_2$ vanish with $n \to \infty$.

For $\mathcal{J}_1$, write out $\kappa_0$ and $\kappa_1$ to get

$$\mathcal{J}_1 = \frac{\mathsf{E} \max_{1\leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{li}^2}{n^{2v}} \int_1^\infty \frac{8d^2\lambda \log n + 12d^2 \log \tilde{\kappa}_0}{\tilde{\kappa}_0^3} d\tilde{\kappa}_0.$$

The integrand is a sum, so write the integral as a sum of two integrals such that $\mathcal{J}_1 \leq \mathcal{J}_1^\dagger + \mathcal{J}_1^\ddagger$ and use $\mathsf{E} \max_{1\leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{li}^2 = \mathrm{O}(n^\varsigma)$ to get

$$\mathcal{J}_1^\dagger = \mathrm{O}\left( \frac{n^\varsigma \log n}{n^{2v}} \int_1^\infty \frac{1}{\tilde{\kappa}_0^3} d\tilde{\kappa}_0 \right), \quad \mathcal{J}_1^\ddagger = \mathrm{O}\left( \frac{n^\varsigma}{n^{2v}} \int_1^\infty \frac{\log \tilde{\kappa}_0}{\tilde{\kappa}_0^3} d\tilde{\kappa}_0 \right).$$

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

For $\mathcal{J}_2$, use the bound in (1.23) and by our choice of $\kappa_0$ and $\kappa_1$ we get

$$\mathcal{J}_2 \leq 2L \int_1^\infty \tilde{\kappa}_0 \exp\left\{ -\frac{\kappa_0^2}{4\kappa_1} \right\} d\tilde{\kappa}_0 = 2L \int_1^\infty n^{-2\lambda} \tilde{\kappa}_0^{-2} d\tilde{\kappa}_0.$$

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

### 1.4.4 Proofs of the empirical process results

**Proof of Theorem 1.3.4**. Let $N_a = n^{-1/2}\sigma^{-1}$. We take the variation in parameter $a$ and merge it with the variation over the parameter $c$, such that $\mathbb{F}_n(a, 0, c) = \mathbb{F}_n(0, 0, c + N_a ac)$. Introduce $M_n$ as an alternative way to write $R_n$ as

$$\begin{aligned} M_n(c, c + N_a ac) = R_n(a, c) &= \mathbb{F}_n(0, 0, c + N_a ac) - \mathbb{F}_n(0, 0, c) \\ &= n^{-1/2} \sum_{i=1}^n \{ J_i(c, c + N_a ac) - \mathsf{E}_{i-1} J_i(c, c + N_a ac) \}, \end{aligned}$$

where $J_i$ is defined in (1.21). We show $\mathcal{M}_n = \sup_{c\in\mathbb{R}} \sup_{|a|\leq n^{1/4-\eta_B}} M_n(c, c + N_a ac)$ to vanish for large $n$.

**1. The chaining setup.** We will employ a chaining argument where we chain over $c$. While we consider uniform convergence over $a, c$, we chain only over $c$ because we will show that the uniform convergence over $a, c$ can be rewritten as a uniform convergence over $c$ only. We will later, in point 2.2, choose a small $\delta > 0$ which will be used to form the covering balls for $c$ in point 1.1. It is important to note that $\delta > 0$ will be chosen independently of the sample size $n$.

1.1 Construct the c-balls. Choose a $\delta > 0$ mentioned in point 1. For $\delta, n > 0$ partition the $c$-axis as outlined in (1.22). The construction of the grid points $c_k$

makes $\mathsf{H}(c_k)-\mathsf{H}(c_{k-1})$ equidistant, not the distance $c_k-c_{k-1}$. We set the number of grid points to $K=\text{int}(n^{1/2}/\delta)$, where int refers to the integer component discarding the decimals. As $\mathsf{H}(\infty)=1$, the distance $\mathsf{H}(c_k)-\mathsf{H}(c_{k-1})$ is fixed at $1/K\sim n^{-1/2}\delta$. For each $c>-\infty$ there exists $c_{k-1}$ and $c_k$ so that $c_{k-1}<c\leq c_k$. Let $c_{k(c)}$ and $c_{k(a,c)}$ denote the nearest right grid points to $c$ and $c+N_aac$, respectively.

1.2. Apply chaining. Add and subtract $1_{(\nabla\tilde{\varepsilon}_i\leq c_{k(a,c)})}$ and $1_{(\nabla\tilde{\varepsilon}_i\leq c_{k(c)})}$ so that

$$J_i(c,c+N_aac)=1_{(\nabla\tilde{\varepsilon}_i\leq c+N_aac)}-1_{(\nabla\tilde{\varepsilon}_i\leq c_{k(a,c)})}\\ +1_{(\nabla\tilde{\varepsilon}_i\leq c_{k(a,c)})}-1_{(\nabla\tilde{\varepsilon}_i\leq c_{k(c)})}+1_{(\nabla\tilde{\varepsilon}_i\leq c_{k(c)})}-1_{(\nabla\tilde{\varepsilon}_i\leq c)}.$$

By the triangle inequality we then get

$$|M_n(c,c+N_aac)|=|M_n(c,c_{k(c)})|+|M_n(c_{k(c)},c_{k(a,c)})|+|M_n(c_{k(a,c)},c+N_aac)|.$$

Because $|M_n(c,c_{k(c)})|$ is a special case where $a=0$ in $|M_n(c_{k(a,c)},c+N_aac)|$, it suffices show that the oscillation term $|M_n(c_{k(a,c)},c+N_aac)|$ and the discrete point term $|M_n(c_{k(c)},c_{k(a,c)})|$ vanish uniformly in $a,c$. We consider them in turn.

**2. The oscillation term $M_n(c_{k(a,c)},c+N_aac)$.** As $c_{k(a,c)}$ is the nearest right grid points to $c+N_aac$ we can write the uniform convergence of the oscillation term over $a,c$ as the uniform convergence over the grid points $k$ as follows

$$\sup_{c\in\mathbb{R}}\sup_{|a|\leq n^{1/4-\eta}B}|M_n(c_{k(a,c)},c+N_aac)|\leq\max_{1\leq k\leq K}\sup_{c:c_{k-1}\leq c\leq c_k}|M_n(c,c_k)|.$$

Since $c_{k-1}<c\leq c_k$ we bound $J_i(c,c_k)\leq z_{ik}=J_i(c_{k-1},c_k)$. Thus, $|M_n(c,c_k)|\leq n^{-1/2}\sum_{i=1}^n(z_{ik}+\mathsf{E}_{i-1}z_{ik})$ uniformly in $c$.

2.1 The martingale decomposition. Add and subtract $\mathsf{E}_iJ_i(c_{k-1},c_k)$ and note

$$|M_n(c,c_k)|\leq M_n^a(c_{k-1},c_k)+M_n^b(c_{k-1},c_k)+2\overline{M}_n(c_{k-1},c_k),$$

where

$$M_n^a(c_{k-1},c_k)=n^{-1/2}\sum_{i=1}^n(z_{ik}-\mathsf{E}_iz_{ik}),\ M_n^b(c_{k-1},c_k)=n^{-1/2}\sum_{i=1}^n(\mathsf{E}_iz_{ik}-\mathsf{E}_{i-1}z_{ik}),$$
$$\overline{M}_n(c_{k-1},c_k)=n^{-1/2}\sum_{i=1}^n\mathsf{E}_{i-1}z_{ik}.$$

Here, $M_n^1$ and $M_n^2$ are $\mathcal{F}_{i+1}$ and $\mathcal{F}_i$-adapted martingales. It suffices to show, for $j = a, b$

$$\mathcal{M}_n^j = \max_{1\le k\le K} |M_n^j(c_{k-1}, c_k)| = \mathrm{o_P}(1), \qquad \overline{\mathcal{M}}_n = \max_{1\le k\le K} |\overline{M}_n(c_{k-1}, c_k)| = \mathrm{o_P}(1).$$

2.2. The compensator term $\overline{\mathcal{M}}_n$. We show that the compensator term is $\overline{\mathcal{M}}_n = \mathrm{o}(1)$. By the partitioning of the $c$-axis, we have with $K = \mathrm{int}(n^{1/2}/\delta)$ and uniformly in $k$,

$$\mathsf{E}_{i-1} z_{ik} \overset{a.s.}{=} \mathsf{E} J_i(c_{k-1}, c_k) = K^{-1} \le \delta n^{-1/2}, \tag{1.24}$$

Here we used that $\nabla\varepsilon_i$ is independent from $\mathcal{F}_{i-1}$. Therefore, $\overline{\mathcal{M}}_n \le \delta$ almost surely. As $\delta$ can be chosen arbitrarily small, we have $\overline{\mathcal{M}}_n = \mathrm{o}(1)$ almost surely.

2.3. The martingale term $\mathcal{M}_n^a$. We show that $\mathcal{M}_n^a = \mathrm{o_P}(1)$, by applying Theorem 1.3.11 to the family of martingales. We set $v_a = 1/2$ and index $\ell = k$ and consider

$$z_{i\ell,a} = z_{ik} = J_i(c_{k-1}, c_k)$$

which is $\mathcal{F}_{i+1}$-adapted. Note that $|z_{i\ell,a} - \mathsf{E}_i z_{i\ell,a}| \le 1$ as $0 \le J_i(c_{k-1}, c_k) \le 1$ and $0 \le \mathsf{E}_i J_i(c_{k-1}, c_k) \le 1$. To verify the conditions of Theorem 1.3.11, we show that we have $\lambda_a = 1/2$ and $\varsigma_a = 1/2$.

*The parameter* $\lambda_a$. The set of indices $\ell$ has size $L_a = K \sim n^{1/2}$, hence $\lambda = 1/2$.

*The parameter* $\varsigma_a$. We show that $\mathcal{E}_a = \mathsf{E}\max_{1\le\ell\le L}\sum_{i=1}^n \mathsf{E}_i z_{i\ell,a}^2 = \mathrm{O}(n^{1/2})$. Because $J_i$ is an indicator function and $c_{k-1} < c_k$, we have $J_i^2 = J_i$. It thus suffices to consider $\mathsf{E}_i z_{i\ell,a}$. Add and subtract $\mathsf{E}_{i-1} z_{i\ell,a}$ and use the triangle inequality to get

$$\mathcal{E}_a \le \mathsf{E}\max_{1\le\ell\le L}\left|\sum_{i=1}^n (\mathsf{E}_i z_{i\ell,a} - \mathsf{E}_{i-1} z_{i\ell,a})\right| + \mathsf{E}\max_{1\le\ell\le L}\left|\sum_{i=1}^n \mathsf{E}_{i-1} z_{i\ell,a}\right|. \tag{1.25}$$

As $z_{i\ell,a} = z_{ik}$, for the second term in (1.25) we use the bound in (1.24), noting that $\delta$ is fixed and we only require the rate, to get that it is of order $\mathrm{O}(n^{1/2})$. For the first term in (1.25), apply $\mathsf{E}[X] = \int_0^\infty \{1 - F(x)\}dx$, which holds for any non-negative random variable $X$ and its distribution function $F(x)$, to get

$$\mathsf{E}\max_{1\le\ell\le L}\left|\sum_{i=1}^n (\mathsf{E}_i z_{i\ell,a} - \mathsf{E}_{i-1} z_{i\ell,a})\right| \le \int_0^\infty \mathsf{P}\left\{\max_{1\le\ell\le L}\left|\sum_{i=1}^n (\mathsf{E}_i z_{i\ell,a} - \mathsf{E}_{i-1} z_{i\ell,a})\right| > \kappa\right\} d\kappa.$$

Apply Theorem 1.3.12 with $v'_a = 1/2$ and parameters $\lambda'_a$ and $\varsigma'_a$.

*The parameter* $\lambda'_a$. Given the same set-up, $\lambda'_a = \lambda_a$.

*The parameter* $\varsigma'_a$. We show that $\mathcal{E}'_a = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} \mathsf{E}_i^2 z_{i\ell,a} = \mathrm{O}(n^{1/2})$. Use Jensen's inequality and the law of iterated expectations to show $\mathsf{E}_{i-1} \mathsf{E}_i^2 z_{i\ell,a} \leq \mathsf{E}_{i-1} \mathsf{E}_i z_{i\ell,a}^2 = \mathsf{E}_{i-1} z_{i\ell,a}^2$. Because $J_i$ is an indicator function and $c_{k-1} < c_k$, we have $J_i^2 = J_i$ and thus it suffices to consider $\mathsf{E}_{i-1} z_{i\ell,a}$. As $z_{i\ell,a} = z_{ik}$, we use the bound in (1.24) to get that $\mathcal{E}'_a$ is of order $\mathrm{O}(n^{1/2})$.

*The condition* $\varsigma'_a < 2v'_a$. It holds since $\varsigma'_a = 1/2 < 1 = 2v'_a$ since $v'_a = 1/2$.

As both terms in (1.25) are of order $\mathrm{O}(n^{1/2})$ we have that $\mathcal{E}_a = \mathrm{O}(n^{\varsigma_a})$ where $\varsigma_a = 1/2$.

*The condition* $\varsigma_a < 2v_a$. It holds since $\varsigma_a = 1/2 < 1 = 2v_a$ since $v_a = 1/2$.

2.4. The martingale term $\mathcal{M}_n^b$. We show that $\mathcal{M}_n^b = \mathrm{o}_{\mathsf{P}}(1)$ by applying Theorem 1.3.11 to it. We set $v_b = 1/2$ and index $\ell = k$ and consider

$$z_{i\ell,b} = \mathsf{E}_i z_{i\ell,a} = \mathsf{E}_i J_i(c_{k-1}, c_k),$$

which is $\mathcal{F}_i$-adapted. Note that $|\mathsf{E}_i z_{i\ell,b} - \mathsf{E}_{i-1} z_{i\ell,b}| \leq 1$ as $0 \leq \mathsf{E}_{i-1} J_i(c_{k-1}, c_k) \leq 1$ and $0 \leq \mathsf{E}_i J_i(c_{k-1}, c_k) \leq 1$. We show that the parameters are $\lambda_b = 1/2$ and $\varsigma_b = 1/2$. We verify the conditions of Theorem 1.3.11.

*The parameter* $\lambda_b$. We have $\lambda_b = 1/2$, as the set-up is identical to point 2.2.

*The parameter* $\varsigma_b$. We show that $\mathcal{E}_b = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{i\ell,b}^2 = \mathrm{O}(n^{1/2})$. Note that $z_{i\ell,b}^2 = \mathsf{E}_i^2 z_{i\ell,a}$ Use Jensen's inequality and the law of iterated expectations to show $\mathsf{E}_{i-1} \mathsf{E}_i^2 z_{i\ell,a} \leq \mathsf{E}_{i-1} \mathsf{E}_i z_{i\ell,a}^2 = \mathsf{E}_{i-1} z_{i\ell,a}^2$. Because $J_i$ is an indicator function, we have $J_i^2 = J_i$ and thus it suffices to consider $\mathsf{E}_{i-1} z_{i\ell,a}$. As in point 2.3, because $z_{i\ell,a} = z_{ik}$, we use the bound in (1.24) to get that $\mathcal{E}_b = \mathrm{O}(n^{\varsigma_b})$ where $\varsigma_b = 1/2$.

*The condition* $\varsigma_b < 2v_b$. It holds since $\varsigma_b = 1/2 < 1 = 2v_b$ since $v_b = 1/2$.

**3. The discrete point term** $M_n(c_{k(c)}, c_{k(a,c)})$**.** By Assumption 1.2.2 and 1.2.3(ib) we apply Theorem 1.4.2 with $K \sim n^{1/2}$ to get the uniform covering number. That is there exists a constant $C > 0$ so that

$$\sup_{c \in \mathbb{R}} \sup_{a \leq n^{1/4-\eta} B} |k(a,c) - k(c)| \leq D_n = C n^{1/4-\eta}.$$

As a consequence, for generic grid points $c_{k^*}$, $c_k$, so that $k^* : k \leq k^* \leq k + D_n$, we can bound,

$$|M_n(c_{k(c)}, c_{k(a,c)})| \leq \widetilde{\mathcal{M}}_n = \max_{1 \leq k \leq K} \max_{k^* : k \leq k^* \leq k + D_n} |M_n(c_k, c_{k^*})|,$$

which is uniform in $a$ and $c$. It suffices to show that $\widetilde{\mathcal{M}}_n = \mathsf{o}_{\mathsf{P}}(1)$.

3.1. A martingale decomposition. We need to show $M_n = \mathsf{o}_{\mathsf{P}}(1)$ uniformly in $k, k^*$. To do so, we decompose it into two martingales. Add and subtract $n^{-1/2} \sum_{i=1}^n \mathsf{E}_i J_i(c_k, c_{k^*})$ to $M_n$ to get $M_n = \widetilde{M}_n^c + \widetilde{M}_n^d$, where

$$\widetilde{M}_n^c = n^{-1/2} \sum_{i=1}^n \{J_i(c_k, c_{k^*}) - \mathsf{E}_i J_i(c_k, c_{k^*})\},$$
$$\widetilde{M}_n^d = n^{-1/2} \sum_{i=1}^n \{\mathsf{E}_i J_i(c_k, c_{k^*}) - \mathsf{E}_{i-1} J_i(c_k, c_{k^*})\}.$$

Thus, it suffices to show that

$$\widetilde{\mathcal{M}}_n^c = \max_{1 \leq k \leq K} \max_{k^* : k \leq k^* \leq k + D_n} \widetilde{M}_n^c = \mathsf{o}_{\mathsf{P}}(1), \quad \widetilde{\mathcal{M}}_n^d = \max_{1 \leq k \leq K} \max_{k^* : k \leq k^* \leq k + D_n} \widetilde{M}_n^d = \mathsf{o}_{\mathsf{P}}(1).$$

3.2. The martingale $\widetilde{\mathcal{M}}_n^d$. We show that $\widetilde{\mathcal{M}}_n^d = \mathsf{o}_{\mathsf{P}}(1)$ by applying Theorem 1.3.11 to it. We set $v_d = 1/2$ and $\ell = \{k, k^*\}$ and consider $z_{i\ell,d} = \mathsf{E}_i J_i(c_k, c_{k^*})$, which is $\mathcal{F}_i$-adapted. Note that $|\mathsf{E}_i z_{i\ell,d} - \mathsf{E}_{i-1} z_{i\ell,d}| \leq 1$ as $0 \leq \mathsf{E}_{i-1} J_i(c_k, c_{k^*}) \leq 1$ and $0 \leq \mathsf{E}_i J_i(c_k, c_{k^*}) \leq 1$. We show that the parameters are $\lambda_d = \varsigma_d = 3/4 - \eta$. We verify the conditions of Theorem 1.3.11.

*The parameter $\lambda_d$.* The set of indices $\ell$ has size $L = KD_n \sim n^{3/4-\eta}$. Therefore $\lambda_d = 3/4 - \eta$.

*The parameter $\varsigma_d$.* We show that $\mathcal{E}_d = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{i\ell,d}^2 = \mathrm{O}(n^{3/4-\eta})$. Note that $z_{i\ell,d}^2 = \mathsf{E}_i^2 J_i(c_k, c_{k^*})$ Use Jensen's inequality and the law of iterated expectations to show $\mathsf{E}_{i-1} \mathsf{E}_i^2 J_i(c_k, c_{k^*}) \leq \mathsf{E}_{i-1} \mathsf{E}_i J_i^2(c_k, c_{k^*}) = \mathsf{E}_{i-1} J_i^2(c_k, c_{k^*})$. Because $J_i$ is an indicator function and $c_k \leq c_{k^*}$, we have $J_i^2 = J_i$ and thus it suffices to consider $\mathsf{E}_{i-1} J_i(c_k, c_{k^*})$. By construction since $k^* : k \leq k^* \leq k + D_n$ (see also point 3), we have

$$\mathsf{E}_{i-1} J_i(c_k, c_{k^*}) \leq \mathsf{H}(c_{k^*}) - \mathsf{H}(c_k) \leq D_n/K, \tag{1.26}$$

where $D_n \sim n^{1/4-\eta}$. The terms $D_n$ and $K$ are deterministic and do not depend on $i$, $k$ and $k^*$, therefore noting $D_n/K \sim n^{-1/4-\eta}$, we get with $\varsigma = 3/4 - \eta$

$$\mathsf{E} \max_{1\leq k\leq K} \max_{k^*:k\leq k^*\leq k+D_n} \sum_{i=1}^{n} \mathsf{E}_{i-1} J_i(c_k, c_{k^*}) \leq nD_n/K = \mathrm{O}(n^{\varsigma}).$$

*The condition* $\varsigma_d < 2v_d$. It holds since $\varsigma_d = 3/4 - \eta < 1 = 2v_d$ since $v_d = 1/2$.

3.3. The martingale $\widetilde{\mathcal{M}}_n^c$. We show that $\widetilde{\mathcal{M}}_n^c = \mathrm{o}_{\mathsf{P}}(1)$ by applying Theorem 1.3.11 to it. We set $v_c = 1/2$ and $\ell = \{k, k^*\}$ and consider $z_{i\ell,c} = J_i(c_k, c_{k^*})$, which is $\mathcal{F}_{i+1}$-adapted. Note that $|z_{i\ell,c} - \mathsf{E}_i z_{i\ell,c}| \leq 1$ as $0 \leq J_i(c_k, c_{k^*}) \leq 1$ and $0 \leq \mathsf{E}_i J_i(c_k, c_{k^*}) \leq 1$. We show that the parameters are $\lambda_c = 3/4 - \eta$ and $\varsigma_c = 3/4 - \eta$. We verify the conditions of Theorem 1.3.11.

*The parameter* $\lambda_c$. We have $\lambda_c = 3/4 - \eta$, as the set-up is identical to point 3.2.

*The parameter* $\varsigma_c$. We show that $\mathcal{E}_c = \mathsf{E} \max_{1\leq\ell\leq L} \sum_{i=1}^{n} \mathsf{E}_i z_{i\ell,c}^2 = \mathrm{O}(n^{3/4-\eta})$. Because $J_i$ is an indicator function and $c_k \leq c_{k^*}$, we have $J_i^2 = J_i$ and thus it suffices to consider $\mathsf{E}_i z_{i\ell,c}$. Add and subtract $\mathsf{E}_{i-1} z_{i\ell,c}$ and use the triangle inequality to get

$$\mathcal{E}_c \leq \mathsf{E} \max_{1\leq\ell\leq L} \left| \sum_{i=1}^{n} (\mathsf{E}_i z_{i\ell,c} - \mathsf{E}_{i-1} z_{i\ell,c}) \right| + \mathsf{E} \max_{1\leq\ell\leq L} \left| \sum_{i=1}^{n} \mathsf{E}_{i-1} z_{i\ell,c} \right|. \tag{1.27}$$

As $z_{i\ell,c} = J_i(c_k, c_{k^*})$, for the second term in (1.27) we use the bound in (1.26) to get that it is of order $\mathrm{O}(n^{3/4-\eta})$. For the first term in (1.27), use $\mathsf{E}[X] = \int_0^\infty \{1 - F(x)\} dx$ on a non-negative random variable $X$, to get

$$\mathsf{E} \max_{1\leq\ell\leq L} \left| \sum_{i=1}^{n} \mathsf{E}_i z_{i\ell,c} - \mathsf{E}_{i-1} z_{i\ell,c} \right| \leq \int_0^\infty \mathsf{P} \left\{ \max_{1\leq\ell\leq L} \left| \sum_{i=1}^{n} \mathsf{E}_i z_{i\ell,c} - \mathsf{E}_{i-1} z_{i\ell,c} \right| > \kappa \right\} d\kappa.$$

Apply Theorem 1.3.12 with the parameters $\lambda_c'$, $\varsigma_c' = 3/4 - \eta$ and $v_c' = 1/2$.

*The parameter* $\lambda_c'$. Given the same set-up, $\lambda_c' = \lambda_c$.

*The parameter* $\varsigma_c'$. We show that $\mathcal{E}_c' = \mathsf{E} \max_{1\leq\ell\leq L} \sum_{i=1}^{n} \mathsf{E}_{i-1} \mathsf{E}_i^2 z_{i\ell,c} = \mathrm{O}(n^{3/4-\eta})$. Use Jensen's inequality and the law of iterated expectations to show $\mathsf{E}_{i-1} \mathsf{E}_i^2 z_{i\ell,c} \leq \mathsf{E}_{i-1} \mathsf{E}_i z_{i\ell,c}^2 = \mathsf{E}_{i-1} z_{i\ell,c}^2$. Because $J_i$ is an indicator function and $c_k \leq c_{k^*}$, we have $J_i^2 = J_i$ and thus it suffices to consider $\mathsf{E}_{i-1} z_{i\ell,c}$. As $z_{i\ell,c} = J_i(c_k, c_{k^*})$, we use the bound in (1.26) to get that it is of order $\mathrm{O}(n^{3/4-\eta})$.

*The condition* $\varsigma_c' < 2v_c'$. It holds since $\varsigma_c' = 3/4 - \eta < 1 = 2v_c'$ since $v_c' = 1/2$. As both terms in (1.25) are of order $\mathrm{O}(n^{3/4-\eta})$ we have that $\mathcal{E}_c = \mathrm{O}(n^{\varsigma_c})$ where $\varsigma_c = 3/4 - \eta$.

*The condition* $\varsigma_c < 2v_c$. It holds since $\varsigma_c = 3/4 - \eta < 1 = 2v_c$ since $v_c = 1/2$.

**4. Conclusion.** In point 3 we have shown that the two terms $\widetilde{\mathcal{M}}_n^c$ and $\widetilde{\mathcal{M}}_n^d$ vanish uniformly in $k, k^*$. As $\sup_{c\in\mathbb{R}} \sup_{a\leq n^{1/4-\eta_B}} |M_n(c_{k(c)}, c_{k(a,c)})| \leq \widetilde{\mathcal{M}}_n^c + \widetilde{\mathcal{M}}_n^d$ this implies that the discrete point term $M_n(c_{k(c)}, c_{k(a,c)})$ vanishes in $a, c$. In point 2, we have considered the oscillation term $M_n(c_a, c_{k(a,c)})$. This term vanished uniformly in $a, c$ as $\widetilde{\mathcal{M}}_n^a$ and $\widetilde{\mathcal{M}}_n^b$ vanish uniformly in $k$ and because we can choose $\delta$ small such that the compensator term $\overline{\mathcal{M}}_n$ vanishes. Therefore, by the chaining argument over $c$, we have that $R_n(a, c) = \mathrm{o_P}(1)$. ■

**Proof of Theorem 1.3.5**. Combine the two estimation errors into one vector and create a corresponding expanded vector of regressors as

$$u = (a, b')' \quad \text{and} \quad \nabla x_{inu} = (\sqrt{2}n^{-1/2}c, \nabla x_{in}')'.$$

As with other variables we denote $\nabla\tilde{x}_{inu} = 2^{-1/2}\sigma^{-1}\nabla x_{inu}$ and $\nabla\check{x}_{inu} = \sigma^{-1}\nabla x_{inu}$. Recall the definition of $\mathbb{F}_n$ from (1.18) and write our object of interest as

$$\begin{aligned} R_n(u, c) &= \mathbb{F}_n(a, b, c) - \mathbb{F}_n(0, 0, c) \\ &= n^{-1/2} \sum_{i=1}^{n} \{J_i(c, c + \nabla\tilde{x}_{inu}'u) - \mathsf{E}_{i-1} J_i(c, c + \nabla\tilde{x}_{inu}'u)\}, \end{aligned}$$

where $J_i$ is defined as in (1.21). We show $\mathcal{R}_n = \sup_{|u|\leq n^{1/4-\eta_B}} |R_n(u, c)| = \mathrm{o_P}(1)$.

This proof is structured into three parts. The first part implements the chaining over $u$, in which we place grid points $u_m$ in regular intervals within $u$. In part two, we show that our empirical process vanishes on the grid points $u_m$

$$\mathcal{R}_{n,1} = \max_{1\leq m\leq M} |R_n(u_m)| = \mathrm{o_P}(1). \tag{1.28}$$

Then, in the final third part, we show that our empirical process vanishes on the points in-between our grid points $u_m$

$$\mathcal{R}_{n,2} = \max_{1\leq m\leq M} \sup_{|u-u_m|\leq\delta} |R_n(u) - R_n(u_m)|. \tag{1.29}$$

**1. The chaining setup.** In this section, we chain over $u$, by covering them with balls of radius $\delta > 0$. We will choose $\delta$ independently of the sample size $n$. We think of $\delta$ as a small parameter which we set to control our process in point 3.5 below.

1.1. Construct the u-balls. Consider the $\delta$ from in point 1. For $\delta, n > 0$, cover the set $|u| \leq n^{1/4-\eta}B$ with balls of radius $\delta$ which centre around grid points $u_m$. The number of balls necessary to cover $u$ is $M \sim (n^{1/4-\eta}B/\delta)^{\dim x+1} \sim n^{(1/4-\eta)\dim x}/\delta^{\dim x+1}$. Thus, for any $u$ there exists a $u_m$ so that $|u - u_m| \leq \delta$.

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 $\mathcal{R}_{n,1}$ and $\mathcal{R}_{n,1}$ are defined in (1.28) and (1.29).

**2. The discrete point term $\mathcal{R}_{n,1}$.** In this section we show that $\mathcal{R}_{n,1}$ is $o_{\mathsf{P}}(1)$. The proof relies on the iterated martingale inequality in Theorem 1.3.11. Because $R_n$ is not a martingale with $J_i$ being $\mathcal{F}_{i+1}$ adapted, in the first step we need a martingale decomposition of $R_n$. Then, we apply Theorem 1.3.11 on the martingales constructed in the previous point, by showing that its conditions are satisfied.

2.1. Martingale decomposition. Let $z_{im} = J_i(c, c + \nabla\tilde{x}'_{inu}u_m)$ and define the martingales $R^a_n$ and $R^b_n$ as

$$R^a_n(u_m) = n^{-1/2}\sum_{i=1}^{n}(z_{im} - \mathsf{E}_i z_{im}), \quad R^b_n(u_m) = n^{-1/2}\sum_{i=1}^{n}(\mathsf{E}_i z_{im} - \mathsf{E}_{i-1}z_{im})$$

so that $R_n = R^a_n + R^b_n$. Thus, it suffices to show that

$$\mathcal{R}^a_{n,1} = \max_{1\leq m\leq M}|R^a_n(u_m)| = o_{\mathsf{P}}(1), \quad \mathcal{R}^b_{n,1} = \max_{1\leq m\leq M}|R^b_n(u_m)| = o_{\mathsf{P}}(1).$$

2.2. The martingale $\mathcal{R}^a_{n,1}$. We show that $\mathcal{R}^a_{n,1} = o_{\mathsf{P}}(1)$, by applying Theorem 1.3.11 to it. We set $v_a = 1/2$ and index $\ell = m$ and consider

$$z_{i\ell,a} = z_{im} = J_i(c, c + \nabla\tilde{x}'_{inu}u_m)$$

which is $\mathcal{F}_{i+1}$-adapted. Note that $|\mathsf{E}_i z_{i\ell,a} - \mathsf{E}_{i-1}z_{i\ell,a}| \leq 2$ as $|J_i| \leq 1$. We show that the parameters are $\lambda_a = (1/4-\eta)(\dim x + 1)$ and $\varsigma_a = 3/4 - \eta$. We verify the conditions of Theorem 1.3.11.

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

*The parameter $\varsigma_a$.* We show that $\mathcal{E}_a = \mathsf{E} \max_{1\leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_i z_{i\ell,a}^2 = \mathrm{O}(n^{3/4-\eta})$. We have that $\mathsf{E}_i z_{i\ell,a}^2 = \mathsf{E}_i |z_{i\ell,a}|$. Note that

$$\mathsf{E}_i |z_{i\ell,a}| = |\mathsf{P}(\check{\varepsilon}_{i+1} \leq \check{\varepsilon}_i - \sqrt{2}c | \mathcal{F}_i) - \mathsf{P}(\check{\varepsilon}_{i+1} \leq \check{\varepsilon}_i - \sqrt{2}c - \nabla \check{x}'_{inu} u_m | \mathcal{F}_i)|$$

where $z_{i\ell,a}$ contains four random variables that are adapted to three different filtrations: we have $\mathcal{F}_{i-1}$-adapted $x_i$; $\mathcal{F}_i$-adapted $\varepsilon_i$ and $x_{i+1}$; and $\mathcal{F}_{i+1}$-adapted $\varepsilon_{i+1}$. Given $\mathcal{F}_i$, only $\varepsilon_{i+1}$ remains random. Because only $\varepsilon_{i+1}$ is independent of $\mathcal{F}_i$, we can write the probability as an integral over the density $\mathsf{f}$ of the error term $\check{\varepsilon}_{i+1}$ as

$$\mathsf{E}_i z_{i\ell,a} = \int_{\check{\varepsilon}_i - \sqrt{2}c - \nabla \check{x}'_{inu} u_m}^{\check{\varepsilon}_i - \sqrt{2}c} \mathsf{f}(v) dv.$$

apply the Taylor expansion around $\check{\varepsilon}_i - \sqrt{2}c$ to get $\mathsf{E}_i |z_{i\ell,a}| \leq |\nabla \check{x}_{inu}||u_m| \mathsf{f}(v^*)$ where $|v^* - \check{\varepsilon}_i + \sqrt{2}c| \leq |\check{x}'_{inu} u_m|$. Use $\mathsf{f}(v^*) < \sup_{v\in\mathbb{R}} \mathsf{f}(v)$ and $|u_m| \leq n^{1/4-\eta} B$ to bound $\mathsf{E}_i |z_{i\ell,a}|$ uniformly in $m$ as

$$\mathsf{E}_i |z_{i\ell,a}| \leq n^{1/4-\eta} B |\nabla \check{x}_{inu}| \sup_{v\in\mathbb{R}} \mathsf{f}(v). \tag{1.30}$$

We get

$$\mathcal{E}_a \leq \mathrm{O}(n^{1/4-\eta}) \sup_{v\in\mathbb{R}} \mathsf{f}(v) \mathsf{E} \sum_{i=1}^n \mathsf{E}_i |\nabla x_{inu}|.$$

Apply the law of iterated expectations to get $\mathsf{E} \sum_{i=1}^n \mathsf{E}_i |\nabla x_{inu}| = \mathsf{E} \sum_{i=1}^n |\nabla x_{inu}|$. We consider the two components of $\nabla x_{inu}$ in turn. For the regular regressors $\nabla x_{in}$ we appeal to Assumption 1.2.3(iii) such that $\mathsf{E} \sum_{i=1}^n |\nabla x_{in}| = \mathrm{O}(n^{1/2})$. For the term related to $a$ we have for $c \in \mathbb{R}$ that $\mathsf{E} \sum_{i=1}^n n^{-1/2} \sigma^{-1} c = \mathrm{O}(n^{1/2})$. Combine both to get $\mathsf{E} \sum_{i=1}^n |\nabla x_{inu}| = \mathrm{O}(n^{1/2})$. Further, by Assumption 1.2.3(ib) we have $\sup_{v\in\mathbb{R}} \mathsf{f}(v) < \infty$. Therefore $\mathcal{E}_a = \mathrm{O}(n_a^{\varsigma})$ where $\varsigma_a = 3/4 - \eta$.

*The condition $\varsigma_a < 2v_a$.* This holds since $0 < \eta$ so that $\varsigma_a = 3/4 - \eta < 1 = 2v_a$ since $v_a = 1/2$.

2.3. The martingale term $\mathcal{R}_{n,1}^b$. We show that $\mathcal{R}_{n,1}^b = \mathrm{o}_{\mathsf{P}}(1)$ by applying Theorem 1.3.11 to it. We set $v_b = 1/2$ and index $\ell = m$ and consider

$$z_{i\ell,b} = \mathsf{E}_i z_{i\ell,a} = \mathsf{E}_i J_i(c, c + \nabla \tilde{x}'_{inu} u_m)$$

as defined in point 2.1. Note that $z_{i\ell,b}$ is $\mathcal{F}_i$-adapted and that $|z_{i\ell,b} - \mathsf{E}_i z_{i\ell,b}| \leq 2$ as $|J_i| \leq 1$. We show that the parameters are $\lambda_b = (1/4 - \eta)(\dim x + 1)$ and $\varsigma_b = 3/4 - \eta$. We verify the conditions of Theorem 1.3.11.

*The parameter* $\lambda_b$. We have $\lambda_b = (1/4 - \eta)(\dim x + 1)$, as in point 2.2.

*The parameter* $\varsigma_b$. We show that $\mathcal{E}_b = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{i\ell,b}^2 = \mathrm{O}(n^{3/4-\eta})$. Note that $z_{i\ell,b}^2 = \mathsf{E}_i^2 z_{i\ell,a}$. Use Jensen's inequality to show $\mathsf{E}_{i-1}\mathsf{E}_i^2 z_{i\ell,a} \leq \mathsf{E}_{i-1}\mathsf{E}_i z_{i\ell,a}^2$. We have that $\mathsf{E}_i z_{i\ell,a}^2 = \mathsf{E}_i |z_{i\ell,a}|$. In point 2.2, we have derived a bound for $\mathsf{E}_i z_{i\ell,a}$ in (1.30). We therefore have

$$\mathsf{E}_i |z_{i\ell,a}| \leq n^{1/4-\eta} B \mathsf{E}_i |\nabla \check{x}_{in}| \sup_{v \in \mathbb{R}} \mathsf{f}(v).$$

We get

$$\mathcal{E}_b \leq \mathrm{O}(n^{1/4-\eta}) \sup_{v \in \mathbb{R}} \mathsf{f}(v) \mathsf{E} \sum_{i=1}^n \mathsf{E}_{i-1} |\nabla x_{inu}|.$$

As in point 2.2 apply the law of iterated expectations to the expectation of the sum and use Assumption 1.2.3(ib) and Assumption 1.2.3(iii) to get that $\mathcal{E}_b = \mathrm{O}(n_b^{\varsigma})$ where $\varsigma_b = 3/4 - \eta$.

*The condition* $\varsigma_b < 2v_b$. It holds since $0 < \eta$ so that $\varsigma_b = 3/4 - \eta < 1 = 2v_b$ since $v_b = 1/2$.

**3. The oscillation term $\mathcal{R}_{n,2}$.** In this section we show that $\mathcal{R}_{n,2}$ is $\mathrm{o}_{\mathsf{P}}(1)$. The proof relies on finding a bound of $|R_n(u) - R_n(u_m)|$ that is a function of the grid points $u_m$. We then apply Theorem 1.3.11 to handle each of the resulting families of martingales.

3.1 The term $S_n$. Let $r_i(u_m, u, c) = J_i(c, c + \nabla \tilde{x}'_{inu} u) - J_i(c, c + \nabla \tilde{x}'_{inu} u_m)$ and write $R_n(u, c) - R_n(u_m, c) = n^{-1/2} \sum_{i=1}^n \{r_i(u_m, u, c) - \mathsf{E}_{i-1} r_i(u_m, u, c)\}$. Note that the two indicator functions $1_{(\nabla \tilde{\varepsilon}_i \leq c)}$ cancel so that

$$r_i(u_m, u, c) = s_i(u_m, u, c) = J_i(c + \nabla \tilde{x}'_{inu} u_m, c + \nabla \tilde{x}'_{inu} u).$$

Let

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

and note $R_n(u,c) - R_n(u_m,c) = S_n(u_m,u,c)$. Therefore $\mathcal{R}_{n,2}$ can be written as

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

3.2. Bounding the function $s_i(u_m,u,c)$. Let $\underline{z}_{im} = J_i(\underline{c}_{im}, \overline{c}_{im})$, where

$$\underline{c}_{im} = c + \nabla\tilde{x}'_{inu}u_m - \delta|\nabla\tilde{x}_{inu}|, \quad \overline{c}_{im} = c + \nabla\tilde{x}'_{inu}u_m + \delta|\nabla\tilde{x}_{inu}|. \tag{1.31}$$

From $c - |\nabla\tilde{x}_{inu}||\delta| < c + \nabla\tilde{x}'_{inu}(u-u_m) \le c + |\nabla\tilde{x}_{inu}||\delta|$ it follows by adding $\nabla\tilde{x}'_{inu}u_m$ that

$$\underline{c}_{im} \le c + \nabla\tilde{x}'_{inu}u \le \overline{c}_{im}$$

As a consequence

$$|s_i(u_m,u,c)| \le \underline{z}_{im} = J_i(\underline{c}_{im}, \overline{c}_{im}). \tag{1.32}$$

3.3. Bounding the function $|S_n(u_m,u,c)|$. The triangle inequality gives

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

Using the bound (1.32) to the $|s_i|$ in point 3.1 we get the bound

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

which is uniform in $u$, $c$. Thus, to show that $\mathcal{R}_{n,2}$ is $\mathsf{o_P}(1)$ it suffices to show $M_{mn} = \mathsf{o_P}(1)$ uniformly in $m$.

3.4. Martingale decomposition. We need to show $M_{mn} = \mathsf{o_P}(1)$ uniformly in $m$. To do so, we decompose it into two martingale and a compensator term. Add and subtract $n^{-1/2}\sum_{i=1}^n \mathsf{E}_i\underline{z}_{im}$ 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*

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

and a *compensator term* $\overline{M}_{mn} = n^{-1/2}\sum_{i=1}^n \mathsf{E}_{i-1}\underline{z}_{im}$. Thus, it suffices to show that $\widetilde{\mathcal{M}}^c_n = \max_{1\le m\le M}\widetilde{M}^c_{mn}$, $\widetilde{\mathcal{M}}^d_n = \max_{1\le m\le M}\widetilde{M}^d_{mn}$, $\overline{\mathcal{M}}_n = \max_{1\le m\le M}\overline{M}_{mn}$, are $\mathsf{o_P}(1)$.

3.5. The compensator term $\overline{M}_{mn}$. We show that $\overline{\mathcal{M}}_n = \mathrm{o}_{\mathsf{P}}(1)$. We start by bounding $\mathsf{E}_{i-1}\underline{z}_{im}$. Recall from (1.31) that $\underline{c}_{im} = c + \nabla\tilde{x}'_{inu}u_m - \delta|\nabla\tilde{x}_{inu}|$ and $\overline{c}_{im} = c+\nabla\tilde{x}'_{inu}u_m+\delta|\nabla\tilde{x}_{inu}|$. Use the law of iterated expectations to get $\mathsf{E}_{i-1}\underline{z}_{im} = \mathsf{E}_{i-1}\mathsf{E}_i\underline{z}_{im}$. Note that

$$\mathsf{E}_i\underline{z}_{im} = \mathsf{P}(\check{\varepsilon}_i - \sqrt{2}\overline{c}_{im} \leq \check{\varepsilon}_{i+1} \leq \check{\varepsilon}_i - \sqrt{2}\underline{c}_{im}|\mathcal{F}_i)$$

where $\underline{z}_{im}$ contains four random variables that are adapted to three different filtrations: we have $\mathcal{F}_{i-1}$-adapted $x_i$; $\mathcal{F}_i$-adapted $\varepsilon_i$ and $x_{i+1}$; and $\mathcal{F}_{i+1}$-adapted $\varepsilon_{i+1}$. Given $\mathcal{F}_i$, only $\varepsilon_{i+1}$ remains random. Because only $\varepsilon_{i+1}$ is independent of $\mathcal{F}_i$, we can write the conditional expectation as an integral over the density $\mathsf{f}$ of the error term $\check{\varepsilon}_{i+1}$ to get

$$\mathsf{E}_i\underline{z}_{im} = \int_{\check{\varepsilon}_i-\sqrt{2}\overline{c}_{im}}^{\check{\varepsilon}_i-\sqrt{2}\underline{c}_{im}} \mathsf{f}(v)dv,$$

to which we apply the Taylor expansion around $\check{\varepsilon}_i - \sqrt{2}\overline{c}_{im}$ with $\sqrt{2}\underline{c}_{im} - \sqrt{2}\overline{c}_{im} = 2\delta|\nabla\check{x}_{inu}|$ to get $\mathsf{E}_i\underline{z}_{im} = 2\delta|\nabla\check{x}_{inu}|\mathsf{f}(v^*)$, where $\check{\varepsilon}_i - \sqrt{2}\overline{c}_{im} \leq v^* \leq \check{\varepsilon}_i - \sqrt{2}\underline{c}_{im}$. Use $\mathsf{f}(v^*) < \sup_{v\in\mathbb{R}}\mathsf{f}(v)$ to bound $\mathsf{E}_i\underline{z}_{im}$ uniformly in $m$ as

$$\mathsf{E}_i\underline{z}_{im} \leq 2\delta|\nabla\check{x}_{inu}|\sup_{v\in\mathbb{R}}\mathsf{f}(v). \tag{1.33}$$

By the law of iterated expectation on $\mathsf{E}_{i-1}\mathsf{E}_i|\nabla\check{x}_{inu}| = \mathsf{E}_{i-1}|\nabla\check{x}_{inu}|$ we have

$$\overline{\mathcal{M}}_n \leq \delta\mathrm{O}(n^{-1/2})\sup_{v\in\mathbb{R}}\mathsf{f}(v)\sum_{i=1}^n \mathsf{E}_{i-1}|\nabla x_{inu}|.$$

We consider the two components of $\nabla x_{inu}$ in turn. For the regular regressors $\nabla x_{in}$ we appeal to Assumption 1.2.3(iii) and the Markov Inequality such that $\sum_{i=1}^n \mathsf{E}_{i-1}|\nabla x_{in}| = \mathrm{O}_{\mathsf{P}}(n^{1/2})$. For the term related to $a$ we have for $c \in \mathbb{R}$ that $\sum_{i=1}^n \mathsf{E}_{i-1}n^{-1/2}\sigma^{-1}c = \mathrm{O}(n^{1/2})$. Combine both to get $\sum_{i=1}^n \mathsf{E}_{i-1}|\nabla x_{inu}| = \mathrm{O}_{\mathsf{P}}(n^{1/2})$. Therefore, with Assumption 1.2.3(ib) we have $\sup_{v\in\mathbb{R}}\mathsf{f}(v) < \infty$ and thus $\overline{\mathcal{M}}_n = \delta\mathrm{O}_{\mathsf{P}}(1)$ where the $\mathrm{O}_{\mathsf{P}}(1)$-term does not depend on $\delta$.

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 $n$. Thus, we consider an arbitrary, but fixed $\gamma > 0$. We are still free to choose $\delta$ which will be exploited now. Given $\gamma > 0$ we can find a constant $C$ not depending on $\delta$ 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 $\widetilde{\mathcal{M}}_n^c$. We show $\widetilde{\mathcal{M}}_n^c = \mathsf{o}_{\mathsf{P}}(1)$, using Theorem 1.3.11. We set $v_c = 1/2$ and index $\ell = m$ and consider

$$z_{i\ell,c} = \underline{z}_{im} = J_i(\underline{c}_{im}, \overline{c}_{im}),$$

which is $\mathcal{F}_{i+1}$-adapted. Note that $|z_{i\ell,c} - \mathsf{E}_i z_{i\ell,c}| \leq 1$ as $0 \leq J_i(\underline{c}_{im}, \overline{c}_{im}) \leq 1$ and $0 \leq \mathsf{E}_i J_i(\underline{c}_{im}, \overline{c}_{im}) \leq 1$. We have parameters $L = M$ and $\lambda_c = (1/4 - \eta)(\dim x + 1)$ and $\varsigma_c = 1/2$. We verify the conditions of Theorem 1.3.11.

*The parameter* $\lambda_c$. We have $\lambda_c = (1/4-\eta)(\dim x+1)$, as the set-up is identical to point 2.2.

*The parameter* $\varsigma_c$. We show that $\mathcal{E}_c = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_i z_{i\ell,c}^2 = \mathsf{o}_{\mathsf{P}}(1)$. Because $J_i$ is an indicator function and $\underline{c}_{im} \leq \overline{c}_{im}$, we have $J_i^2 = J_i$ and thus it suffices to consider $\mathsf{E}_i z_{i\ell,c}$. In (1.33) we have found a bound on $\mathsf{E}_i \underline{z}_{im}$ which is uniform in $u$. As $z_{i\ell,c} = \underline{z}_{im}$ we have

$$\mathcal{E}_c \leq \mathrm{O}(\delta) \sup_{v \in \mathbb{R}} \mathsf{f}(v) \mathsf{E} \sum_{i=1}^n |\nabla x_{inu}|.$$

As in point 2.2, we have by Assumption 1.2.3(iii) that the expectation of the sum is of order $\mathrm{O}(n^{1/2})$. By Assumption 1.2.3(ib), we have $\sup_{v \in \mathbb{R}} \mathsf{f}(v) < \infty$. As $\delta$ is a constant, we find $\mathcal{E}_c = \mathrm{O}(n^{\varsigma_c})$ where $\varsigma_c = 1/2$.

*The condition* $\varsigma_c < 2v_c$. This holds since $0 < \eta$ so that $\varsigma_c = 1/2 < 1 = 2v_c$ since $v_c = 1/2$.

3.7. The martingale $\widetilde{\mathcal{M}}_n^d$. We show $\widetilde{\mathcal{M}}_n^d = \mathsf{o}_{\mathsf{P}}(1)$ using Theorem 1.3.11. We set $v_d = 1/2$ and index $\ell = m$ and consider

$$z_{i\ell,d} = \mathsf{E}_i \underline{z}_{im} = \mathsf{E}_i J_i(\underline{c}_{im}, \overline{c}_{im}),$$

which is $\mathcal{F}_i$-adapted. Note that $|z_{i\ell,c} - \mathsf{E}_{i-1} z_{i\ell,c}| \leq 1$ as $0 \leq \mathsf{E}_i J_i(\underline{c}_{im}, \overline{c}_{im}) \leq 1$ and $0 \leq \mathsf{E}_{i-1} J_i(\underline{c}_{im}, \overline{c}_{im}) \leq 1$. We have parameters $L = M$ and $\lambda_d = (1/4-\eta)(\dim x+1)$ and $\varsigma_d = 1/2$. We verify the conditions of Theorem 1.3.11.

*The parameter* $\lambda_d$. We have $\lambda_c = (1/4-\eta)(\dim x+1)$, as the set-up is identical to point 2.2.

*The parameter* $\varsigma_d$. We show that $\mathcal{E}_d = \mathsf{E} \max_{1 \leq \ell \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{i\ell,d}^2 = \mathsf{o}_{\mathsf{P}}(1)$. Note that $z_{i\ell,d}^2 = \mathsf{E}_i^2 z_{i\ell,c}$. Use Jensen's inequality to show $\mathsf{E}_{i-1}\mathsf{E}_i^2 z_{i\ell,c} \leq \mathsf{E}_{i-1}\mathsf{E}_i z_{i\ell,c}^2 =$

$\mathsf{E}_{i-1}z_{i\ell,c}^2$. Because $J_i$ is an indicator function and $\underline{c}_{im} < \overline{c}_{im}$, we have $J_i^2 = J_i$ and thus it suffices to consider $\mathsf{E}_{i-1}\mathsf{E}_i z_{i\ell,c}$. In (1.33) we have found a bound on $\mathsf{E}_i \underline{z}_{im}$ which is uniform in $u$. As $\mathsf{E}_i z_{i\ell,c} = \mathsf{E}_i \underline{z}_{im}$ we have

$$\mathcal{E}_d \leq \mathrm{O}(\delta) \sup_{v\in\mathbb{R}} \mathsf{f}(v) \mathsf{E} \sum_{i=1}^{n} \mathsf{E}_{i-1} |\nabla x_{inu}|.$$

Apply the law of iterated expectations to get $\mathsf{E}\sum_{i=1}^n \mathsf{E}_{i-1}|\nabla x_{inu}| = \mathsf{E}\sum_{i=1}^n |\nabla x_{inu}|$. The bound on $\mathcal{E}_d$ is the same as the bound on $\mathcal{E}_c$. We therefore get with Assumption 1.2.3(ib) and (iii) that $\mathcal{E}_d = \mathrm{O}(n^{\varsigma_d})$ where $\varsigma_d = 1/2$.

*The condition* $\varsigma_d < 2v_d$. This holds since $0 < \eta$ so that $\varsigma_d = 1/2 < 1 = 2v_d$ since $v_d = 1/2$.

**4. Conclusion.** We have shown that $\widetilde{\mathcal{M}}_n^c = \mathrm{o}_\mathsf{P}(1)$ and that $\widetilde{\mathcal{M}}_n^d = \mathrm{o}_\mathsf{P}(1)$. Since also $\overline{\mathcal{M}}_n = \mathrm{o}_\mathsf{P}(1)$ as shown in point 3.5, we have that $\mathcal{R}_{n,2}$ is $\mathrm{o}_\mathsf{P}(1)$. In points 2 it was shown that $\mathcal{R}_{n,1} = \mathrm{o}_\mathsf{P}(1)$. In combination, $\mathcal{R}_n = \mathrm{o}_\mathsf{P}(1)$. ■

Finally, we prove Theorem 1.3.1 by decomposing the two-sided indicator functions into two one-sided indicator functions.

**Proof of Theorem 1.3.1**. Let $c^* = c + n^{-1/2}\sigma^{-1}ac$. Because

$$1_{(|\nabla\tilde{\epsilon}_i - \nabla\tilde{x}_{in}'b| \leq c^*)} = 1_{(\nabla\tilde{\epsilon}_i \leq c^* + \nabla\tilde{x}_{in}'b)} - 1_{(\nabla\tilde{\epsilon}_i < -c^* + \nabla\tilde{x}_{in}'b)}$$

and by the definition of the one and two-sided empirical process, we have

$$\mathbb{G}_n(a,b,c) = \mathbb{F}_n(a,b,c) - \lim_{c^\dagger \downarrow c} \mathbb{F}_n(a,b,-c^\dagger)$$

for any $c > 0$. Using this and the triangle inequality, we have

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

The terms on the right hand side vanish uniformly in $a, b$ by Theorem 1.3.5 using Assumptions 1.2.2, 1.2.3 (ib,iii). ■

### 1.4.5 Proofs of the results of compensator terms

As for the proof of Theorem 1.3.1, we proceed by showing first the one-sided equivalent of Theorem 1.3.2. Recall the definition of the pseudo compensator $\overline{\mathsf{F}}_n$ in (1.17) and of $Q_n(a,b,c)$ in (1.19).

Define a term which sets $a=0$ in $Q_n(a,b,c)$ as

$$Q_n(b,c) = n^{1/2}\{\overline{\mathsf{F}}_n(0,b,c) - \overline{\mathsf{F}}_n(0,0,c)\} - \mathsf{h}(c)\xi'(c)b, \tag{1.34}$$

a corresponding term which sets $b=0$ as

$$Q_n(a,c) = n^{1/2}\{\overline{\mathsf{F}}_n(a,0,c) - \overline{\mathsf{F}}_n(0,0,c)\} - ac\mathsf{h}(c)/\sigma. \tag{1.35}$$

and a remainder term with normalisation $N_a = n^{-1/2}\sigma^{-1}$

$$V_n(a,b,c) = \mathsf{h}(c+N_aac)\xi'(c+N_aac)b - \mathsf{h}(c)\xi'(c)b \tag{1.36}$$

Note $\overline{\mathsf{F}}_n(a,b,c) = \overline{\mathsf{F}}_n(0,b,c+N_aac)$. Add and subtract $\mathsf{h}(c+N_aac)\xi'(c+N_aac)b$ and $n^{1/2}\overline{\mathsf{F}}_n(a,0,c) = n^{1/2}\overline{\mathsf{F}}_n(0,0,c+N_aac)$. from $Q_n(a,b,c)$ and apply the triangle inequality to get

$$|Q_n(a,b,c)| \leq |Q_n(b,c+N_aac)| + |Q_n(a,c)| + |R_n(a,c)| \tag{1.37}$$

We therefore prove Theorem 1.3.6 by considering $Q_n(b,c)$ and $Q_n(a,c)$.

**Theorem 1.4.3.** *Suppose Assumptions 1.2.2 and 1.2.3 (ii,iii) hold. Then, for all $B>0$ and $0<\eta<1/4$, we get with $Q_n(b,c)$ defined as in* (1.34) *that* $\sup_{|b|\leq n^{1/4-\eta}B}\sup_{c\in\mathbb{R}}|Q_n(b,c)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta})$.

**Proof of Theorem 1.4.3**. Write $Q_n(b,c) = n^{-1/2}\sum_{i=1}^n \mathcal{P}_i(b,c)$, where

$$\mathcal{P}_i(b,c) = \mathsf{E}_{i-1}\{1_{(\nabla\tilde{\varepsilon}_i - \nabla\tilde{x}'_{in}b\leq c)} - 1_{(\nabla\tilde{\varepsilon}_i\leq c)}\} - \mathsf{h}(c)\mathsf{E}_{i-1}(\nabla\tilde{x}'_{in}|\nabla\tilde{\varepsilon}_i = c)b.$$

We apply Theorem 1.3.7 to $\mathcal{P}_i(b,c)$ as follows. First note that the conditional expectation can be written as a conditional probability

$$\mathsf{E}_{i-1}\{1_{(\nabla\tilde{\varepsilon}_i - \nabla\tilde{x}'_{in}b\leq c)} - 1_{(\nabla\tilde{\varepsilon}_i\leq c)}\} = \mathsf{P}(\nabla\tilde{\varepsilon}_i - \nabla\tilde{x}'_{in}b\leq c|\mathcal{F}_{i-1}) - \mathsf{P}(\nabla\tilde{\varepsilon}_i\leq c|\mathcal{F}_{i-1}).$$

Note that $\mathcal{P}_i(b,c)$ contains four random variables that are adapted to three different filtrations: we have $\mathcal{F}_{i-1}$-adapted $x_{i+1}$; $\mathcal{F}_i$-adapted $\varepsilon_i$ and $x_{i+1}$; and $\mathcal{F}_{i+1}$-adapted

$\varepsilon_{i+1}$. Let $Y_i = \nabla\tilde{\varepsilon}_i$ and $X_i = \nabla\tilde{x}_i$ and let $\mathsf{m}_i(y,x)$ denote the joint conditional density of $Y_i, X_i$ given $\mathcal{F}_{i-1}$ (see equation (1.8)). By Assumption 1.2.3(iii) we have that $\mathsf{m}_i(y,x)$ is differentiable with respect to $y$. We get the bound

$$|\mathcal{P}_i(b,c)| \leq 2^{-1}|b|^2 \sup_{y\in\mathbb{R},x\in\mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)|\,\mathsf{E}_{i-1}|\nabla\tilde{x}_{in}|^2,$$

where $\dot{\mathsf{m}}_i(y|x)$ is the derivative of the conditional density of $\nabla\tilde{\varepsilon}_i$ given $\nabla\tilde{x}_i$ and $\mathcal{F}_{i-1}$ (see equation (1.8)). Since $|b| \leq n^{1/4-\eta}B$ and using the triangular inequality we get

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

Due to Assumption 1.2.3(iii) and Markov's inequality, $n^{-1}\sum_{i=1}^{n} \mathsf{E}_{i-1}|n^{1/2}\nabla x_{in}|^2 = \mathrm{O}_{\mathsf{P}}(1)$ and by Assumption 1.2.3(ii) we have that $\max_{1\leq i\leq n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}|\dot{\mathsf{m}}_i(y|x)| < \infty$. The upper bound of $|Q_n(b,c)|$ is of order $\mathrm{O}(n^{-2\eta})$ uniformly in $b$ and $c$. ■

While Theorem 1.4.3 controlled for variation in $b$ and $c$, we now analyse variation in $a$ and $c$.

**Theorem 1.4.4.** *Suppose Assumption 1.2.2 and 1.2.3 (ia,ib) hold. Then, for all $B > 0$ and $0 < \eta < 1/4$, we get with $\mathcal{Q}(a,c)$ defined as in* (1.35) *that*

$$\sup_{|a|\leq n^{1/4-\eta}B} \sup_{c\in\mathbb{R}} |\mathcal{Q}(a,c)| = \mathrm{O}(n^{-2\eta}).$$

**Proof of Theorem 1.4.4**. Write $Q_n(a,c) = n^{-1/2}\sum_{i=1}^{n}\mathcal{P}_i(a,c)$, where

$$\mathcal{P}_i(a,c) = \mathsf{E}_{i-1}\{1_{(\nabla\tilde{\varepsilon}_i\leq c+N_a ac)} - 1_{(\nabla\tilde{\varepsilon}_i\leq c)}\} - n^{-1/2}\sigma^{-1}ac\mathsf{h}(c)$$

Note $\nabla\varepsilon_i$ is independent of $\mathcal{F}_{i-1}$ and write the conditional expectation as an integral

$$\mathsf{E}_{i-1}\{1_{(\nabla\tilde{\varepsilon}_i\leq c+N_a ac)} - 1_{(\nabla\tilde{\varepsilon}_i\leq c)}\} = \int_c^{c+N_a ac} \mathsf{h}(u)du.$$

As $\mathsf{f}$ is differentiable by Assumption 1.2.2, the density $\mathsf{h}$ is also differentiable. Thus we can Taylor expand the integral at $c$ and get

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

where $|\tilde{c} - c| \leq |n^{-1/2}\sigma^{-1}ac|$. There exists $n_0 > 0$, so for any $n > n_0$ we have $|n^{-1/2}\sigma^{-1}a| \leq 1/2$. For $n > n_0$ we can therefore apply the second inequality in Theorem 1.4.1 to obtain $n^{-1}\sigma^{-2}a^2c^2 \leq 16a_n^2\tilde{c}^2$. Therefore, with $\tilde{c}^2|\dot{\mathsf{h}}(\tilde{c}^2)| \leq \sup_{v\in\mathbb{R}} v^2|\dot{\mathsf{h}}(v)|$ we have

$$\mathcal{P}_i(a,c) \leq n^{-1}\sigma^{-2}a^2 \sup_{v\in\mathbb{R}} v^2|\dot{\mathsf{h}}(v)|$$

Use $|a| \leq n^{1/4-\eta}B$ and $\sup_{c\in\mathbb{R}} c^2|\dot{\mathsf{h}}(c)| < \infty$ by Assumption 1.2.3(ia,ib) and Theorem 1.2.1(d) to get $|\mathcal{P}_i(a,c)| = \mathrm{O}(n^{-1/2-2\eta})$ uniformly in $a$, $c$ and $i$. Then it follows that $|Q_n(a,c)| = \mathrm{O}(n^{-2\eta})$ uniformly in $a$ and $c$. ■

Using the decomposition of $|Q_n(a,b,c)|$ in (1.37), we combine the results from Theorem 1.4.3 and 1.4.4 in the proof of Theorem 1.3.6.

**Proof of Theorem 1.3.6**. We bound our object of interest $Q_n(a,b,c)$ as in (1.37) with $Q_n(b,c_a)$ and $Q_n(a,c)$ and $V_n(a,b,c)$. We have $Q_n(a,b,c) = \mathrm{O}_{\mathsf{P}}(n^{-2\eta})$, if the following three conditions are satisfied:

$$\begin{aligned}
\mathcal{Q}_{1n} &= \sup_{|b|\leq n^{1/4-\eta}B} \sup_{c\in\mathbb{R}} |Q_n(b,c)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}),\\
\mathcal{Q}_{2n} &= \sup_{|a|\leq n^{1/4-\eta}B} \sup_{c\in\mathbb{R}} |Q_n(a,c)| = \mathrm{O}(n^{-2\eta}),\\
\mathcal{Q}_{3n} &= \sup_{|a|,|b|\leq n^{1/4-\eta}B} \sup_{c\in\mathbb{R}} |V_n(a,b,c)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}).
\end{aligned}$$

We bound the three terms in turn. For $\mathcal{Q}_{1n}$ we apply Theorem 1.4.3 using Assumptions 1.2.2 and 1.2.3(ii,iii). For $\mathcal{Q}_{2n}$ we apply Theorem 1.4.4 using Assumptions 1.2.2 and 1.2.3(ia,ib). Finally, we show that $\mathcal{Q}_{3n}$ vanishes. Write $R_n(a,b,c) = n^{-1/2}\sum_{i=1}^n \mathcal{P}_i(a,b,c)$, where with $c^\dagger = c + n^{-1/2}\sigma^{-1}ac$ we have

$$\mathcal{P}_i(a,b,c) = \mathsf{h}(c^\dagger)\mathsf{E}_{i-1}(\nabla\tilde{x}'_{in}|\nabla\tilde{\varepsilon}_i = c^\dagger)b - \mathsf{h}(c)\mathsf{E}_{i-1}(\nabla\tilde{x}'_{in}|\nabla\tilde{\varepsilon}_i = c)b.$$

Apply the second part of Theorem 1.3.7 as follow. Note that $\mathcal{P}_i(a,b,c)$ contains four random variables that are adapted to three different filtrations: we have $\mathcal{F}_{i-1}$-adapted $x_{i+1}$; $\mathcal{F}_i$-adapted $\varepsilon_i$ and $x_{i+1}$; and $\mathcal{F}_{i+1}$-adapted $\varepsilon_{i+1}$. Let $Y_i = \nabla\tilde{\varepsilon}_i$ and $X_i = \nabla\tilde{x}_i$ and let $\mathsf{m}_i(y,x)$ denote the joint conditional density of $Y_i, X_i$ given $\mathcal{F}_{i-1}$

as in (1.8). By Assumption 1.2.3(ii) we have that $\mathsf{m}_i(y,x)$ is differentiable with respect to $y$. We get the bound

$$|\mathcal{P}_i(a,b,c)| \leq |n^{-1/2}\sigma^{-1}ab| \sup_{y\in\mathbb{R},x\in\mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)|\mathsf{E}_{i-1}|\nabla\tilde{x}_{in}|,$$

where $\dot{\mathsf{m}}_i(y|x)$ is the derivative of the conditional density of $\nabla\tilde{\varepsilon}_i$ given $\nabla\tilde{x}_i$ and $\mathcal{F}_{i-1}$ (see equation (1.8)). Since $|a|,|b| \leq n^{1/4-\eta}B$ and using the triangle inequality we get

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

Due to Assumption 1.2.3(iii) and Markov's inequality, $n^{-1}\sum_{i=1}^{n}\mathsf{E}_{i-1}|n^{1/2}\nabla x_{in}| = \mathrm{O}_{\mathsf{P}}(1)$ and by Assumption 1.2.3(ii) we have that $\max_{1\leq i\leq n}\sup_{y\in\mathbb{R},x\in\mathbb{R}^p}|\dot{\mathsf{m}}_i(y|x)| < \infty$. The upper bound of $|R_n(a,b,c)|$ is of order $\mathrm{O}(n^{-2\eta})$ uniformly in $a$, $b$ and $c$. ■

We can now expand the Theorem 1.3.6 to the absolute empirical processes. In this case we consider the distribution $\mathsf{G}$ as the reference distribution to pick our quantile $c$, as we have absolute differenced error terms.

**Proof of Theorem 1.3.2**. Use the following bound

$$1_{(|\nabla\tilde{\varepsilon}_i-\nabla\tilde{x}'_{in}b|\leq c+n^{-1/2}\sigma^{-1}ac)} \leq 1_{(\nabla\tilde{\varepsilon}_i\leq c+n^{-1/2}\sigma^{-1}ac+\nabla\tilde{x}'_{in}b)} - 1_{(\nabla\tilde{\varepsilon}_i<-c-n^{-1/2}\sigma^{-1}ac+\nabla\tilde{x}'_{in}b)}$$

and write $\overline{\mathsf{G}}_n(a,b,c) = \overline{\mathsf{F}}_n(a,b,c) - \lim_{c^\dagger\downarrow -c}\overline{\mathsf{F}}_n(a,b,c^\dagger)$ for any $c > 0$. By the triangle inequality

$$|\mathcal{G}_n(a,b,c)| \leq |Q_n(a,b,c)| + \lim_{c^\dagger\downarrow -c}|Q_n(a,b,c^\dagger)|,$$

where we use the symmetry of $\mathsf{h}$ for the bias term $\mathcal{B}$. Apply Theorem 1.3.6 with Assumptions 1.2.2 and 1.2.3(ib,ii,iii) to show that $|\mathcal{G}_n(a,b,c)|$ vanishes for large $n$ uniformly in $a$,$b$ and $c$. ■

### 1.4.6 Proof of Theorem 1.3.3

The idea of the following proof is to find a representation of $\widehat{\mathsf{H}}_n$ defined in (1.15) in terms of the empirical process results of $\mathbb{F}_n$ defined in (1.18) and other terms that vanish.

**Proof of Theorem 1.3.3**. Introduce a compensator to $\widehat{\mathsf{H}}_n$ as

$$\overline{\mathsf{H}}_n(a,b,c,\delta_n) = n^{-1}\sum_{i=1}^{n}\mathsf{E}_i 1_{\{|\nabla\tilde{\varepsilon}_i - |c|(1+n^{-1/2}\sigma^{-1}|a|) - \nabla x_{in}'b| \leq |c|n^{-1/2}\sigma^{-1}\delta_n\}}$$

and combine $\widehat{\mathsf{H}}_n$ and $\overline{\mathsf{H}}_n$ to an alternative empirical process

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

**1. Decomposition of $\widehat{\mathsf{H}}_n$.** Write

$$n^{1/2}\widehat{\mathsf{H}}_n(a,b,c,\delta_n) = \mathbb{H}_n(a,b,c,\delta_n) + n^{1/2}\overline{H}_n(a,b,c,\delta_n)$$

by adding and subtracting the compensator term $\overline{\mathsf{H}}_n$. Then, it suffices to show for some $c \in \mathbb{R}$ and $\delta_n \to 0$ as $n \to \infty$ that

$$\sup_{|a|,|b|\leq n^{1/4-\eta}B}\mathbb{H}_n(a,b,c,\delta_n) = \mathsf{o}_\mathsf{P}(1), \quad \sup_{|a|,|b|\leq n^{1/4-\eta}B} n^{1/2}\overline{\mathsf{H}}_n(a,b,c,\delta_n) = \mathsf{o}_\mathsf{P}(1).$$

**2. The compensator term $\overline{\mathsf{H}}_n$.** We show that $n^{1/2}\overline{\mathsf{H}}_n = \mathsf{o}_\mathsf{P}(1)$ uniformly in $a$ and $b$ for $c \in \mathbb{R}$ and $\delta_n \to 0$ as $n \to \infty$.

2.1 A bound on $\overline{\mathsf{H}}_n$. Write the conditional expectation in the compensator term as an integral

$$\mathsf{E}_i 1_{(|\nabla\tilde{\varepsilon}_i - |c|(1+n^{-1/2}\sigma^{-1}|a|) - \nabla\tilde{x}_{in}'b| \leq |c|n^{-1/2}\sigma^{-1}\delta_n)} = \int_{\underline{c}}^{\bar{c}}\mathsf{f}(v)dv,$$

where the bounds of the integral satisfy $|\underline{c} - \bar{c}| = 2\delta_n n^{-1/2}\sigma^{-1}|c|$ with

$$\underline{c} = \check{\varepsilon}_i - \sqrt{2}|c|(1+n^{-1/2}\sigma^{-1}|a|) - \nabla\check{x}_{in}'b - \sqrt{2}|c|n^{-1/2}\sigma^{-1}\delta_n,$$
$$\bar{c} = \check{\varepsilon}_i - \sqrt{2}|c|(1+n^{-1/2}\sigma^{-1}|a|) - \nabla\check{x}_{in}'b + \sqrt{2}|c|n^{-1/2}\sigma^{-1}\delta_n.$$

Apply the Taylor expansion to the integral to get

$$\int_{\underline{c}}^{\bar{c}}\mathsf{f}(v)dv = 2\sqrt{2}\delta_n n^{-1/2}\sigma^{-1}|c|\mathsf{f}(c^*),$$

where $|c^*-\underline{c}| \leq 2\sqrt{2}\delta_n n^{-1/2}\sigma^{-1}|c|$. Note that for $X, Y, Z \in \mathbb{R}$ we have $|X-Y| \leq |Z|$ implies $|X| \leq |Z| + |Y|$ and let $\tilde{c} = \sqrt{2}(1 + n^{-1/2}\sigma^{-1}|a|)$ to get

$$|c^* + \tilde{c}| \leq 3\sqrt{2}n^{-1/2}\sigma^{-1}\delta_n|c| + |\check{\varepsilon}_i - \nabla\check{x}'_{in}b| = \frac{3n^{-1/2}\sigma^{-1}\delta_n}{1 + n^{-1/2}\sigma^{-1}|a|}|\tilde{c}| + |\check{\varepsilon}_i - \nabla\check{x}'_{in}b|.$$

For a sufficiently large $n$, we have $(1 + n^{-1/2}\sigma^{-1}|a|)^{-1}3n^{-1/2}\sigma^{-1}\delta_n < 1/2$ and we apply Theorem 1.4.1 to get $|\tilde{c}| \leq 2(|c^*| + |\varepsilon_i| + |\nabla\check{x}'_{in}b|)$. Use this bound on $|\tilde{c}| = \sqrt{2}|c|(1 + n^{-1/2}\sigma^{-1}|a|)$ and get

$$\int_{\underline{c}}^{\bar{c}} \mathsf{f}(v)dv \leq 4\delta_n n^{-1/2}\sigma^{-1}(1 + n^{-1/2}\sigma^{-1}|a|)^{-1}(|c^*| + |\check{\varepsilon}_i| + |\nabla\check{x}'_{in}b|)\mathsf{f}(c^*).$$

Given $|a| \leq n^{1/4-\eta}B$ we have $1 + n^{-1/2}\sigma^{-1}|a| \to 1$ and thus

$$n^{1/2}\overline{\mathsf{H}}_n(a, b, c, \delta_n) \leq \delta_n \mathrm{O}(1)\left[\sup_{v\in\mathbb{R}}|v|\mathsf{f}(v) + n^{-1}\sum_{i=1}^{n}\left\{(|\nabla\check{x}'_{in}b| + |\check{\varepsilon}_i|)\sup_{v\in\mathbb{R}}\mathsf{f}(v)\right\}\right].$$

2.2. Pulling the strings together. We consider the individual terms in the bound of $\overline{\mathsf{H}}_n$ from point 2.1. By the Jensen inequality we have $n^{-1}\sum_{i=1}^n |\nabla\check{x}_{in}| \leq (n^{-1}\sum_{i=1}^n |\nabla\check{x}_{in}|^2)^{1/2}$. Then, by the Markov inequality and Assumption 1.2.3(iii) we have that $n^{-1}\sum_{i=1}^n |\nabla\check{x}_{in}|^2 = \mathrm{O}_\mathsf{P}(n^{-1/2})$. Thus noting that $|b| \leq n^{1/4-\eta}B$, we have $n^{-1}\sum_{i=1}^n |\nabla\check{x}'_{in}b| \leq n^{-1}\sum_{i=1}^n |\nabla\check{x}_{in}||b| = \mathrm{O}_\mathsf{P}(n^{-\eta})$. By Assumption 1.2.3(ia) and a law of large number of independent random variables, we have $n^{-1}\sum_{i=1}^n |\varepsilon_i| \xrightarrow{\mathsf{P}} \mathsf{E}|\varepsilon_i| < \infty$. By Assumption 1.2.3(ib) we have $\sup_{v\in\mathbb{R}}|v|\mathsf{f}(v) < \infty$ and $\sup_{v\in\mathbb{R}}\mathsf{f}(v) < \infty$. Therefore, we have with $\delta_n \to 0$ as $n \to \infty$ that $n^{1/2}\overline{\mathsf{H}}_n(a, b, c, \delta_n)$ vanishes uniformly in $a, b$.

**3. The empirical process term $\mathbb{H}_n$.** We show that $\mathbb{H}_n(a, b, c, \delta_n) = \mathrm{o}_\mathsf{P}(1)$ uniformly in $a$ and $b$ for $c \in \mathbb{R}$ and $\delta_n \to 0$ and $n \to \infty$. Consider the indicator function

$$I_i(a, b, c, \delta_n) = 1_{(\nabla\check{\varepsilon}_i - \nabla\check{x}'_{in}b \leq c[1+n^{-1/2}\sigma^{-1}\{a+\delta_n \mathrm{sign}(c)\}])}$$

and let $I_i^E = I_i(a, b, c, \delta_n) - \mathsf{E}_i I_i(a, b, c, \delta_n)$. Note that

$$\mathbb{H}_n(a, b, c, \delta_n) = n^{-1/2}\sum_{i=1}^{n}\{I_i^E(a, b, c, \delta_n) - \lim_{\delta_n^\dagger \downarrow \delta_n} I_i^E(a, b, c, -\delta_n^\dagger)\}.$$

Add and subtract $I_i^E(0,0,c,0)$ from $\mathbb{H}_n(a,b,c,\delta_n)$. Then use the triangle inequality to get

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

Then for large $n$, we have that $\delta_n$ is sufficiently small such that $a + \delta_n \text{sign}(c) \leq 2a$. We define $\mathcal{J}_i(a,b,c) = J_i\{c, c(1+n^{-1/2}\sigma^{-1}a) + \nabla\tilde{x}_{in}'b\}$ with the $J_i$ function defined as in (1.21). It suffices to show

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

As $\mathcal{J}_i(a,b,c) = J_i(c, c + \nabla\tilde{x}_{inu}'u)$ where $u = (a,b')'$ and $\nabla x_{inu} = (\sqrt{2}n^{-1/2}c, \nabla x_{in}')'$ are as in the proof of Theorem 1.3.5 we can apply the same chaining argument. The key difference to Theorem 1.3.5 is that we subtract $\mathsf{E}_i\mathcal{J}_i(a,b,c)$ instead of $\mathsf{E}_{i-1}J_i(c, c+\nabla\tilde{x}_{inu}'u)$; the conditional expectations are in regard to different filtrations. We would therefore only need to consider $\mathcal{R}_{n,1}^a$ in point 2.1 and $\widetilde{M}_{mn}^c$ in point 3.4 of Theorem 1.3.5. Another minor difference relates to the bound of $|a|, |b|$ and it requires us to pick $B = 2\widetilde{B}$ in applying the proof of Theorem 1.3.5 to our present problem. Therefore, our object of interest vanishes uniformly over $a, b$ given Assumptions 1.2.2 and 1.2.3(ib,iii). ■

# 1.5 Conclusion

This chapter introduced a series of empirical process results for a class of empirical processes with differences of residuals. We developed mathematical techniques to overcome the temporal dependence structure introduced by the correlated error terms $\nabla\varepsilon_i$ and regressors $\nabla x_i$.

The results were based on a chaining argument and we applied the chaining in two different variations: First, in Theorem 1.3.4 we chained over the $c$, by placing $K \sim n^{1/2}$ grid points along the $c$-axis. Second, in Theorem 1.3.5 we implemented a chaining over both $a$ and $b$, by placing $M \sim n^{(1/4-\eta)\dim x}$ balls of radius $\delta > 0$ to

cover the space spanned by $a$ and $b$. Within the chaining arguments, we handled the temporal dependence structure by decomposing our empirical process $\mathbb{F}_n$ by inserting a $\mathcal{F}_i$-adapted compensator term to create the families of martingales.

We also made extensive use of the exponential martingale inequality, which allowed us to handle the tail probabilities of families of martingales that arose within the chaining argument. The temporal dependence of the indicators with differenced residuals required us to write a variation of Theorem 1.3.11, which is stated in Theorem 1.3.12. We used our own application of the exponential martingale inequality to handle the $\mathcal{F}_i$-adapted families of martingales that arose when we chained over $c$ in Theorem 1.3.4.

The dependence of the differenced residuals also changed the Taylor expansion of the compensator terms which we use to obtain their first-order bias term. Conditional on the filtration $\mathcal{F}_{i-1}$ we had three random variables within the indicator functions which remain random: $\varepsilon_i$, $\varepsilon_{i+1}$ and $x_{i+1}$. In Theorem 1.3.7 we developed a technique to handle the compensator terms, albeit at the cost of Assumption 1.2.3(ii).

There are a number of extensions to the proofs that we did not yet solve. For example, we plan to state Theorem 1.3.1 uniformly in $a, b, c$ instead of just $a, b$. Here the temporal dependence of the indicator functions has proven to be difficult to get around. Another avenue of future research, is to generalise our results to indicator functions involving general moving averages of the residuals. One example for such an algorithm is linear broken trend saturation as discussed in Ericsson (2017) which saturates the regression equation with broken linear trends. These variables are defined as $T_{ik} = i - k + 1$ for $i \geq k$ and zero otherwise. Our current results are already more general than sufficient for the study of the gauge of Step-indicator Saturation which we will turn to in the next chapter.

# Chapter 2

# Gauging the Step-indicator Saturation Algorithm

ABSTRACT. We investigate the asymptotic properties of the gauge of Step-indicator Saturation. The algorithm is used to detect and model location shifts in time series data. We consider a stylized version of Step-indicator Saturation that uses the split-half algorithm. The gauge of Step-indicator Saturation relates to the frequency of falsely detected location shifts when, in fact, the data generating process has no shifts. Our results offer an asymptotic justification to use the gauge to choose the tuning parameter in this statistical procedure.

## 2.1 Introduction

*Step-indicator Saturation* detects and models location shifts which are changes to the previously unconditional mean of time-series data. While we had an illustration of a location shift in Figure 1, more formally, we think of location shifts as non-zero coefficients $\delta_\ell$ in the data generating process,

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

where the $x_i$ and $\varepsilon_i$ are as in model (1.1). The number of location shifts $L$ in the data generating process and the breakpoints $\{\tau_\ell\}_{\ell=0}^{L}$ are typically unknown to the researcher.

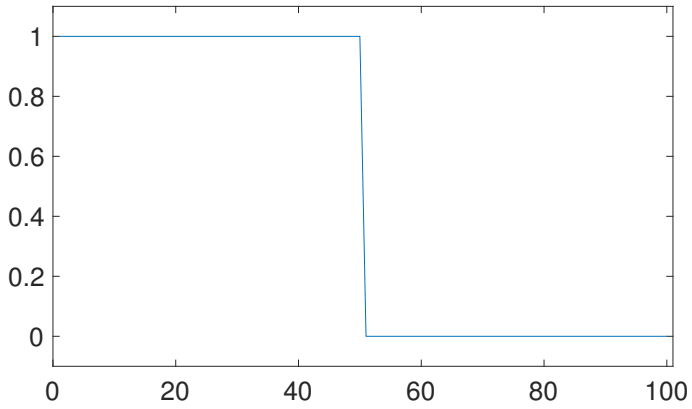



Figure 2.1: An illustration of a step-indicator.

Macroeconomic data are packed with location shifts. Bai et al. (2006) find empirical evidence for location shifts in the US ex-post real interest rate (the three-month treasury bill rate deflated by the CPI inflation rate taken from the Citibase data) in 1966, 1972 and 1980. Hansen (2001) finds empirical evidence of location shifts in US labour productivity in 1982 and 1995. Perron (1989) finds that most US macroeconomic variables can be modelled as trend stationary with two location shifts in 1929 and 1973 which represents the financial crash of the Great Depression and an oil price shock following the OPEC oil embargo.

Unmodelled location shifts give rise to three distinct statistical problems. First, model selection procedures are affected that start from a general unrestricted model and target a parsimonious and encompassing empirical model. Castle and Hendry (2014) document the effect of unmodelled location shifts on selection probabilities of the variables, their lags and non-linear functions. Second, parameter estimation is distorted. Hendry and Mizon (2011) show that an unmodelled location shift caused by the food stamp program of the Great Depression leads to implausible regression coefficients in their model for the US food expenditure. Third, as described in Clements and Hendry (1998), unanticipated location shifts at or near the forecast origin lead to forecast failure.

The way location shifts are detected with Step-indicator Saturation algorithms is by selecting over step-indicators in the model selection stage. We model location shifts with step-indicators. A step-indicator is a variables that has value one up

until a change point, when it switches to value zero. Figure 2.1 illustrates. We create a saturating set of step-indicators which means that we create one step-indicator per observation. This saturating set is then added to the existing set of candidate variables. This regression model cannot be estimated, because it has more explanatory variables than regressors. Yet, by adding the regressors in blocks we can use iterative procedures to eliminate unnecessary regressors, ideally retaining only those which are present in the data generating process.

Because the unnecessary step-indicators can be eliminated using a host of different statistical procedures, a number of variations of Step-indicator Saturation algorithms have been proposed, see Castle et al. (2015). On the two extremes there is the split-half algorithm and Autometrics. The split-half algorithm is the simplest way to implement the elimination of unnecessary step-indicators as it splits the saturating set of step-indicators into just two blocks. In contrast, Autometrics searches through expanding and contracting blocks of regressors and it is robust to a wide range of contamination with location shifts. Due to its robustness, it is the primary implementation of the algorithm that practitioners use. Doornik (2009) contains a more detailed discussion of Autometrics. Another implementation of Step-indicator Saturation which has been popular with practitioners is implemented in the R package *gets* as discussed in Pretis, Reade and Sucarrat (2016). They implement Step-indicator Saturation with a fixed block size, where indicators retained in each block are re-combined and selected over to yield the terminal models.

In this chapter we study the simplest implementation of Step-indicator Saturation that uses the split-half algorithm. For this algorithm, the decision rule to categorise period $i$ as a breakpoint hinges on some preliminary estimator $(\hat{\beta}, \hat{\sigma}^2)$ which is estimated on those observations that are not dummied out by adding half of the saturating set of step-indicators. We consider the test statistic $\sqrt{2}\hat{\sigma}\nabla\hat{\varepsilon}_i = (\nabla y_i - \nabla x_i'\hat{\beta})$, where the symbol $\nabla$ denotes the forward difference, such that $\nabla y_i = y_i - y_{i+1}$. A period $i$ is classified as a breakpoint if $|\nabla\hat{\varepsilon}_i| \geq \sqrt{2}\hat{\sigma}c$, where c is a suitable cut-off value. The variable $c$ is the only tuning parameter of the algorithm and therefore the question arises how to choose the tuning parameter.

We will choose the cut-off $c$ to control the frequency of false positives, which are retained step-indicators when the data generating process contains no location

shifts. The lower the frequency of false positives, the higher the probability that a location that we have identified is indeed a structural break in the process that has generated the data. As the idea of indicator saturation methods is to find breaks and to attribute narratives to each of the identified breaks it is important that the identified break are not spurious. We therefore choose $c$ to use Step-indicator Saturation with confidence.

We formalise the idea of the frequency of false positives by introducing the empirical gauge. When deciding on breakpoints, let $v_i = 0$ denote an observation classified as 'normal' and $v_i = 1$ an observation classified as an 'location shift'. Then, under the null of no location shift in the data generating process, the empirical gauge is defined as the fraction of breakpoints found

$$\hat{\gamma}_n = n^{-1} \sum_{i=1}^{n} v_i.$$

The empirical gauge is similar, but different to the concept of the size used in hypothesis testing, as it is contaminated by estimation error within the decision rule for breakpoints. The prefix 'empirical' is used to distinguish the empirical gauge from the population gauge, with is the asymptotic limit of the empirical gauge.

The history of the concept of the gauge is as follows. In a simulation study of the properties of a general-to-specific algorithm for model selection, Hoover and Perez (1999) considered different measures for the performance of the algorithm that are related to what is now called the gauge. While they referred to one of their measures of the size, it was defined as the number of falsely significant variables divided by the difference between the total number of variables and the number of variables with non-zero coefficients. Hendry and Krolzig (2005); Doornik (2009); Hendry and Doornik (2014) extend the Hoover-Perez idea for model selection in the *PcGets* and *Autometrics* algorithms. In a series of simulation studies, the tuning parameter of these algorithms have been calibrated in terms of the false detection rates for irrelevant regressors. The name 'gauge' was assigned to the false detection rate in Santos and Hendry (2010) and Castle et al. (2011). The first theoretical study of the gauge was in Johansen and Nielsen (2016*b*) who study its properties for the one-step Huber-skip estimator by Huber (1964) and the forward search

algorithm first suggested by Hadi (1992) and developed by Atkinson and Riani (2000).

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

$$\hat{\gamma}_n \xrightarrow{\mathsf{P}} \mathsf{P}(|\nabla\varepsilon_i| \geq \sqrt{2}\sigma c) = \gamma.$$

This establishes an asymptotic relationship between the cut-off value $c$ and the selection probability of an individual step-indicator. We will show that this relationship holds under general assumptions about the regressors. We allow for stationary regressors with arbitrary time-series dependence and also allow for unit-root and trend stationary regressors.

We also analyse the asymptotic distribution of the empirical gauge of Step-indicator Saturation. The asymptotic distribution can be used to understand the likelihood of large deviations of the empirical gauge from its population gauge. We find that in contrast to the analysis in Johansen and Nielsen (2016*b*), where the asymptotic expansion of the empirical gauge depends exclusively on the estimation error $\hat{\sigma} - \sigma$, the asymptotic distribution of the empirical gauge of Step-indicator Saturation depends on both estimation errors $\hat{\sigma} - \sigma$ and $\hat{\beta} - \beta$.

The gauge of step-indicator saturation cannot be studied with the classical asymptotic tool-kit. One of the reasons is that the empirical gauge is a function of the estimation errors $\hat{\beta} - \beta$ and $\hat{\sigma} - \sigma$ which are random variables themselves. We therefore had to develop our own empirical process results of the first chapter which we will use extensively in this second chapter. In fact, we will use all of the results from the first chapter, except of Theorem 1.3.4.

We proceed as follows. First, we explain the Step-indicator Saturation algorithm in section 2.2. Second, we discuss our assumptions in section 2.4. Third, we introduce the concept of the gauge in section 2.3. Fourth, we present the asymptotic results on the empirical gauge of Step-indicator Saturation in section 2.5. Six, we show the results from simulations related to the asymptotic theory in section 2.6. The proofs of this chapter are collected in section 2.7.

## 2.2 Algorithm

Before we explain the individual steps of the split-half algorithm for Step-indicator Saturation that we analyse in this chapter, we take a step back and consider the taxonomy of split-half algorithms. This exercise helps us to understand the differences to Impulse-indicator Saturation and to understand the different variations of Step-indicator Saturation.

### 2.2.1 The taxonomy of split-half algorithms

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

To set up the split-half algorithm we need some notation. We divide the sample in two parts: the first $n_1 = \text{int}(n/2)$ observations and the last $n_2 = n - n_1$ observations. We write the regression model (1.1) conformably in matrix notation

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

We choose a set of dummy variables of the form

$$\begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix},$$

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

We start the split-half algorithm by estimating $\beta$ from the second sample giving the estimators $\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'Y_2$ and $\hat{\sigma}_2^2 = (Y_2 - X_2\hat{\beta}_2)'(Y_2 - X_2\hat{\beta}_2)/n_2$. From this we compute residuals for the first sample, $\hat{\varepsilon}_1 = Y_1 - X_1\hat{\beta}_2$ and regress these on $D_1$ giving $\hat{\delta}_1 = (D_1'D_1)^{-1}D_1'\hat{\varepsilon}_1$, which reduced to $\hat{\delta}_1 = D_1^{-1}\hat{\varepsilon}_1$, since $D_1$ is

invertible. We will now form t-statistics for each of the coefficients in $\hat{\delta}_1$. To do this we compute the variance of $\hat{\delta}_1$ as if the regressors are fixed. This gives

$$\begin{aligned} \mathsf{V}(\hat{\delta}_1|X) &= \mathsf{V}\{D_1^{-1}(Y_1 - X_1\hat{\beta}_2)|X\} \\ &= \mathsf{V}[D_1^{-1}\{\varepsilon_1 - X_1(X_2'X_2)^{-1}X_2'\varepsilon_2\}|X] \\ &= \sigma^2 D_1^{-1}\{I + X_1(X_2'X_2)^{-1}X_1'\}(D_1^{-1})'. \end{aligned} \tag{2.1}$$

Here $\sigma^2$ is estimated by $\hat{\sigma}_2^2$. For each component of $\hat{\delta}_1$ we get the t-statistic

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

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

We swap the role of the two sub-samples, compute the same quantities $\hat{\beta}_1$, $\hat{\sigma}_1^2$ to get $\mathsf{V}(\hat{\delta}_2|X)$ and $v_{2,\ell}$, which are defined as before. We select those dummy variables where $|v_{2,\ell}| > c$. This concludes the split-half procedure.

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

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

Their inverses, used for the computation of $\hat{\delta}_i$, are

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

Both set of step-indicators $D_\nabla$ and $D_\Delta$ span the same space. However, they represent re-parametrizations. Either choice results in a single element of $\hat{\delta}_2$ that is computed differently from the remaining ones. For the choice of $D_2 = D_\nabla$, say, the estimate of the last dummy in half sample is $\hat{\delta}_{2,n} = \hat{\varepsilon}_n$ while it is $\hat{\delta}_{2,\ell} = \hat{\varepsilon}_\ell - \hat{\varepsilon}_{\ell+1}$ for the previous dummies. The variance of $\hat{\delta}_{2,\ell}$ changes accordingly from $\mathsf{V}(\hat{\delta}_{2,\ell}|X) =$

$\sigma^2\{2 + \nabla x'_{2,\ell}(X'_1X_1)^{-1}\nabla x_{2,\ell}\}$ for $n_1 + 1 \leq \ell \leq n - 1$ to $\mathsf{V}(\hat{\delta}_{2,n}|X) = \sigma^2\{1 + x'_{2,n}(X'_1X_1)^{-1}x_{2,n}\}$ for the final dummy. In the case of $D_2 = D_\Delta$, the first dummy is estimated differently as $\hat{\delta}_{2,1} = \hat{\varepsilon}_{n_1+1}$ while the latter ones are estimated as $\hat{\delta}_{2,\ell} = \hat{\varepsilon}_{2,\ell} - \hat{\varepsilon}_{2,\ell-1}$. For the variance of $\hat{\delta}_{2,\ell}$ we get $\mathsf{V}(\hat{\delta}_{2,1}|X) = \sigma^2\{1+x'_{2,1}(X'_1X_1)^{-1}x_{2,1}\}$ for the first dummy and $\mathsf{V}(\hat{\delta}_{2,\ell}|X) = \sigma^2\{2 + \nabla x'_{2,\ell}(X'_1X_1)^{-1}\nabla x_{2,\ell}\}$ for the subsequent ones. Both $D_\nabla$ and $D_\Delta$ are valid choices for $D_i$. The choice on the $D_i$ depends on the application. If the interest is to have a consistent decision rule at the end of the sample, say to detect a location shift near the forecast origin to avoid forecast failure, then a consistent decision rule to the previous observations might be appropriate. In this case, $D_\Delta$ may be preferable. In practice, both $D_\Delta$ and $D_\nabla$ are implement in the software packages. The implementation of Step-indicator Saturation in Autometrics uses $D_\Delta$. In contrast, $D_\nabla$ is used in the R package *gets*.

In Castle et al. (2015) the set of dummy variables is of the form

$$\begin{pmatrix} D_\nabla & F \\ 0 & D_\nabla \end{pmatrix} \tag{2.2}$$

where $F$ is a matrix of ones. They arrive at matrix (2.2) by saturating the regression equation with $n$ step-indicators which form the set $\mathcal{S} = \{1_{(i\leq j)}, j = 1, ..., n\}$. The split-half algorithm with this variation of dummy variables faces multicollinearity of the last step-indicator with the intercept, as the last column of matrix (2.2) is a column of ones. There are two ways to address the multicollinearity. In the first, the final step-indicator is dropped. The second is to modify the final step-indicator to be $1_{(n_1<i\leq n)}$ instead. However, this unnecessarily complicates the asymptotic theory.

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

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

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

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

and

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

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

### 2.2.2 The split-half algorithm with step-indicators

Given the general discussion of the split-half algorithm, in this section we will proceed to spell out the individual steps of the split-half algorithm with Step-indicator Saturation in greater detail. Because we develop asymptotic theory and therefore consider large sample sizes, we will ignore the varying decision rules on the boundaries of the sub-samples.

**Algorithm 2.2.1.** *The split-half step-indicator saturation algorithm.*

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

2. *Split the data into two sets $I_1$ and $I_2$ of $n_1$ and $n_2$ observations.*

3. *Calculate the least squares estimators for $(\beta, \sigma^2)$ based on sample $I_j$ as*

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

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

$$v_i = 1_{(i \in I_1)} \frac{(\nabla y_i - \hat{\beta}_2' \nabla x_i)}{\hat{\omega}_{2,i}} + 1_{(i \in I_2)} \frac{(\nabla y_i - \hat{\beta}_1' \nabla x_i)}{\hat{\omega}_{1,i}} \tag{2.3}$$

   *where $\hat{\omega}_{j,i}^2$ could be chosen as in (2.1) as $\hat{\sigma}_j^2 \{2 + \nabla x_i' (\sum_{k \in I_j} x_k x_k')^{-1} \nabla x_i\}$ for fixed regressors.*

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

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

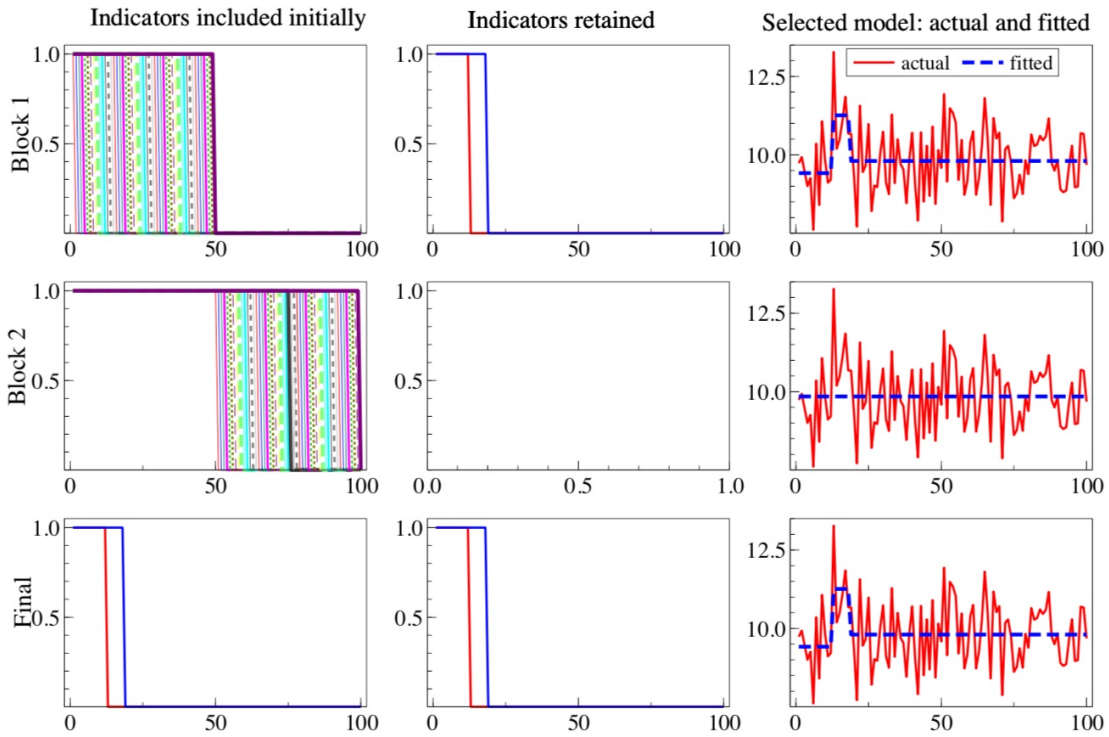



Figure 2.2: An illustration of the split-half algorithm with Step-indicator Saturation when there is no break in the data generating process

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

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

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

In Figure 2.2 we present a graphical illustration of the split-half algorithm for Step-indicator Saturation. The figure was first presented in Castle et al. (2015). The Algorithm 2.2.1 concerns the first two rows on the plots. We do not consider a reselection step that is performed after the split-half algorithm, which is illustrated in the third row. To explain the figure, we discuss it row by row. In the first row we consider adding half of the step-indicators to a regression equation. That is equivalent to dummying out the first half of observations. In this example the

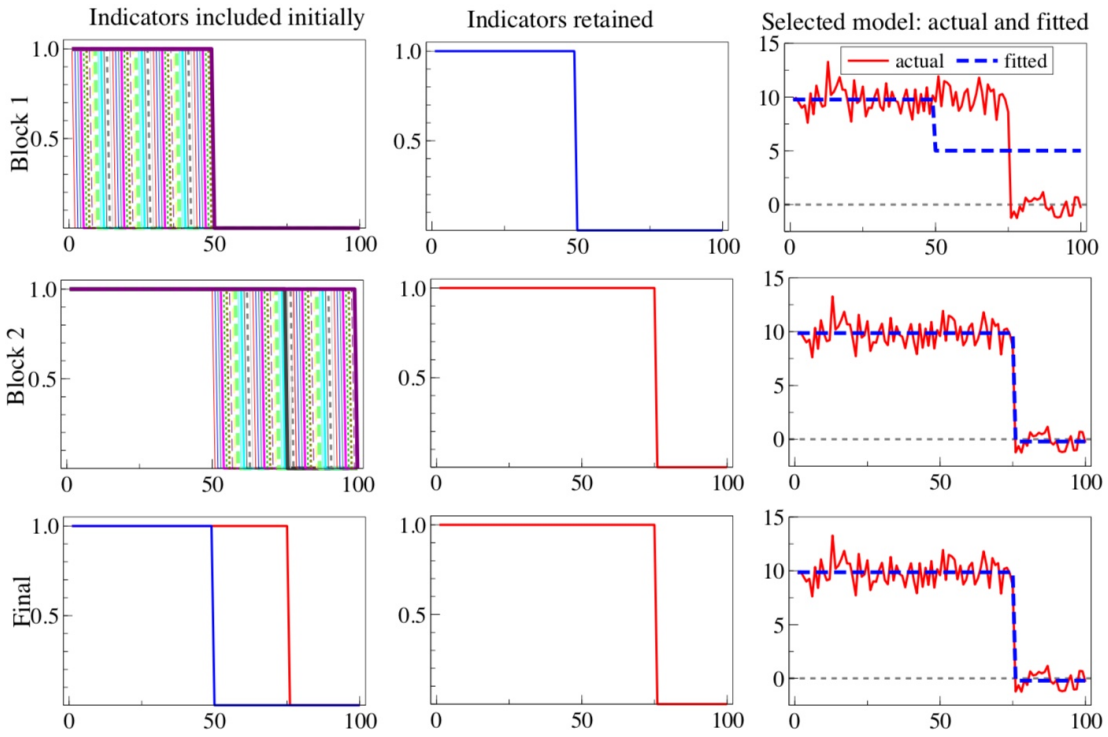



Figure 2.3: An illustration of the split-half algorithm with Step-indicator Saturation when there is a break in the data generating process in period 75

only regressor is the intercept $\mu$ and it is constant over time as we consider the null that it does not contain any location shifts. The parameters $\hat{\mu}$ and $\hat{\sigma}$ are estimated on the second half sample. Based on these estimates we compute the t-ratios as in (2.3). In this example, we have two observations for which the t-ratio is larger than the cut-off value $c$ and we thus select two step-indicators. This is illustrated in the second column. The third column displays the data and the mean that is fitted with the two retained step-indicators. In the second row, we repeat this procedures with the second half of the saturating set of step-indicators added. In this example, none of the t-ratios exceed the cut-off value $c$ and no step-indicators are retained for the second block of step-indicators. Therefore, as the data generating process did not contain any location shifts, but as we have retained two step-indicators, we have two false positives. In the third row, we reselect over the previously selected step-indicators. In this example this does not change anything as both of them are still retained.

In Figure 2.3 we present an analogous example under the alternative of a location shift. As before, the only regressor is the intercept, yet it changes in period 75 from $\mu_0 = 10$ to $\mu_1 = 0$. This example aims to illustrate how Step-indicator Saturation is able to pick up location shifts. Again we discuss row by row. In the first row, we add the first block of step-indicators to the regression equation. The mean that we estimate over the second subsample will be in-between $\mu_0$ and $\mu_1$. The t-ratio of the final observation in the first subsample that is added is then highly significant and we retain this step-indicator. In the second row, we select the step-indicator during the period of the break. As Algorithm 2.2.1 only considers the first two columns, we would have correctly modelled the change in the mean at the cost of one false positive. A reselection step, which is illustrated in the third column, would have eliminated the false positive in our example.

As discussed in section 2.2.1, Step-indicator Saturation is related to Impulse-indicator Saturation where we extend the feature space with impulse-indicators instead of step-indicators. Because we run simulations in section 2.6, where we compare Step-indicator Saturation to Impulse-indicator Saturation, we state for the sake of completeness the split-half algorithm for Impulse-indicator Saturation.

**Algorithm 2.2.2.** *The split-half Impulse-indicator Saturation algorithm. Step 1 to 3 are as in Algorithm 2.2.1.*

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

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

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

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

## 2.3 The gauge

Before we introduce the gauge it is useful to consider model selection in a stylised regression model with $K$ orthogonal regressors. In this case, for each regressor a t-statistic can be computed based on $t_j = (\text{se}(\hat{\beta}_j))^{-1} \hat{\beta}_j$. Asymptotically, each t-statistic will be standard normal distributed and independent from each other.

We consider the null model where all $K$ regressors are irrelevant. The t-tests are at a significance level $\alpha$. We retain a variable if t-statistic is greater than the critical value $|t_j| > c_\alpha$. As discussed on page 104 of Hendry and Doornik (2014), the average number of irrelevant variables retained is then

$$k = \sum_{i=0}^{K} i \frac{K!}{i!(K-i)!} \alpha^i (1-\alpha)^{K-i} = K\alpha.$$

Correspondingly, the average frequency of irrelevant variables being retained is $k/K = \alpha$. The size of the selection procedure however, is much larger, as it is $1-(1-\alpha)^N$; it is the probability that any regressor is retained.

The empirical gauge is defined similarly to the average frequency of irrelevant variables being retained in the model selection context. Specifically, we define it for step-indicator saturation as

$$\hat{\gamma}_n(c) = n^{-1} \sum_{i=1}^{n} 1_{(|v_i|>c)}, \tag{2.4}$$

where $v_i$ is defined in equation (2.3) of Algorithm 2.2.1. If we ignore the split-half algorithm for the moment and leave open the source of the initial least squares estimates for $\hat{\beta}$ and $\hat{\sigma}$, we can write out the gauge as a function of the estimation error of the location parameter $\beta$ and the scale parameter $\sigma$ as

$$\hat{\gamma}_n(c) = n^{-1} \sum_{i=1}^{n} 1_{\{|\nabla\varepsilon_i - \nabla x_i'(\hat{\beta}-\beta)| \leq \sqrt{2}\sigma c + \sqrt{2}(\hat{\sigma}-\sigma)c\}}.$$

We will show that despite the dependence of the individual indicator functions deriving from the $\nabla\varepsilon_i$ and the estimation error of $\hat{\beta}$ and $\hat{\sigma}$, the empirical gauge still converges to the population gauge, which we define as $\mathsf{P}(|\nabla\tilde{\varepsilon}_i| > c_\alpha) = \alpha$ and which is equivalent to the average frequency of irrelevant variables being retained in the model selection context. Another perspective to see why the example of orthogonal regressors discussed in Hendry and Doornik (2014) does not apply for the empirical gauge of Step-indicator Saturation is because we have more regressors than observations, as we have added one step-indicator per observation in addition to the original regressors $x_i$. Therefore the regressors cannot be transformed into an orthogonal set of regressors.

Understanding the asymptotic theory of the empirical gauge is important, because the only tuning parameter of Step-indicator Saturation is related to it: We set the cut-off value $c$ to target a population gauge and the asymptotic results relate both versions of the gauge. For example, if we choose $c_\alpha$ such that $\alpha = 0.01$ this corresponds with a sample size of $n = 100$ to one false positive that we would expect on average. A low number of false positives are important to have high confidence that the detected location shifts are breaks in the data generating process and not statistical artefacts. This is of particular importance to the research agenda to attribute location shifts to historical events and policy changes.

While we value a large $c$ to reduce the errors of the first type where we find a location shift while the data generating process had a constant mean, a larger $c$ will come with the trade-off that the power of the algorithm to find location shifts in the data generating process will decrease. Instead of optimising between the frequency of false positives and the frequency to find true positives, in Step-indicator Saturation we calibrate the algorithm only with the frequency of false positives in mind. The idea is to calibrate the algorithms to have the same population gauge and possibly the same variance around the population gauge, before comparing their frequency to retain true positives.

A related but distinct idea to the gauge is the false discovery rate for multiple tests of Benjamini and Hochberg (1995). Consider the testing of $m$ hypothesis. The false discovery rate concerns those tests that reject the hypothesis. Say $R$ of these exists, of which $V$ are false rejections. Then the false discovery rate is $\mathsf{E}(V/R)$. In contrast, the gauge concerns those observations that don't contain a location shift. Under the null, there no location shift in the data generating process, so $m = n$. The gauge is then the share of the number of falsely declared location shifts $V$ and the number of hypothesis $m$: $\mathsf{E}(V/m)$.

## 2.4 Assumptions

In this section, we present a set of assumptions for the asymptotic theory of the gauge of Step-indicator Saturation. When we use the Step-indicator Saturation estimator, we assume the density $\mathsf{f}$ of $\sigma^{-1}\varepsilon_i$ is known. In most applications, $\mathsf{f}$

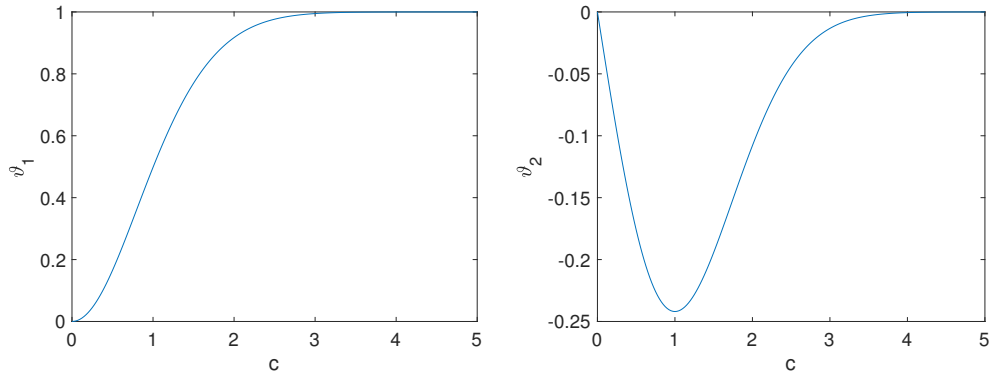



Figure 2.4: An numerical approximation of $\vartheta_1$ and $\vartheta_2$ for a standard normal $\mathsf{f}$ and varying $c$

will be the normal density. But our assumptions allow for a range of alternative densities such as the $t$-density. In fact, we could even use a $\mathsf{f}$ that is non-symmetric.

Yet, before we discuss the assumptions, we state a couple of definitions of densities, distribution functions and moments. Recall the normalisations for $\varepsilon_i$ and $x_i$ that we have introduced in the first chapter, where $\check{\varepsilon}_i = \sigma^{-1}\varepsilon_i$ and $\tilde{\varepsilon}_i = (2\sigma^2)^{-1/2}\varepsilon_i$. We will use this notation throughout this chapter.

We repeat a number of definitions for the densities, distributions and moments of our random variables. The $\nabla\tilde{\varepsilon}_i$ have density and distribution function

$$\mathsf{h}(x) = \sqrt{2}\int_{-\infty}^{\infty} \mathsf{f}(y)\mathsf{f}(\sqrt{2}x + y)\,\mathrm{d}y, \quad \mathsf{H}(x) = \mathsf{P}(\nabla\tilde{\varepsilon}_i \leq c).$$

and similarly, $|\nabla\tilde{\varepsilon}_i|$ has distribution function

$$\psi = 1 - \gamma = \mathsf{G}(c) = \mathsf{P}\left(|\tilde{\varepsilon}_i - \tilde{\varepsilon}_{i+1}| < c\right) = \int_{-c}^{c} \mathsf{h}(x)dx.$$

Note that the properties of $\mathsf{h}$ derive from the properties of $\mathsf{f}$ as in Theorem 1.2.1. The conditional joint density $\mathsf{m}_i(y, x)$ given $\mathcal{F}_{i-1}$ is the density of the two random variables $\nabla\tilde{\varepsilon}_i$ and $\nabla\tilde{x}_i$. Let $\mathsf{k}_i(x)$ be the conditional marginal density of the regressors $\nabla x_i$ conditional on $\mathcal{F}_{i-1}$. Then the conditional density $\mathsf{m}_i(x|y)$ given $\mathcal{F}_{i-1}$ and the derivative of the conditional density $\mathsf{m}_i(y|x)$ given $\mathcal{F}_{i-1}$ with respect to the differenced error term, if it exists, is

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

Table 2.1: Numerical approximations of $\vartheta_1$ and $\vartheta_2$ for varying $\gamma$

| $\gamma$ | 5% | 2.5% | 1% | 0.1% |
|---|---|---|---|---|
| $\vartheta_1$ | 0.909 | 0.954 | 0.981 | 0.998 |
| $\vartheta_2$ | -0.115 | -0.073 | -0.037 | -0.006 |

Finally, define the third and fourth moment of the scaled innovations $\check{\varepsilon}_i$ as

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

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

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

$$\begin{aligned}
\vartheta_1 &= \mathsf{E}\{1_{(|\nabla\check{\varepsilon}_i|\le c)} 1_{(|\nabla\check{\varepsilon}_{i+1}|\le c)}\} \\
&= \int_{-\sqrt{2}c}^{\sqrt{2}c} \int_{-\sqrt{2}c}^{\sqrt{2}c} \int_{-\infty}^{\infty} \mathsf{f}(x+y+z)\mathsf{f}(y+z)\mathsf{f}(z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x, \qquad (2.5) \\
\vartheta_2 &= \mathsf{E}\{1_{(|\nabla\check{\varepsilon}_i|\le c)}(\check{\varepsilon}_{i+1}^2 - 1)\} = \mathsf{E}\{1_{(|\nabla\check{\varepsilon}_i|\le c)}(\check{\varepsilon}_i^2 - 1)\} \\
&= \int_{-\infty}^{\infty} \int_{-\sqrt{2}c}^{\sqrt{2}c} (x^2 - 1)\mathsf{f}(x)\mathsf{f}(y+x)\,\mathrm{d}y\,\mathrm{d}x, \qquad (2.6) \\
\vartheta_3 &= \mathsf{E}\{1_{(|\nabla\check{\varepsilon}_i|\le c)} x_{i+1}\check{\varepsilon}_{i+1}\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\sqrt{2}c}^{\sqrt{2}c} xy\mathsf{f}(y)\mathsf{k}(x, z+y)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x. \qquad (2.7) \\
\vartheta_4 &= \mathsf{E}\{\check{\varepsilon}_i 1_{\{|\nabla\check{\varepsilon}_i|\le c\}}\} = \int_{-\infty}^{\infty} \int_{-\sqrt{2}c}^{\sqrt{2}c} x\mathsf{f}(x)\mathsf{f}(x-y)dydx \qquad (2.8)
\end{aligned}$$

We plot $\vartheta_1$ and $\vartheta_2$ for a standard normal density $\mathsf{f}$ in Figure 2.4. The values for $\vartheta_3$ depend on the regressors $x_i$. Note that $\vartheta_3$ is a vector of dimension $\dim x$. In the special case of a first-order autoregression $y_i = \mu + \alpha_1 y_{i-1} + \varepsilon_i$ we find $\vartheta_3 = -\sigma\vartheta_2$. The value of $\vartheta_4$ is zero, for all symmetric densities $\mathsf{f}$, because then

$$\int_{-\sqrt{2}c}^{\sqrt{2}c} \mathsf{f}(x)\mathsf{f}(x-y)dy = \int_{-\sqrt{2}c}^{\sqrt{2}c} \mathsf{f}(-x)\mathsf{f}(-x-y)dy.$$

As in the first chapter, the Assumption 1.2.2 is used throughout the second chapter. We therefore have the same properties regarding the filtration over the $\varepsilon_i$ and $x_i$ and the basic characteristics of the density function of the error terms

$\mathsf{f}$. The next set of assumptions strengthens of the assumptions in section 1.2 Chapter 1. The generality has two sources: first, we aim to make the assumptions more tractable for applied researchers and second, the proofs require stronger assumptions on moments of $\varepsilon_i$ and $x_i$.

**Assumption 2.4.1.** *For some values of* $0 \leq \kappa < \eta \leq 1/4$*, suppose*
(*i*) *the density* $\mathsf{f}$ *satisfies*
  (*a*) *for* $c > 0$, $c\mathsf{f}(c)$, $|\dot{\mathsf{f}}(c)|$ *are decreasing large* $c$ *and increasing for small* $c$*;*
  (*b*) $\int_{-\infty}^{\infty} u\mathsf{f}(u)du = 0$ *and* $\int_{-\infty}^{\infty} |u^{2+\phi}|\mathsf{f}(u)du < \infty$*, for some small* $\phi > 0$*;*
(*ii*) *the conditional density* $\mathsf{m}_i(y|x)$ *given* $\mathcal{F}_{i-1}$*, see* (1.8)*, exists and satisfies* $\max_{1\leq i\leq n} \sup_{y\in\mathbb{R},x\in\mathbb{R}^p} |\dot{\mathsf{m}}_i(y|x)| < \infty$*;*
(*iii*) *the regressors* $x_i$ *and the normalisation matrix* $N$ *satisfy*
  (*a*) $\sum_{i=1}^{n} Nx_ix_i'N \xrightarrow{\mathsf{D}} \Sigma \overset{a.s.}{>} 0$*;*
  (*b*) $n^{-1}\sum_{i=1}^{n} \mathsf{E}|n^{1/2}N\nabla x_i^{2+\phi}| < \infty$*, for some small* $\phi > 0$*;*
(*iv*) *the estimators* $(\hat{\beta}, \hat{\sigma}^2, \hat{\omega}_i)$ *satisfy*
  (*a*) $N^{-1}(\hat{\beta} - \beta) = \mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$*;*
  (*b*) $n^{1/2}(\hat{\sigma}^2 - \sigma^2) = \mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$*;*
  (*c*) $\max_{1\leq i\leq n} n^{1/2}|\hat{\omega}_i - \hat{\sigma}| = \mathrm{o}_{\mathsf{P}}(1)$.

**Remark 2.4.1.** *The conditions* (*i*) *and* (*iii*) *are satisfied in a wide range of models, as discussed in Johansen and Nielsen (2013, 2016*a,b*). The normal and t-distribution both satisfy condition* (*i*) *on the density. Stationary, random walk and deterministically trending regressors satisfy condition* (*iii*)*. Condition* (*iv*) *allows the standardised estimation errors to diverge at the rate of* $n^{1/4-\eta}$ *instead of being bounded in probability. By choosing* $\eta = 1/4$*, we would consider estimators with the standard rate of convergence.*

## 2.5 Asymptotic results

The results of this chapter relate to the empirical gauge of split-half Step-indicator Saturation. Because we analyse the empirical gauge, we only need to consider the steps 1 to 5 in Algorithm 2.2.1. The first theorem states that the empirical gauge converges in probability to the population gauge $\gamma = \mathsf{P}(|\nabla\tilde{\varepsilon}_i| \geq c)$. The second

theorem provides an asymptotic distribution result of the empirical gauge. The proofs are given in the section 2.7.

**Theorem 2.5.1.** *Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2.1. Suppose Assumptions 1.2.2 and 2.4.1(i,iii,iv) hold. Then, for fixed $c \in \mathbb{R}$, we have $\hat{\gamma}_n(c) \xrightarrow{\mathsf{P}} \gamma$.*

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

This theorem is used for the calibration of the Step-indicator Saturation algorithm. The unique tuning parameter of the algorithm is $c$. The theorem provides the rational to choose $c$ to match the desired population gauge. The empirical gauge $\hat{\gamma}_n$ converges to $\gamma$ which in turn depends on $c$ by (1.7). Thus, we first specify our tolerance for false positives expressed by $\gamma$, which given our assumed distribution of the error term, results in a selection quantile $c$. For example, let's assume a standard normal reference distribution $\mathsf{f}$ and a sample size $n = 100$. When we tolerate on average 1 falsely detected location shift, then we want the gauge to be $\hat{\gamma}_{100} = 1\%$. Thus we choose $\psi = 0.99$ and $c_{0.99} = 2.58$.

We present results related to the asymptotic distribution of the gauge. These results are organised as follows. We first state in Theorem 2.5.2 an asymptotic expansion of $\sqrt{n}\{\hat{\gamma}_n(c) - \gamma(c)\}$. This asymptotic expansion is used to derive the asymptotic distribution of the gauge, as the terms of this expansion satisfy the conditions of a mixingale (see Theorem 2.7.4). The explicit asymptotic variance terms are then presented for stationary regressors in Theorem 2.5.3 and for unit-root and trend stationary regressors in Theorem 2.5.4.

Consider limits of the first two normalised sample moments

$$n^{-1/2} \sum_{i=1}^{n} N' x_i \xrightarrow{\mathsf{D}} \mu, \quad \sum_{i=1}^{n} N' x_i x_i' N \xrightarrow{\mathsf{D}} \Sigma, \tag{2.9}$$

and let $\Sigma_j$ with $j = 1, 2$ be the subsample analogue of $\Sigma$. Similarly, let $N_j$ be the subsample analogue of $N$ which we have introduced in the first chapter as

the normalisation for $x_i$ that depends on whether the regressors are stationary, random walk or trend stationary. For example, we have

$$N_j = \begin{pmatrix} n_j^{-1/2} & 0 \\ 0 & n_j^{-3/2} \end{pmatrix} \text{ when } x_i = \begin{pmatrix} 1 \\ i \end{pmatrix}. \tag{2.10}$$

Recall the definition of $\xi_c$ from (1.12). Define subsample analogues of $\xi_c$, such that $\xi_{cj} = \xi_j(c) - \xi_j(-c)$ with $j = 1, 2$ and

$$n_j^{-1/2} \sum_{i \in I_j} \mathsf{E}_{i-1}(N_j \nabla \tilde{x}_i | \nabla \tilde{\varepsilon}_i = c) \xrightarrow{\mathsf{D}} \xi_j(c).$$

Further, introduce variables that emerge due to the split-half approach as

$$H_i = n_1^{1/2} n_2^{-1/2} 1_{(i \in I_2)} + n_2^{1/2} n_1^{-1/2} 1_{(i \in I_1)}, \tag{2.11}$$

$$K_i = n_1^{1/2} \xi_{c1}' \Sigma_2^{-1} N_2 1_{(i \in I_2)} + n_2^{1/2} \xi_{c2}' \Sigma_1^{-1} N_1 1_{(i \in I_1)}. \tag{2.12}$$

In the special case when the two sub-samples satisfy $n_1 = n_2$ and where the regressors are stationary, we have $H_i = 1$ and $K_i = \xi_c' \Sigma^{-1}$ as in this case $\Sigma_j = \Sigma$ and $\xi_{cj} = \xi_c$.

**Theorem 2.5.2.** *Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2.1. Suppose the Assumptions 1.2.2 and 2.4.1(i,ii,iiib,iva) hold. Then, we get for fixed $c \in \mathbb{R}$ that*

$$\begin{aligned} n^{1/2}\{\hat{\gamma}_n(c) - \gamma(c)\} = &-n^{1/2}\{\widehat{G}_n(0,0,c) - \psi\} \\ &- \mathsf{h}(c) n^{-1/2} \sum_{i=1}^{n} \left\{ \left( \check{\varepsilon}_i^2 - 1 \right) c H_i + K_i x_i \varepsilon_i \right\} + \mathrm{o}_{\mathsf{P}}(1). \end{aligned}$$

The next theorem uses the asymptotic expansion in Theorem 2.5.2 to derive the asymptotic distribution of the gauge with stationary regressors. For the asymptotic variance, we will use Theorem 2.7.3. Our terms simplify somewhat due to the assumed stationarity of the regressors $x_i$.

Recall that the expressions for $\vartheta_1$ to $\vartheta_4$ are given in equation (2.5) to (2.8). We also tabulated the resulting values for $\vartheta_1$ and $\vartheta_2$ for a standard normal density $\mathsf{f}$ in Table 2.1.

**Theorem 2.5.3.** *Let the regressors $x_i$ be stationary and let $n_j/n \to \lambda_j > 0$ for $j = 1, 2$. Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2.1. Suppose the Assumptions 1.2.2 and 2.4.1 hold. Then, for fixed $c \in \mathbb{R}$ and*

$$\tilde{\lambda} = (\lambda_1 \lambda_2)^{1/2} \tag{2.13}$$

*we have $\sqrt{n}\{\hat{\gamma}_n(c) - \gamma\} \xrightarrow{\mathsf{D}} \mathsf{N}\{0, \gamma(1-\gamma) + 2(\vartheta_1 - \psi^2) + \omega_* + \omega_{**}\}$, where*

$$\begin{aligned}
\omega_* &= c\mathsf{h}(c)\{c\mathsf{h}(c)(\varkappa_4 - 1) + 8\tilde{\lambda}\vartheta_2\}, \\
\omega_{**} &= \mathsf{h}(c)[4\sigma\tilde{\lambda}\xi_c'\Sigma^{-1}\mu\{\vartheta_4 + c\mathsf{h}(c)\varkappa_3\} + \sigma^2\mathsf{h}(c)\xi_c'\Sigma^{-1}\xi_c + 4\sigma\tilde{\lambda}\xi_c'\Sigma^{-1}\vartheta_3].
\end{aligned}$$

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

**Theorem 2.5.4.** *Let the regressors $x_i$ be random walk or trend stationary regressors, i.e. consider the models $y_i = y_{i-1} + \varepsilon_i$ or $Y_i = AY_{i-1} + \Phi d_{i-1} + e_i$ with*

$$Y_{i-1} = \begin{pmatrix} y_{i-1} \\ \vdots \\ y_{i-k} \end{pmatrix}, \quad A = \begin{pmatrix} (\delta_1, \ldots, \delta_{k-1}) & \delta_k \\ I_{k-1} & 0 \end{pmatrix}, \quad \Phi = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}, \quad e_i = \begin{pmatrix} \varepsilon_i \\ 0 \end{pmatrix}$$

*where $d_i \in \mathbb{R}^\ell$ are deterministic terms and where $|eigen(A)| < 1$. The deterministic component satisfies $d_i = Dd_{i-1}$ with $|eigen(D)| = 1$ and $rank(d_1, \ldots, d_\ell) = \ell$. Let $n_j/n \to \lambda_j > 0$ for $j = 1, 2$. Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2.1. Suppose the Assumptions 1.2.2 and 2.4.1 hold. Then, for fixed $c \in \mathbb{R}$ we have $n^{1/2}\{\hat{\gamma}_n(c) - \gamma\} \xrightarrow{\mathsf{D}} \mathsf{N}\{0, \gamma(1-\gamma) + 2(\vartheta_1 - \psi^2) + \omega\}$ where $\omega = c\mathsf{h}(c)\{c\mathsf{h}(c)(\varkappa_4 - 1) + 8\tilde{\lambda}\vartheta_2\}$.*

**Example 2.5.1.** *We consider the asymptotic distribution of the gauge for standard normally distributed error terms $\check{\varepsilon}_i$. Further assume that the sample size in the*

*two sub-samples is equal and that the regressors $x_i$ are stationary. The asymptotic distribution of the empirical gauge then simplifies to*

$$n^{1/2}\{\hat{\gamma}_n^{sn}(c) - \gamma\} \xrightarrow{\mathsf{D}} \mathsf{N}\{0, \gamma(1-\gamma) + 2(\vartheta_1 - \psi^2) + \omega_*^{sn} + \omega_{**}^{sn}\} \tag{2.14}$$

*with* $\omega_*^{sn} = 2c\mathsf{h}(c)\{c\mathsf{h}(c) + 2\vartheta_2\}$ *and* $\omega_{**}^{sn} = \mathsf{h}^2(c)\xi_c'\Sigma^{-1}\xi_c + 2\mathsf{h}(c)\xi_c'\Sigma^{-1}\vartheta_3$. *Here we have used that* $\varkappa_3 = 0$, $\varkappa_4 = 3$ *and* $\vartheta_4 = 0$ *for the standard normal density* $\mathsf{f}$ *and that* $\tilde{\lambda} = 0.5$ *for split-half with equal sample sizes in sub-samples. The asymptotic variance simplifies further for exogenous regressors* $x_i$, *as in that case* $\xi_c = \omega_{**}^{sn} = 0$ *and thus*

$$n^{1/2}\{\hat{\gamma}_n^{exo}(c) - \gamma\} \xrightarrow{\mathsf{D}} \mathsf{N}\{0, \gamma(1-\gamma) + 2(\vartheta_1 - \psi^2) + \omega_*^{sn}\}$$

*Finally, in the special case of a first order autoregression* $y_i = \mu + \alpha_1 y_{i-1} + \varepsilon_i$, *we have*

$$n^{1/2}\{\hat{\gamma}_n^{ar}(c) - \gamma\} \xrightarrow{\mathsf{D}} \mathsf{N}\{0, \gamma(1-\gamma) + 2(\vartheta_1 - \psi^2) + \omega_*^{ar} + \omega_{**}^{ar}\},$$

*where* $\omega_*^{ar} = 2c\mathsf{h}(c)\{c\mathsf{h}(c) + 2\vartheta_2\}$ *and* $\omega_{**}^{ar} = 4c\mathsf{h}(c)(\alpha_1 - \alpha_1^3)\{\alpha_1 c\mathsf{h}(c) + \vartheta_2\}$.

## 2.6 Simulation results

We illustrate the asymptotic results using simulations which have been coded in MATLAB. We replicate each simulation $M = 10,000$ times. For the first-order autoregression, the observation $y_0$ is on the long-run mean of the time series.

### 2.6.1 Simulations on the convergence

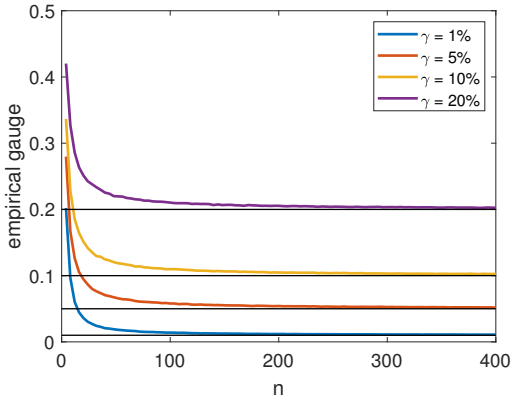
Consider first the simulations on the convergence of the gauge of Step-indicator Saturation and the finite sample properties of Step-indicator Saturation. In each of the simulations, we run Algorithm 2.2.1 with $n_1 = n_2$.

In Figure 2.5a we plot the empirical gauge for an increasing sample size and different selection probabilities $\gamma$, to analyse the small sample properties. The DGP is $y_i = \varepsilon_i$ where $\varepsilon$ has standard normal density $\mathsf{f}$. The sample size varies from 10 to 300 and we increase the sample size one-by-one. Each time we increase the sample size, we redraw all $n$ observations. We run Algorithm 2.2.1 on the regression equation

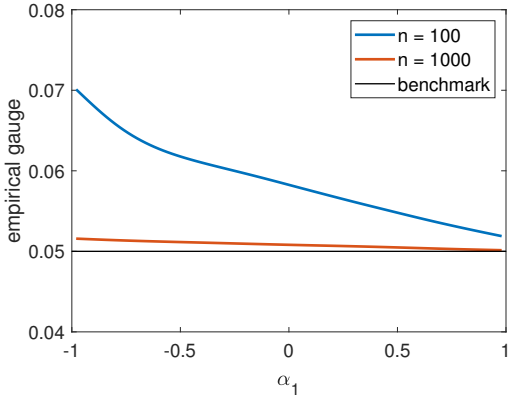
$$y_i = \beta x_i + \varepsilon_i, \tag{2.15}$$

Figure 2.5: Simulations on the small sample bias of the gauge

(a) Bias with varying $n$ and $\gamma$




(b) Bias with varying $\alpha_1$ and $n$



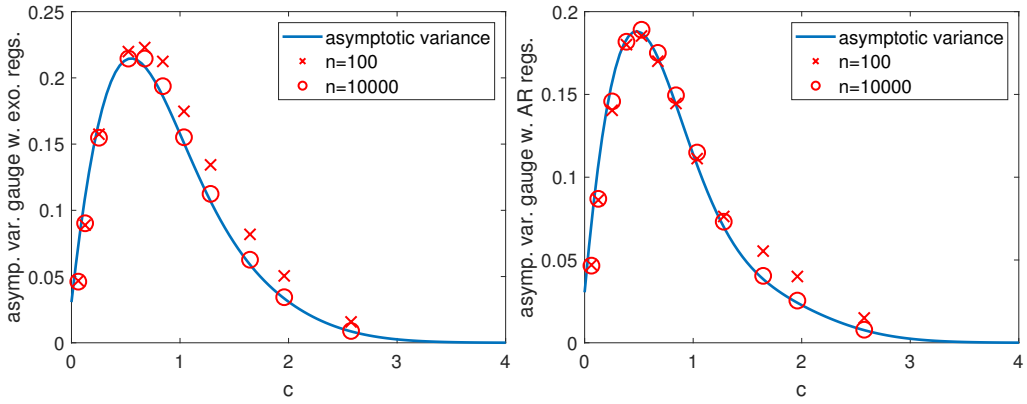

for a strictly exogenous one-dimensional regressor $x_i$ which are i.i.d. draws from a standard normal distribution. We observe that the small sample bias of the gauge is positive. On average we find more false positives if the sample sizes are small. For example, if the sample size is limited to $n = 20$ observations and when we select at a 5% level then on average we are retaining around 10% of observations. The empirical gauge is double the population gauge. However, the bias vanishes quickly with growing samples sizes. Moreover, we observe that the small sample bias depends on the selection quantile $c$. In absolute value, the small sample bias is larger for bigger selection probabilities $\gamma$.

Figure 2.5b shows the relationship between the first-order autoregressive coefficient and the small sample bias of the gauge. The DGP is

$$y_i = \alpha_1 y_{i-1} + \varepsilon_i, \tag{2.16}$$

where $-1 < \alpha_1 < 1$. We increase the coefficients by 0.1 in every step. The errors are standard normal. We run Algorithm 2.2.1 on the correctly specified first-order autoregressive model for two sample sizes $n = 100$ and $n = 1000$. In each iteration of the simulation, we redraw all observations. The figure illustrates that for constant $n$, the small sample bias increases for negative $\alpha_1$ and is the largest for $\alpha_1$ close to $-1$. In contrast, a positive $\alpha_1$ reduces the small sample bias. However, the small sample bias vanishes asymptotically, as predicted by the consistency of the gauge.

Figure 2.6: Simulations for the asymptotic variance with varying $c$

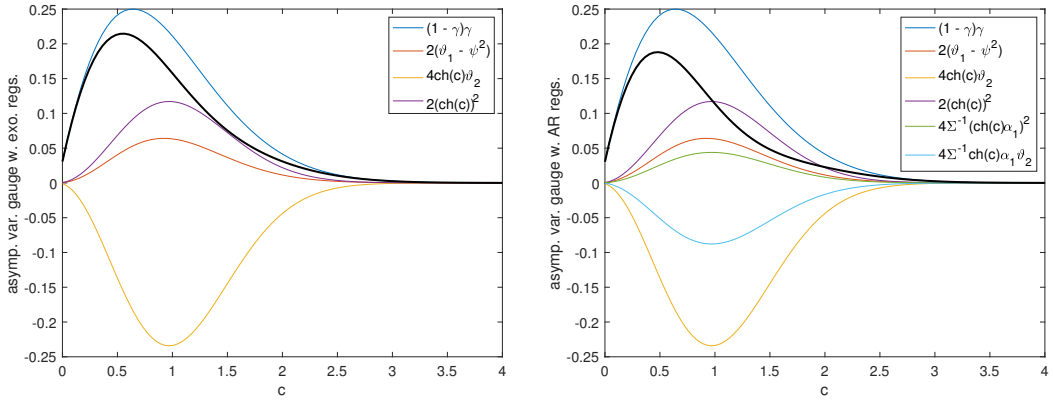



### 2.6.2 Simulations on the distribution

In this section we provide simulation results on the asymptotic distribution of the empirical gauge. We first study the asymptotic variance for small sample sizes. Figure 2.6 shows the small sample bias of the estimated variance of the gauge of Step-indicator Saturation. The simulation is based two different DGPs. On the left panel, we consider DGP $y_i = \varepsilon_i$, where the error terms are standard normal. In the right panel, we consider the DGP in (2.16). We run Algorithm 2.2.1 on the model (2.15). We plot the results for two different sample sizes $n = 100$ and $n = 10000$. In each iteration of the simulation, we redraw all $n$ observations. While the small sample bias of the gauge is both negative for small cut-off values $c$ and positive for large $c$, the range of $c$'s that are usually used in applied work have a positive bias. The larger asymptotic variance of the gauge translates into a larger uncertainty about the number of wrongly retained step-indicators. The large variance of the empirical gauge for smaller $c$ in the range from $[0.5, 2]$ is another reason, why we recommend researchers to use small selection probabilities $\gamma$ with Step-indicator Saturation.

Figure 2.7 presents a decomposition of the individual terms of the asymptotic variance of the gauge for a single exogenous and autoregressive regressor. The two terms that don't depend on estimation errors $a, b$ are $(1-\gamma)\gamma$ and $2(\vartheta_1 - \psi^2)$; the terms that depends on $a$ are $4c\mathsf{h}(c)\vartheta_2$ and $2(c\mathsf{h}(c))^2$ and the terms that depend on $b$ are $4\Sigma^{-1}(c\mathsf{h}(c)\alpha_1)^2$ and $4\Sigma^{-1}c\mathsf{h}(c)\alpha_1\vartheta_2$. Not all terms increase the asymptotic variance. There is one term from the estimation error in $a$ and $b$ respectively that

Figure 2.7: Decomposition of the asymptotic variance of the gauge




decrease the asymptotic variance for all $c$, because they contain $\vartheta_2$ which is negative for all $c$. Another conclusion from this exercise is that a larger autoregressive coefficient in model (2.16), $\alpha_1$, will decrease the asymptotic variance of the gauge.

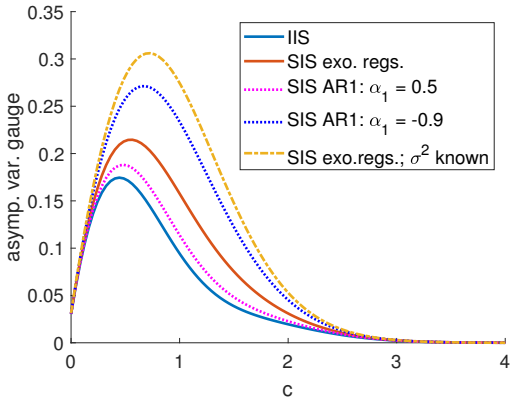
We compare the asymptotic variance of the gauge of the Impulse-indicator Saturation (IIS), as discussed in section 2.2.1, to Step-indicator Saturation (SIS). Johansen and Nielsen (2016*b*) derive in their Corollary 5 with normally distributed errors terms the asymptotic distribution of the empirical gauge of Impulse-indicator Saturation as

$$n^{1/2}\{\hat{\gamma}_n^{\mathrm{IIS}}(c) - \gamma\} \xrightarrow{\mathsf{D}} \mathsf{N}\{0, \gamma(1-\gamma) + 2c\mathsf{h}(c)\varkappa_2 + 2c^2\mathsf{h}^2(c)\}, \tag{2.17}$$

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

In Figure 2.8, we display the different asymptotic variance curves of the gauge as a function of the selection quantile $c$ for IIS and different variations of SIS. The DGP is $y_i = \varepsilon_i$ with standard normal $\varepsilon_i$. We consider three distinct models for SIS. First, we run Algorithm 2.2.1 on the model $y_i = \beta x_i + \varepsilon_i$, where $x_i$ are i.i.d. draws from a standard normal distribution and where $\hat{\omega}_{j,i}^2 = \hat{\sigma}_j^2$. We therefore have from Theorem 2.5.2 that $\omega_{**} = 0$. Second, we run Algorithm 2.2.1 with model (2.16) and $\hat{\omega}_{j,i}^2 = \hat{\sigma}_j^2$. In the DGP, we set $\alpha_1 = 0.5$. Third, we run Algorithm 2.2.1 with $x_i$ as in the first case, but with known error variance $\hat{\omega}_{j,i}^2 = \sigma^2 = 1$. We therefore have from Theorem 2.5.2 that $\omega_* = 0$. We run the corresponding IIS algorithm, as outlined in section 2.2.1 on the model $y_i = \beta x_i + \varepsilon_i$, where $x_i$ are i.i.d. draws from a standard normal distribution and where $\hat{\omega}_{j,i}^2 = \hat{\sigma}_j^2$.

Figure 2.8: Comparison of the asymptotic variance of the gauge




The following facts stand out. First, for all $c$ the asymptotic variance of the gauge in IIS is lower than all four competing SIS models. Second, running SIS knowing the variance of the error term $\sigma$ results in a higher asymptotic variance of the gauge. Third, in the autoregressive model the $\alpha_1$ coefficient changes the asymptotic variance. The asymptotic variance is the largest for $\alpha_1$ close to $-1$. And it is the smallest for $\alpha_1$ close to 1. This is different to IIS, as their theoretical asymptotic variance does not include terms that depend on the regressors. Finally, we observe that the asymptotic variance of the gauge falls rapidly for growing $c$. This gives a rational to choose a large $c$ corresponding to selection probabilities of 1% or lower, as recommended by Castle et al. (2015), as the large $c$ reduces the uncertainty on the empirical gauge.

In Table 2.2 we tabulate the asymptotic variance of the empirical gauge of Step-indicator Saturation for different selection probabilities $\psi = \{5\%, 1\%, 0.5\%, 0.1\%\}$ and compare to the asymptotic variance of the empirical gauge of Impulse-indicator Saturation as stated in (2.17). The data generating process is $y_i = \varepsilon_i$ with standard normal $\varepsilon_i$. We consider four models, which are the same models as in Figure 2.8. We find that at a five percent selection probability, the variance of Step-indicator Saturation with a strictly exogenous regressor and with unknown variance of the error term is over 60 percent higher than for a similar setting with Impulse-indicator Saturation. This difference vanishes for smaller selection probabilities.

Table 2.2: Table of the asymptotic variance of the empirical gauge of Step-indicator Saturation (in %) for different selection probabilities and different DGPs

| | Type of data generating process | | | |
|---|---|---|---|---|
| $\psi$ | IIS -AR -$\sigma^2$ | SIS -AR -$\sigma^2$ | SIS +AR -$\sigma^2$ | SIS -AR +$\sigma^2$ |
| 5% | 2.13 | 3.48 | 2.16 | 4.75 |
| 1% | 0.71 | 0.89 | 0.75 | 0.99 |
| 0.5% | 0.40 | 0.48 | 0.43 | 0.50 |
| 0.1% | 0.01 | 0.01 | 0.01 | 0.01 |

Note: The expression'-AR' refers to a DGP with strictly exogenous regressors which are drawn from an i.i.d. standard normal distribution. The expression '+AR' refers to a DGP with autoregressive regressor. The expression '-$\sigma^2$' refers to an unknown error variance. The expression '+$\sigma^2$' refers to a known error variance.

## 2.7 Proofs

In this section we prove the asymptotic properties of the gauge of Step-indicator Saturation. We do this in three steps. We first discuss a number of auxiliary theorems that we will use in our later proofs. Then, we present the empirical process representation of the gauge of Step-indicator Saturation. And finally, we show the proofs of the asymptotic convergence and distribution results.

### 2.7.1 A collection of preliminary theorems

This section lists theorems used in the proof of the convergence and asymptotic distribution result of the Step-indicator Saturation estimator. First, consider some process $M_n(\hat{\theta}, c)$ depending on estimation errors $\hat{\theta}$. In this case $\hat{\theta}$ is a random variable. If $\hat{\theta}$ is bounded in probability, then $M_n(\hat{\theta}, c)$ can be analysed by treating the estimation error as a constant and showing uniform convergence of $M_n(\theta, c)$ over $\theta \in \Theta$ for a compact $\Theta$. This is due to the following result.

**Theorem 2.7.1.** *(Berenguer-Rico and Nielsen (2017), Lemma A.1.) If* $\forall \epsilon > 0$ *a compact set* $\Theta$ *exists so* $\lim_{n\to\infty} \mathsf{P}(\hat{\theta} \in \Theta^c) < \epsilon$ *then*

$$\mathsf{P}\{|M_n(\hat{\theta}, c)| > \epsilon\} \leq \mathsf{P}\{\sup_{\theta\in\Theta} |M_n(\theta, c)| > \epsilon\} + \epsilon.$$

The next theorem bounds the difference of two indicator functions where a limit on the distance between the boundaries of the indicator functions exists.

**Theorem 2.7.2.** *(Johansen and Nielsen (2009), Lemma 1) For $e < f$, $e_0 < f_0$, and $\max(|e - e_0|, |f - f_0|) \leq \zeta$ we have*

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

The final theorem lists the terms in the covariance matrix that are used in the derivation of the asymptotic distribution of the empirical gauge.

**Theorem 2.7.3.** *Suppose the Assumption 1.2.2, $\mathsf{E}\varepsilon_i^4 < \infty$ and $\mathsf{E}\varepsilon_i^2|x_i|^2 < \infty$ hold. Let*

$$V_i = \left\{1_{(|\nabla\varepsilon_i|\leq\sqrt{2}c\sigma)} - \psi, c\mathsf{h}(c)(\breve{\varepsilon}_i^2 - 1)H_i, \mathsf{h}(c)K_i x_i\varepsilon_i\right\}'.$$

*Recall the definitions of $\gamma$, $\psi$ from (1.7); $\varkappa_3$, $\varkappa_4$ from (1.9); $\vartheta_1$, $\vartheta_2$, $\vartheta_3$, $\vartheta_4$ from (2.5)-(2.8); $\xi_c$ from (1.12); $H_i$, $K_i$ from (2.11),(2.12) and $\mu$, $\Sigma$ from (2.9). Let $\mu_j$, $\Sigma_j$, $\xi_{cj}$ be the subsample analogue of $\mu$, $\Sigma$ and $\xi_c$ with $j = 1, 2$, where each subsample has $n_j$ observations. Recall $n_j n^{-1} \to \lambda_j$ for $n \to \infty$ and have $\tilde{\lambda} = (\lambda_1\lambda_2)^{1/2}$ as in (2.13). Define a diagonal matrix $Q$ which has a diagonal element 1 if the corresponding regressor is stationary and 0 otherwise. Let*

$$\Sigma^\dagger = \xi_{c2}'\Sigma_1^{-1} + \xi_{c1}'\Sigma_2^{-1}, \quad \Sigma^\ddagger = \lambda_1\xi_{c1}'\Sigma_2^{-1}\xi_{c1} + \lambda_2\xi_{c2}'\Sigma_1^{-1}\xi_{c2}.$$

*The vector $V_i$ has mean zero and variance-covariance matrices*

$$n^{-1}\sum_{i=1}^{n}\mathsf{E}(V_iV_i') \to \begin{pmatrix} (1-\gamma)\gamma & 2c\mathsf{h}(c)\vartheta_2\tilde{\lambda} & \sigma\vartheta_4\mathsf{h}(c)\tilde{\lambda}\Sigma^\dagger Q\mu \\ * & c^2\mathsf{h}^2(c)(\varkappa_4 - 1) & \sigma\varkappa_3 c\mathsf{h}^2(c)\tilde{\lambda}\Sigma^\dagger Q\mu \\ * & * & \sigma^2\mathsf{h}^2(c)\Sigma^\ddagger \end{pmatrix},$$

$$n^{-1}\sum_{i=1}^{n}\mathsf{E}(V_iV_{i+1}') \to \begin{pmatrix} \vartheta_1 - \psi^2 & 2c\mathsf{h}(c)\vartheta_2\tilde{\lambda} & \sigma\mathsf{h}(c)\tilde{\lambda}\Sigma^\dagger Q\vartheta_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

**Remark 2.7.1.** *For random walk and trend stationary regressors $x_i$, we have for $n \to \infty$ that $\Sigma^\dagger \to 0$ and $\Sigma^\ddagger \to 0$ which simplifies the variance-covariance matrix accordingly. The terms also simplify for stationary regressors. In particular $\Sigma^\dagger = 2\xi_c'\Sigma^{-1}$ and $\Sigma^\ddagger = \xi_c'\Sigma^{-1}\xi_c$.*

To derive the asymptotic distribution of the empirical gauge, we use the McLeish (1975) mixingale central limit theorem. A mixingale is defined as a sequence of

$\mathcal{F}_i$-adapted random variables $\{z_i\}_{i=1}^n$ for which there exists a positive sequence $\psi_k \to 0$ as $k \to \infty$, such that for all $i \geq 1$, $k \geq 0$, $\mathsf{E}^{1/2}|\mathsf{E}_{i-k}z_i|^2 \leq \psi_k$ and $\mathsf{E}^{1/2}|z_i - \mathsf{E}_{i+k}z_i|^2 \leq \psi_{k+1}$. The sequence $\psi_k$ is considered to be of size $-p$ if there exists a positive sequence $\{L(k)\}$ which satisfies (i) $\sum_{k=1}^n \{kL(k)\}^{-1} < \infty$; (ii) $L(k) - L(k-1) = O\{L(k)/k\}$; (iii) $L(k)$ is eventually non-decreasing; such that $\psi_k = o\{(k^{1/2}L(k))^{-2p}\}$. A simple choice for $L(k)$ that is sufficient for our purpose is $L(k) = k$.

The uniform integrability of a random variable is defined as in Billingsley (1979). We say that a random variable $z_i$ is uniformly integrable, if for $\alpha \to \infty$

$$\sup_{1 \leq i \leq n} \mathsf{E}\{|z_i| 1_{(|z_i| \geq \alpha)}\} \to 0.$$

**Theorem 2.7.4.** *(McLeish (1975), Theorem 2.6) Let $\mathcal{F}_i$-adapted sequence of random variables $\{z_i\}_{i=1}^n$ be a mixingale with $\psi_k$ of size $-1/2$ satisfying $\mathsf{E}z_i = 0$ for all $i$ and $\mathsf{E}S_n^2/n \to \sigma^2$, where $S_n = \sum_{i=1}^n z_i$. If $z_i$ is uniformly square integrable and $\mathsf{E}\{n^{-1}(S_{k+n} - S_k)^2|\mathcal{F}_{k-m}\} \to \sigma^2$ in $L_1(\Omega)$-norm as $\min(m,k,n) \to \infty$, then $n^{-1/2}S_n \xrightarrow{\mathsf{D}} \mathsf{N}(0,\sigma^2)$.*

## 2.7.2 The empirical process representation of the gauge

We show that the empirical gauge can be written in an empirical process representation.

**Theorem 2.7.5.** *Consider the empirical gauge of the split-half Step-indicator Saturation estimator in Algorithm 2.2.1, namely $\hat{\gamma}_n(c) = n^{-1}\sum_{i=1}^n 1_{(|v_i|>c)}$, where the random variable $v_i$ is defined as*

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

*and where the estimated variance is $\hat{\omega}_{j,i}^2 = \hat{\sigma}_j^2\{2 + \nabla x_i'(\sum_{k \in I_j} x_k x_k')^{-1}\nabla x_i\}$. Suppose Assumptions 1.2.2 and 2.4.1 (i,iii,iv) hold. Then, we get*

$$\sup_{c \in \mathbb{R}} \left| \hat{\gamma}_n(c) - \{1 - \lambda_1 \widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \lambda_2 \widehat{\mathsf{G}}_{n_2}(\hat{a}_1, \hat{b}_1, c)\} \right| = \mathrm{o}_{\mathsf{P}}(1). \tag{2.18}$$

**Proof of Theorem 2.7.5**. Introduce the subsample estimation errors

$$\hat{a}_{j,i} = n_j^{1/2}(\hat{\omega}_{j,i} - \sigma), \quad \hat{a}_j = n_j^{1/2}(\hat{\sigma}_j - \sigma), \quad \hat{b}_j = (N_j^{-1})'(\hat{\beta}_j - \beta), \tag{2.19}$$

where $j = 1, 2$ refers to the respective half-samples and where $n_j$ refers to the subsample sample sizes. Note the difference of $\hat{a}_{j,i}$ and $\hat{a}_j$, where the former follows the usual finite sample formula for the distribution of residuals for fixed regressors which we discussed in (2.1). Define also subsample normalised regressors as $\nabla x_{inj} = N_j' \nabla x_i$, where the normalisation $N_j$ is the subsample analogue of $N$ as discussed in 2.10.

We write $v_i$ in terms of the estimation errors $\hat{a}_{j,i}$ and $\hat{b}_j$. To do so, note that

$$\nabla y_i - \hat{\beta}_j \nabla x_i = \nabla \varepsilon_i - \nabla x_{inj}' \hat{b}_j, \quad \sqrt{2}c\hat{\omega}_{j,i} = \sqrt{2}c(n_j^{-1/2}\hat{a}_{j,i} + \sigma),$$

such that $v_i$, with $\tilde{x}_{inj}$ and $\tilde{\varepsilon}_i$ relating to the $\sqrt{2}c$ normalisation, can be written as

$$v_i = \frac{\nabla \tilde{\varepsilon}_i - \nabla \tilde{x}_{in2}' \hat{b}_2}{(n_2^{-1/2}\hat{a}_{2,i} + \sigma)} 1_{(i \in n_1)} + \frac{\nabla \tilde{\varepsilon}_i - \nabla \tilde{x}_{in1}' \hat{b}_1}{(n_1^{-1/2}\hat{a}_{1,i} + \sigma)} 1_{(i \in n_2)}.$$

We then rewrite the empirical gauge $\hat{\gamma}_n(c)$ with these $v_i$ in terms of $\hat{a}_{ij}$ and $\hat{b}_j$ as

$$\hat{\gamma}_n(c) = 1 - \frac{\lambda_1}{n_1} \sum_{i \in I_1} J_i(\underline{c}_{i2}^{\dagger}, \overline{c}_{i2}^{\dagger}) - \frac{\lambda_2}{n_2} \sum_{i \in I_2} J_i(\underline{c}_{i1}^{\dagger}, \overline{c}_{i1}^{\dagger}), \tag{2.20}$$

where $\lambda_j = n_j n^{-1}$, $J_i(x, y) = 1_{(\nabla \tilde{\varepsilon}_i \leq y)} - 1_{(\nabla \tilde{\varepsilon}_i \leq x)}$ and where its boundaries are

$$\overline{c}_{ij}^{\dagger} = \nabla \tilde{x}_{inj}' \hat{b}_j + |c|(1 + n_j^{-1/2}\sigma^{-1}|\hat{a}_{ij}|), \quad \underline{c}_{ij}^{\dagger} = \nabla \tilde{x}_{inj}' \hat{b}_j - |c|(1 + n_j^{-1/2}\sigma^{-1}|\hat{a}_{ij}|).$$

We proceed to replace $\hat{a}_{j,i}$ with $\hat{a}_j$. Introduce a new set of boundaries of $J_i$

$$\overline{c}_{ij} = \nabla \tilde{x}_{inj}' \hat{b}_j + |c|(1 + n_j^{-1/2}\sigma^{-1}|\hat{a}_j|) \quad \underline{c}_{ij} = \nabla \tilde{x}_{inj}' \hat{b}_j - |c|(1 + n_j^{-1/2}\sigma^{-1}|\hat{a}_j|).$$

Then, add and subtract $J_i(\underline{c}_j, \overline{c}_j)$ from (2.20) to get with the triangle inequality

$$\begin{aligned} &|\hat{\gamma}_n(c) - \{1 - \lambda_1 \widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \lambda_2 \widehat{\mathsf{G}}_{n_2}(\hat{a}_1, \hat{b}_1, c)\}| \\ &\quad \leq \left| \frac{\lambda_1}{n_1} \sum_{i \in I_1} \left\{ J_i(\underline{c}_{i2}^{\dagger}, \overline{c}_{i2}^{\dagger}) - J_i(\underline{c}_{i2}, \overline{c}_{i2}) \right\} \right| + \left| \frac{\lambda_2}{n_2} \sum_{i \in I_2} \left\{ J_i(\underline{c}_{i1}^{\dagger}, \overline{c}_{i1}^{\dagger}) - J_i(\underline{c}_{i1}, \overline{c}_{i1}) \right\} \right|, \end{aligned}$$

Without loss of generality, consider the first half-sample. Use Theorem 2.7.2 with $\max_{1\leq i\leq n}|\hat{a}_{2,i}-\hat{a}_2|\leq\delta_n$, where $\delta_n\to 0$ as $n\to\infty$ to find the following upper bound

$$\begin{aligned}\mathcal{H}_{n_1}&=n_1^{-1/2}\sum_{i\in I_1}\left\{J_i(\underline{c}_{i2}^{\dagger},\overline{c}_{i2}^{\dagger})-J_i(\underline{c}_{i2},\overline{c}_{i2})\right\}\\&\leq n_1^{1/2}\left\{\widehat{\mathsf{H}}_{n_1}(\hat{a}_2,\hat{b}_2,c,\delta_n)+\widehat{\mathsf{H}}_{n_1}(\hat{a}_2,\hat{b}_2,-c,\delta_n)\right\},\end{aligned}$$

where $\widehat{\mathsf{H}}_{n_1}(a,b,c,\delta_n)=n^{-1}\sum_{i=1}^n 1_{\{|\nabla\tilde{\varepsilon}_i-|c|(1+n^{-1/2}\sigma^{-1}|a|)-\nabla\tilde{x}_{in}'b|\leq|c|n^{-1/2}\sigma^{-1}\delta_n\}}$ is defined as in (1.15). By Assumption 2.4.1(iv), $\forall\varepsilon,C_n>0$, $\exists B,n_0>0$ so that the sets

$$\mathcal{A}_n=(\max_{1\leq i\leq n}|\hat{a}_{2,i}-\hat{a}_2|\leq\delta_n),\quad \mathcal{B}_n=(|\hat{a}_2|,|\hat{b}_2|\leq Bn^{1/4-\eta})$$

satisfy $\mathsf{P}(\mathcal{A}_n^c)+\mathsf{P}(\mathcal{B}_n^c)<\varepsilon$ for $n>n_0$. By Theorem 2.7.1, we can treat the estimation error as a constant and show uniform convergence of $\widehat{\mathsf{H}}_{n_1}(a,b,c,\delta_n)$ over $a,b$. By Theorem 1.3.3 and for $\delta_n\to 0$ as $n\to\infty$ we get for $c\in\mathbb{R}$

$$\mathcal{H}_{n_1}\leq\sup_{|a|,|b|\leq n^{1/4-\eta}B}2n_1^{1/2}\widehat{\mathsf{H}}_{n_1}(a,b,c,\delta_n)=\mathrm{o}_{\mathsf{P}}(1).$$

Repeat for the second subsample and note that $0<\lambda_j<\infty$ to finish the proof. ■

### 2.7.3 Proof of mean square convergence of gauge

**Proof of Theorem 2.5.1**. We show that $|\hat{\gamma}_n(c)-\gamma|=\mathrm{o}_{\mathsf{P}}(1)$. First rewrite the empirical gauge in the empirical process representation by applying Theorem 2.7.5 with Assumptions 1.2.2 and 2.4.1(i,iii,iv) such that

$$\hat{\gamma}_n(c)=1-\lambda_1\widehat{\mathsf{G}}_{n_1}(\hat{a}_2,\hat{b}_2,c)-\lambda_2\widehat{\mathsf{G}}_{n_2}(\hat{a}_1,\hat{b}_1,c)+\mathrm{o}_{\mathsf{P}}(1).$$

Without loss of generality, consider the first split-sample. Rewrite the empirical gauge of the first split-sample as

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

We argue that $\widehat{\mathsf{G}}_{n_1}(\hat{a}_2,\hat{b}_2,c)-\widehat{\mathsf{G}}_{n_1}(0,0,c)$ vanishes. As $(\hat{a},\hat{b})=\mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$, for all $\epsilon>0$ there exists a $B>0$ so that for large $n$ then $\mathsf{P}\{|(\hat{a},\hat{b})|\leq n^{1/4-\eta}B\}\geq 1-\epsilon$. By Theorem 2.7.1 it suffices to show $\sup_{|a|,|b|\leq n^{1/4-\eta}}|\widehat{\mathsf{G}}_{n_1}(a,b,c)-\widehat{\mathsf{G}}_{n_1}(0,0,c)|=\mathrm{o}_{\mathsf{P}}(1)$.

As in the proof of Theorem 2.5.2, we decompose $\widehat{\mathsf{G}}_{n_1}(\hat{a}_2, \hat{b}_2, c)$ and apply the triangle inequality as

$$\widehat{\mathsf{G}}_{n_1}(a,b,c) - \widehat{\mathsf{G}}_{n_1}(0,0,c) \leq n^{-1/2}\{|\mathcal{G}_{n_1}(a,b,c)| + \left|\overline{\mathcal{G}}_{n_1}(a,b,c)\right| + |\mathcal{B}_1(a,b,c)|\}.$$

By Theorem 1.3.1 and Assumptions 1.2.2 and 2.4.1(ia,iiib) we have that the term $n^{-1/2}|\mathcal{G}_{n_1}(\hat{a}_2, \hat{b}_2, c)|$ vanishes uniformly in $a, b$ for $n_1 \to \infty$. By Theorem 1.3.6 and Assumptions 1.2.2 and 2.4.1(i,ii, iiib), we have that we have that $n^{-1/2}|\overline{\mathcal{G}}_{n_1}|$ vanishes uniformly in $a, b$ for $n_1 \to \infty$. As $|a|, |b| \leq n^{1/4-\eta}$ we have that the term $n^{-1/2}|\mathcal{B}_1(a,b,c)| = n^{-1/2}\{\mathsf{h}(c)(2ca\sigma^{-1} + \xi'_{c1}b)\} = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in $a, b$ for $n_1 \to \infty$.

In the final step, we apply the weak Law of Large Numbers for i.i.d. random variables to $\widehat{\mathsf{G}}_{n_1}(0,0,c)$. We create the independent random variables by splitting the sample along odd and even observations with $\lambda^{odd} = n_1^{odd}/n_1$ and $\lambda^{even} = n_1^{even}/n_1$ where $n_1^{odd}$ refers to odd subsample sample sizes and $n_1^{even}$ to even ones. We collect the odd observations in $I^{odd}$ and even ones in $I^{even}$. Decompose

$$\widehat{\mathsf{G}}_{n_1}(0,0,c) = \frac{\lambda^{odd}}{n_1^{odd}} \sum_{i \in I_1^{odd}} 1_{(|\nabla \varepsilon_i| \leq \sqrt{2}\sigma c)} + \frac{\lambda^{even}}{n_1^{even}} \sum_{i \in I_1^{even}} 1_{(|\nabla \varepsilon_i| \leq \sqrt{2}\sigma c)}.$$

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

### 2.7.4 Proof of convergence in distribution

**Proof of Theorem 2.5.2**. Our object of interest is $G_n = \sqrt{n}\{\hat{\gamma}_n(c) - \gamma\}$, where $\hat{\gamma}_n(c)$ is defined as in (2.18). Noting $\psi = 1 - \gamma$, we then have by Theorem 2.7.5

$$G_n = \sqrt{n}\left\{\psi - \lambda_1 \widehat{G}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \lambda_2 \widehat{G}_{n_2}(\hat{a}_1, \hat{b}_1, c)\right\} + \mathrm{o}_{\mathsf{P}}(1),$$

where $\lambda_j = n_j n^{-1}$ for $j = 1, 2$. We decompose this term into two components

$$G_n = -\sqrt{\lambda_1} G^{\dagger}_{n_1} - \sqrt{\lambda_2} G^{\dagger}_{n_2} \tag{2.21}$$

with $G^{\dagger}_{n_1} = \sqrt{n_1}\left\{\widehat{G}_{n_1}(\hat{a}_2, \hat{b}_2, c) - \psi\right\}$ and $G^{\dagger}_{n_2} = \sqrt{n_2}\left\{\widehat{G}_{n_2}(\hat{a}_1, \hat{b}_1, c) - \psi\right\}$. As before, $n_j$ refers to the sample size of the respective subsample.

Without loss of generality, consider the first split-sample. Bound the empirical distribution function with

$$G_{n_1}^{\dagger} \leq \sqrt{n_1}\left\{\widehat{\mathsf{G}}_{n_1}(0,0,c)-\psi\right\}+\left|\mathcal{G}_{n_1}(\hat{a}_2,\hat{b}_2,c)\right|+\left|\overline{\mathcal{G}}_{n_1}(\hat{a}_2,\hat{b}_2,c)\right|+\left|\mathcal{B}_1(\hat{a}_2,\hat{b}_2,c)\right|,$$

where $\mathcal{G}_{n_1}, \overline{\mathcal{G}}_{n_1}$ and $\mathcal{B}_1$ are defined as in (1.10),(1.13) and (1.14).

*The term* $\left|\mathcal{G}_{n_1}(\hat{a}_2,\hat{b}_2,c)\right| = \mathrm{o}_{\mathsf{P}}(1)$. Because $(\hat{a}_2,\hat{b}_2) = \mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$ by Assumptions 2.4.1(iva,ivb), for all $\epsilon$, there exists a $B$, so that for large $n$ then $\mathsf{P}\{|(\hat{a}_2,\hat{b}_2)| \leq n^{1/4-\eta}B\} \geq 1-\epsilon$. Thus, when we replace the random variables $\hat{a}_2, \hat{b}_2$ with constants $a, b$, then by Theorem 2.7.1 it suffices to show that $\sup_{|a|,|b|\leq n^{1/4-\eta}} |\mathbb{G}_{n_1}(a,b,c) - \mathbb{G}_{n_1}(0,0,c)| = \mathrm{o}_{\mathsf{P}}(1)$, which follows from Theorem 1.3.1 and Assumptions 1.2.2 and 2.4.1(ia,ivc,ivd).

*The term is* $\left|\overline{\mathcal{G}}_{n_1}(\hat{a}_2,\hat{b}_2,c)\right| = \mathrm{o}_{\mathsf{P}}(1)$. Use Theorem 1.3.2 with Assumptions 1.2.2 and 2.4.1(i,ii,iiib) and the same truncation argument on $(\hat{a}_2,\hat{b}_2)$.

*For the term* $\mathcal{B}_1(\hat{a}_2,\hat{b}_2,c)$, we Taylor expand the estimation errors $\hat{a}_2, \hat{b}_2$ as

$$\begin{aligned}(N_2^{-1})'(\hat{\beta}_2-\beta) &= \Sigma_2^{-1}N_2\sum_{i\in I_2}x_i\varepsilon_i+\mathrm{o}_{\mathsf{P}}(1),\\ n_2^{1/2}(\hat{\sigma}_2-\sigma) &= 2^{-1}n_2^{-1/2}\sum_{i\in I_2}(\varepsilon_i^2/\sigma-\sigma)+\mathrm{o}_{\mathsf{P}}(1),\end{aligned}$$

where $\Sigma_2$, $\xi_{c1}$ and $N_2$ are the subsample analogue of $\Sigma$, $\xi_c$ and $N$. Recall $\mathcal{B}_1(\hat{a}_2,\hat{b}_2,c) = \mathsf{h}(c)(2\hat{a}_2c\sigma^{-1}+\xi_{c1}'\hat{b}_2)$ and insert the expansions into the bias term to get

$$\mathcal{B}_1(\hat{a}_2,\hat{b}_2,c) = \mathsf{h}(c)\sum_{i\in I_2}\left\{\left(\check{\varepsilon}_i^2-1\right)cn_2^{-1/2}+\xi_{c1}'\Sigma_2^{-1}N_2x_i\varepsilon_i\right\}+\mathrm{o}_{\mathsf{P}}(1).$$

*The term* $G_n$. We plug our subsample expressions for $G_{n_1}^{\dagger}$ and $G_{n_2}^{\dagger}$ into (2.21) and we study the empirical distribution term and the bias term separately. For the former we get

$$\sum_{j=1}^{2}-\lambda_j^{1/2}{n_j}^{1/2}\{\widehat{\mathsf{G}}_{n_j}(0,0,c)-\psi\} = -\sqrt{n}\{\widehat{\mathsf{G}}_n(0,0,c)-\psi\}.$$

For the latter we get

$$\begin{aligned}&-\sqrt{\lambda_1}\mathcal{B}_1(\hat{a}_2,\hat{b}_2,c)-\sqrt{\lambda_2}\mathcal{B}_2(\hat{a}_1,\hat{b}_1,c)\\ &=-\mathsf{h}(c)n^{-1/2}\sum_{i=1}^{n}\left\{\left(\check{\varepsilon}_i^2-1\right)cH_i+K_ix_i\varepsilon_i\right\}+\mathrm{o}_{\mathsf{P}}(1),\end{aligned}$$

where

$$H_i = n_2^{1/2} n_1^{-1/2} 1_{(i \in I_1)} + n_1^{1/2} n_2^{-1/2} 1_{(i \in I_2)},$$
$$K_i = n_1^{1/2} \xi_{c1}' \Sigma_2^{-1} N_2 1_{(i \in I_1)} + n_2^{1/2} \xi_{c2}' \Sigma_1^{-1} N_1 1_{(i \in I_2)}.$$

Combining the terms yields the desired expression. ■

**Proof of Theorem 2.5.3**. Note that $n^{1/2}\{\hat{\gamma}_n(c) - \gamma(c)\} = S_n n^{-1/2}$. We write $S_n = \sum_{i=1}^n X_i$, where

$$X_i = 1_{(|\nabla \varepsilon_i| \leq \sqrt{2}\sigma c)} - \psi + c\mathsf{h}(c)\left(\check{\varepsilon}_i^2 - 1\right) H_i + \mathsf{h}(c) K_i x_i \varepsilon_i,$$

where $H_i$ and $K_i$ are defined as in (2.11) and (2.12). Note that Note that the $X_i$ term consists of a 2-dependent term, an i.i.d. term and a m.d.s. term. Moreover, $X_i$ is $\mathcal{F}_{i+1}$-adapted. We have $\psi_0 = \mathsf{E}|X_{i+1}|^2 < \infty$ by Assumption 1.2.2 and 2.4.1(iiib) and $\psi_1 = \mathsf{E}|\mathsf{E}_i X_{i+1}|^2 \leq 1 + \psi \leq 2$. Moreover $\psi_k = 0$ for $k \geq 2$ as $\mathsf{E}_{i-k+1} X_{i+1} = 0$. The sequence $\{X_i\}_{i=1}^n$ is thus a mixingale with $\psi_k$ of size[1] $-1/2$, satisfying $\mathsf{E}X_i = 0$ for all $i$ and $\mathsf{E}S_n^2/n \to \sigma_X^2$, where $\sigma_X^2 = \mathsf{E}X_1^2 + 2\mathsf{E}(X_1 X_2)$.

It also follows that with $\min(k, m, n) \to \infty$ we have

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

We next show the uniform integrability of $X_i^2$. By Hölder's inequality we have

$$\mathsf{E}X_i^2 1_{(x_i^2 > \alpha)} \leq \{\mathsf{E}X_i^{2p}\}^{1/p} \{\mathsf{E}(1_{(X_i^2 > \alpha)})^q\}^{1/q},$$

for same $1/p + 1/q = 1$ where $p > 1$ small and $q > 1$ large. Note that $\mathsf{E}(1_{(X_i^2 > \alpha)})^q \leq \mathsf{E}(1_{(X_i^2 > \alpha)}) = \mathsf{P}(X_i^2 > \alpha)$. We then have by Assumption 2.4.1(ib) and (iiib) that $\mathsf{E}X_i^2 < \infty$, such that by the Markov inequality $\mathsf{P}(X_i^2 > \alpha) \leq \mathsf{E}X_i^2/\alpha \to 0$ for $\alpha \to \infty$. By the same assumption we have $\mathsf{E}X_i^{2p} < \infty$, such that $\mathsf{E}X_i^2 1_{(X_i^2 > \alpha)} \to 0$. We can then apply Theorem 2.7.4 to get the limiting distribution[2], where we use Theorem 2.7.3 for the covariance terms in $\sigma_X^2$. ■

[1] Following McLeish (1975), for a constant $q > 0$ and a sequence of non-negative constants $\{\psi_k : k \geq 1\}$, we say that $\{\psi_k : k \geq 1\}$ is of size $-q$ if $\psi_k$ converges to zero at a fast enough rate. By Andrews (1993), a sufficient condition is that $\psi_k = \mathrm{O}(k^{-\lambda})$ for some $\lambda > q$. In fact, we have that $\psi_k$ is of any size $q > 0$, because $\psi_k = 0$ for $k \geq 2$.

[2] The index sets $I_1$ and $I_2$ can depend on $n$ and where this is the case, it is necessary to apply the CLT to the two subsets separately and merge them subsequently. In our proof we can think of index sets containing even and odd observations respectively, such that no dependence on $n$ emerges.

The proof of Theorem 2.5.4 follows along the lines of the proof of Theorem 2.5.3, but simplifies as $\xi_c \to 0$ for random walk and trend stationary regressors. Therefore the m.d.s. component drops out of $X_i$.

## 2.8 Conclusion

In this chapter we developed the first asymptotic properties of split-half algorithm with Step-indicator Saturation. The results were limited to the empirical gauge of this algorithm, which is the average frequency of false positives.

Our original goal was to study the effect of Step-indicator Saturation on statistical inference of $\hat{\beta}$ and $\hat{\sigma}$. On this front, unfortunately, we have not progressed very far, because we did not find a representation of the ordinary least squares estimates that we could apply our empirical process result to. In Johansen and Nielsen (2009) the authors were able to study the effects of Impulse-indicator Saturation, because their single regressor least squares estimator could be written as

$$\hat{\beta}_{\text{IIS}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})1_{(|\hat{\varepsilon}_i| \leq \hat{\sigma}c)}}{\sum_{i=1}^{n}(x_i - \bar{x})^2 1_{(|\hat{\varepsilon}_i| \leq \hat{\sigma}c)}}.$$

In contrast to them, in the presence of Step-indicator Saturation and single regressor least squares we get for a single breakpoint in $n_1$ that

$$\hat{\beta}_{\text{SIS}} = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)(y_i - \bar{y}_1) + \sum_{i=n_1+1}^{n}(x_i - \bar{x}_2)(y_i - \bar{y}_2)}{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + \sum_{i=n_1+1}^{n}(x_i - \bar{x}_2)^2}.$$

Moreover, while the impulse indicator saturation does not introduce a bias if applied to a autoregressive model, Step-indicator Saturation produces a small bias for the time-series coefficients in autoregressive regression models, such as (2.16). To illustrate this, we run a simulation with model (2.16), the time-series coefficient $\alpha_1 = 0.8$, the sample size $n = 1,000$ and the number of replications $M = 10,000$. In the first case, we use Step-indicator Saturation at a 5% significance level and collect the estimates for the time-series coefficient. In the second case, we estimate regular least squares regression coefficients. The result of the simulation is plotted in Figure 2.9.

We observe a small downwards bias of the time-series coefficient in the autoregressive regression model. We do not expect this bias to vanish asymptotically.

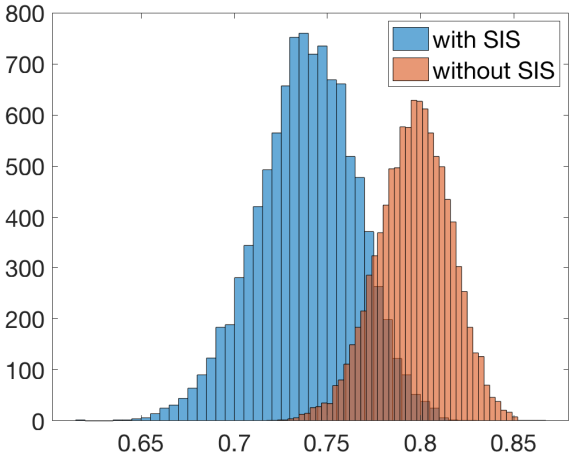



Figure 2.9: A histogram of regression estimates of autoregressive regression models with and without Step-indicator Saturation. The x-axis shows $\alpha_1$ and the y-axis shows count numbers.

This bias emerges due to the difficulty to distinguish a random walk from a stationary time-series with location shifts. In our simulations, we have $\alpha_1 = 0.8$ and therefore a strong time-series persistence of our process. The spurious step-indicators pick up a share of the time-series persistence, reducing the remaining time-series correlation that we can estimate. In our future work, we aim to get an analytical expression for the asymptotic bias that Step-indicator Saturation introduces to time-series models.

Over the past two chapters, we first developed the mathematical engine to study the empirical gauge of Step-indicator Saturation. Then, we applied our results to derive asymptotic convergence and distribution results for the empirical gauge. We will leave behind the theoretical considerations of Step-indicator Saturation and turn to an application of this procedure to study the relationship between deindustrialisation and real interest rates.

# Chapter 3

# Deindustrialisation: A secular force towards lower real rates

Abstract. Since the inception of the Bank of England the interest rates in the United Kingdom have never been as low as during the economic malaise that followed the Great Recession of 2007-09. This chapter argues that deindustrialisation, which we define as the declining output share of manufacturing, is a key driver for lower long-run real rates. We motivate the relationship using a general equilibrium model in the context of a trade shock which originates from the rise of China as a manufacturing power house. We test our hypothesis empirically using the catalyst method, which exploits the causal transmission of large exogenous shocks, to identify the unique direction of causation. From 1990 to 2017, we estimate that deindustrialisation has reduced the British long-run real rates by 419 basis points.

## 3.1 Introduction

Over the last 30 years we have witnessed a large decline in the real interest rate around the world (King and Low (2014); Holston et al. (2017)). The dominant explanations for this phenomenon follow the "New Secular Stagnation" hypothesis by Summers (2014), which explains lower real rates with structural changes in the global economic system. These include demographic change (Giuseppe et al. (2017)), international patterns of saving flows (Bernanke (2005)) and the lower cost of capital (Sajedi and Thwaites (2016)). However, none of the proposed secular

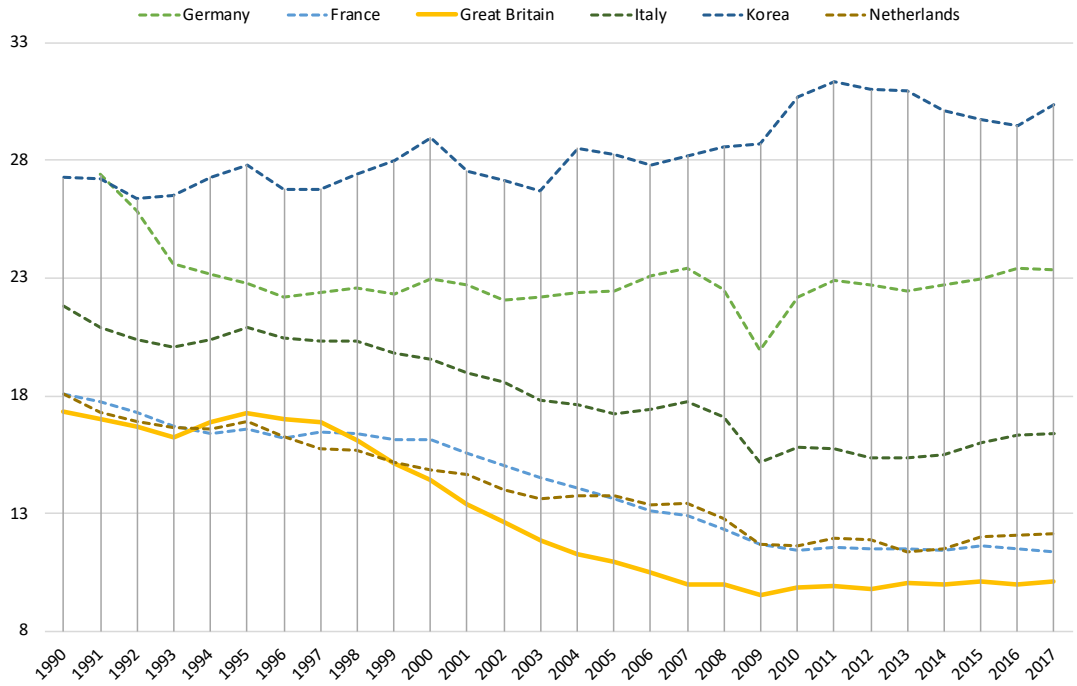



Figure 3.1: A comparison of the deindustrialisation of Great Britain with other highly industrialised countries from 1990 to 2017. The y-axis shows the gross value add of manufacturing to total output in percent.

forces have considered the structural change within developed economies, namely the shift away from manufacturing towards the service sector.

In this chapter, we study this empirical relationship between the secular decline of the share of manufacturing output, a trend we call deindustrialisation, and the real interest rate in the context of the United Kingdom. First, Britain experienced the most pronounced deindustrialisation among diversified economies and it has the lowest share of manufacturing output among highly industrialised countries as of 2017. Using panel data from OECD (2017) this is illustrated in a comparison of the British deindustrialisation to other OECD member countries in Figure 3.1. Second, Britain has the longest availability of a historical time series of a market driven real rate, as it has been the first country to issue an index-linked government bond in 1981. These securities allow researchers to compute a range of *ex ante* real rates which are based on forward looking inflation expectations.

We explain the observation that deindustrialisation is associated with lower real interest rates using trade shocks which originate far away from Britain in the industrialising emerging economies of East Asia. To do so, we make use of the 1-2-3 model proposed by Devarajan et al. (1990) which is a general equilibrium model with 3 goods, 2 domestic sectors and 1 country. In that model, the effect of the industrialisation of emerging countries is a lower price of manufacturing goods in international markets. We assume that this disproportionately affects the price of exports relative to the price of imports. We show that this deterioration in the terms of trade can explain why deindustrialisation is coupled with lower real rates. In section 3.3 we describe and solve the 1-2-3 model.

The central contribution of this chapter is the identification of a causal effect of the fall in the share of manufacturing output on real rates. The causality is established with the application of the 'catalyst method' by Bazinas and Nielsen (2015). In general, the deduction of a causal ordering between two observed variables from their joint distribution is difficult. However, if we can find a third variable, which represents large exogenous shocks we name catalysts, we can determine the order in which this shock transmits between the two variables of interest. A set of testable conditions can then be used to establish their causal ordering, by pinning down a unique direction of the influence from one variable to another. Section 3.4.1 and 3.4.2 explain the catalyst method in greater detail.

Based on our theoretical framework, which suggests that terms of trade shocks can give rise to a both deindustrialisation and lower real rates, we use them as our catalyst. Our empirical results establish the causal transmission of our terms of trade catalyst. The causal transmission suggests a unique causal ordering in which terms of trade shocks first result in deindustrialisation that in turn culminate in lower real rates. On average a one percent point decline of the share of manufacturing output resulted in a 79 basis point decline in the real rate. From 1990 to 2017, this accounts for a 418 basis point decline of the real rate. The quantitative impact is somewhat larger than the effect of the savings glut, which Beltran et al. (2013) and Warnock and Warnock (2009) estimate to be between 40 to 80 basis points. In section 3.5 we discuss our empirical results in detail.

This chapter aims to showcase the use of Step-indicator Saturation in an empirical setting. Our methodological innovation is the joint use of the catalyst method

and Step-indicator Saturation. We use the catalyst method as an alternative way to study causal relationships. As discussed above, the catalyst method is based on three testable hypothesis which can be used to establish causal transmission. The open question is how to construct the catalyst in the first place. Bazinas and Nielsen (2015) have used indicator saturation methods to identify outliers. They then attributed these outliers to specific exogenous forces, such as oil shocks and fiscal shocks, and these dummy variables are used as catalysts. Instead, in this chapter, we use a different strategy to construct the catalyst. We start with a theoretical model that argues that terms of trade shocks can explain both deindustrialisation and a lower real interest rate. We then run Step-indicator Saturation to find the location shifts in the terms of trade variable. Out of this large collection of location shifts, we select the subset of structural breaks that best fit the conditions of causal transmission.

What are the policy implications of this study? This study explains low real interest rates with the rapid industrialisation of emerging economies which accelerated deindustrialisation in the United Kingdom. It is the structural change in the pattern of the global manufacturing supply chain that can explain the low rates of returns for savers. According to this narrative, the downwards pressure on real rates should subside as the deindustrialisation of advanced economies decelerates and as populist backlashes against the outsourcing of manufacturing employment grows. Indeed, to increase the real interest rate to a higher level, it may be beneficial to implement a government policy to support the investment in capital intensive manufacturing industries. This could help lift the economy out of the secular stagnation and solve the problems of monetary policy at the zero lower bound.

## 3.2 Preliminary evidence

In this section, we turn to the empirical analysis of the effect of the decline of the manufacturing sector on the financial markets. We explain and justify our choice of the variables that we use to measure deindustrialisation and the real rate.

### 3.2.1 Our measure of the deindustrialisation

There are two broad categories of measures for deindustrialisation that are used in the literature. First, there are employment share based measures of deindustrialisation (Rowthorn and Ramaswamy (1997, 1998); Lawrence et al. (2013)). They are defined as the share of manufacturing employment to the total employed workforce. Second, there are measures based on the share of gross value added in manufacturing to total output (Nickell et al. (2008); Rodrik (2016)).[1] These output share based measures use either current or constant prices.[2]

We use the share of manufacturing output at constant prices as our measure of deindustrialisation. We use an output based measure, because they are less sluggish than the employment based measures. For example, exogenous shocks such as a foreign financial crisis can markedly reduce manufacturing output within the same quarter by reducing the demand for exports. However, due to employment laws, the adjustment of the workforce can take several quarters.

We construct our measure using quarterly Bloomberg data on the gross value added in manufacturing in constant 2013 prices and the corresponding measure on GDP. The original source of the data is the Office of National Statistics in the United Kingdom. We call this variable DEIND.

Figure 3.2, panel c and d, plot the levels and first-differences of the share of the manufacturing output at constant 2013 prices. From 1990 to 2017, this share has declined from 14.2 percent to 8.7 percent. We apple to the difference of our measure of deindustrialisation the Step-indicator Saturation algorithm at a 1% selection probability. We find a number of breaks and different regimes in which the United Kingdom has deindustrialised. Notably, since the 1990s deindustrialisation occurred at a very high pace of about 0.2 % per year. Following the devaluation of the pound when the United Kingdom left the European Rate Mechanism in 1992, the trend reversed for about two years, before reverting back to its old mean. Around 2001, when China has joined the world trade organisation, the

[1] The gross value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources.

[2] Rodrik (2016) points out that the level of deindustrialisation differs significantly when using current or constant prices. The former measure shows a stronger decline of manufacturing.

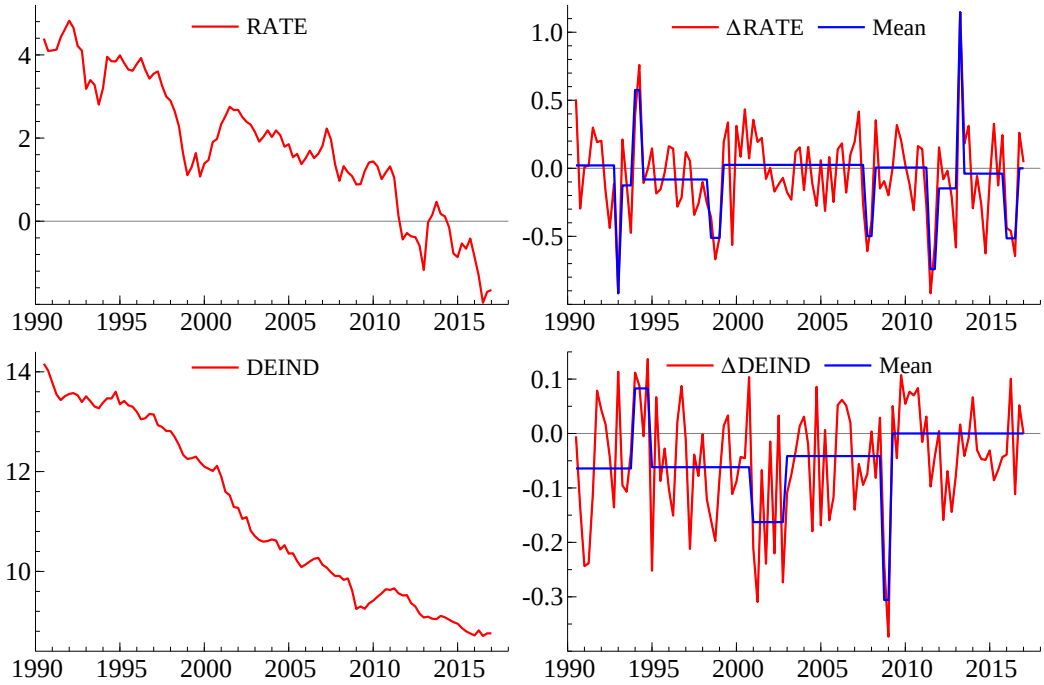



Figure 3.2: Levels and first-differences of the real interest rate and deindustrialisation from 1990:3 to 2017:1. The y-axis is in percent. The time-varying mean is fitted with Step-indicator Saturation at a 1% significance level.

deindustrialisation accelerated markedly, averaging up to 0.7 % per year. However, the mean reverted back close to the level in 1990. After the Great Recession, the rate of deindustrialisation decelerated.

While the share of manufacturing output declined, the absolute value of gross value added in manufacturing in constant 2013 prices has increased 35 percent since 1990. This is the consequence of rapid GDP growth which outweighed the effect of the declining share of manufacturing output.

### 3.2.2 Our measure of the real rate

We pick the five year ahead instantaneous forward rate estimated from the real government liability curve as our measure of the real rate. The forward rate is quoted on an annual basis. It corresponds to the instantaneous real return in five years that an investor would demand today. In this section, we discuss the

derivation of this measure and justify our choice.

The instantaneous forward rate is computed in two steps from the index-linked gilts issued by the UK Debt Management Office. The first step in the calculation of the instantaneous forward rate is to fit the real government liability curve based on index-linked gilts. The main challenge is that the coupon payments of index-linked gilts is indexed to the level of RPI eight months before the cash flow occurs. As a result, the index-linked gilts trade as a purely nominal bond for the final eight months. Evans (1998) derived a relationship between the nominal and real term structures and the term structure of (incompletely) indexed bonds. His methodology is used to fit the real government liability curve.

In the second step, the real government liability curve is expressed as a forward rate.[3] To do so, we define the forward rates as the yield that an investor would agree to today to make an investment over a pre-specified period in the future: for $m$-years beginning $n$ years ahead. The formula for the forward rate, can be expressed as

$$f_i(n,m) = \frac{1}{m}((n+m)y_i(n+m) - ny_i(n)), \tag{3.1}$$

where $y_i(n)$, the continuously compounded yield on a zero-coupon bond that matures in $n$ years and is measured in period $i$, is obtained from the real government liability curve. The instantaneous forward rate $n$ years ahead is the limit of (3.1) when $m$ goes to zero

$$f_i(n,0) = \lim_{m\to 0} f_i(n,m) = y_i(n) + ny_i'(n). \tag{3.2}$$

We call the five year ahead instantaneous forward rate $\text{RATE}_i = f_i(5,0)$. We obtain the data from the Bank of England, which regularly publishes this time series. The advantage of using this measure is that it is a market driven measure of the real rate that does not contain the very short end of the yield curve, which is directly affected by the policy rate decisions of the Bank of England. We choose to use $n = 5$ years, because it is the smallest integer $n$, for which data is continuously available.

Figure 3.2, panel a and b, plot the time series of the forward rate. We see that the forward rate falls from around 4 percent to close to -2 percent. As with

[3]This is a brief summary of the detailed exposition on the forward rates that can be found in the working paper version of Gürkaynak et al. (2007).

our measure for deindustrialisation, we apply Step-indicator Saturation at a 1 % level to the first difference of our measure of the real interest rate. We find that outliers dominate the time-series: around 1992 when the United Kingdom has left the European Rate Mechanism, during the Great Recession in 2008 and then again during the Euro Crisis in 2001 and the Brexit referendum in 2016.

The United Kingdom was the first country to issue an index-linked government bond in 1981.[4] From 1985 onwards a sufficient number of these bonds have been issued to fit the real government liability curve. In contrast, the United States issued its first index-linked bond in 1997. Britain therefore has the best historical data on market based measures of the real rates.

### 3.2.3 Reduced form regressions

Before we discuss our theoretical macroeconomic model and turn to a causal relationship between the variables, we show the reduced form evidence. This is based on a reduced form regression model of $\Delta\text{RATE}_i$ which is

$$\begin{aligned}\Delta\text{RATE}_i = \mu_{r\cdot d} + \phi\Delta\text{DEIND}_i + \beta_{rr\cdot d}\Delta\text{RATE}_{i-1} + \beta_{rd\cdot d}\Delta\text{DEIND}_{i-1} \\ +\gamma_{rr\cdot d}\text{RATE}_{i-1} + \gamma_{rd\cdot d}\text{DEIND}_{i-1} + \delta_1 S2000Q4_i + \delta_2 S2014Q4_i + \varepsilon_i,\end{aligned}$$

where $S2000Q4_i$ and $S2014Q4_i$ are step-indicators which have been illustrated in Figure 2.1 and which have breakpoints in Q4 2000 and Q4 2014 respectively.

Table 3.1 showcases the results of the reduced form regressions. In the individual columns we add different subsets of the step-indicators. We find that the inclusion of the step-indicators is essential in uncover the relationship between deindustrialisation and the real rate. Without any step-indicator, the coefficient on $\Delta\text{DEIND}_i$ is highly insignificant with a p-value of 36%. This changes dramatically, when we include either just the 2000:4 step-indicator, or both the 2000:4 and the 2015:4 step-indicator, as $\Delta\text{DEIND}_i$ becomes highly significant with a p-value

[4]The index-linked gilts were issued following the Wilson Report which was published in 1980. It recommended the government to issue inflation-indexed gilts which would be exclusively sold to pension funds. We see the introduction of the index-linked gilts associated with the desire of governments that have experienced the inflation surge of the 1970s, to introduce a commitment device to back their own determination to reduce inflation.

Table 3.1: Regression table with the reduced form regressions. Sample from 1990:3 to 2017:1. P-values in reported parenthesis.

| **The dependent variable is $\Delta$RATE$_i$** | | | |
|---|---|---|---|
| $\Delta$DEIND$_i$ | 0.28 | 0.69* | 0.77* |
| | (0.36) | (0.03) | (0.01) |
| RATE$_{i-1}$ | -0.10* | -0.21** | -0.26** |
| | (0.03) | (0.00) | (0.00) |
| DEIND$_{i-1}$ | 0.09* | 0.35** | 0.38** |
| | (0.02) | (0.00) | (0.00) |
| SIS dummies: | | 2000:4 | 2000:4 |
| | | | 2015:4 |
| **Diagnostic Tests (p-value)** | | | |
| AR 1-5 test: | 0.23 | 0.95 | 0.95 |
| ARCH 1-4 test: | 0.95 | 0.93 | 0.78 |
| Normality test: | 0.01* | 0.01* | 0.03* |
| Hetero test: | 0.33 | 0.54 | 0.34 |
| RESET23 test: | 0.89 | 0.43 | 0.41 |
| Adj. $R^2$: | 0.03 | 0.12 | 0.17 |
| Log-Likelihood: | -27.03 | -21.28 | -17.70 |
| Short-run fluctuations: | 1 lag | 1 lag | 1 lag |
| Constant: | Yes | Yes | Yes |

** significance at 1%, * significance at 5%

of 3% and 1% respectively. As the diagnostic test show, the models are reasonably specified. Only the normality test rejects at 3%, once we include the two step-indicators. Adding an impulse indicator for 2013:2 is sufficient to overcome a rejection of normality.

These preliminary results highlight the usefulness of Step-indicator Saturation in empirical studies. The relationship between deindustrialisation and the real rate would otherwise be hidden in a univariate regression equation. By searching for location shifts, we have uncovered this formerly hidden relationship.

The coefficients of $\text{RATE}_{i-1}$ and $\text{DEIND}_{i-1}$ are highly significant which hints at co-integration. The statistical inference about the presence of co-integration is non-standard because we have searched for location shifts with Step-indicator Saturation. While it is a fascinating topic, it is outside the scope of this paper to develop the theory for testing for co-integration while we apply Step-indicator Saturation.

Figure 3.3 shows a scatter-plot of the $\Delta\text{RATE}_i$ and $\Delta\text{DEIND}_i$ variable. Both axis are in percent. The year labels refer to the first quarter of the year. The dotted lines refer to the three conjectured long-run trends in the co-integrating relationship which are identified by Step-Indicator Saturation in 2000:4 and 2015:4. There appear to be three regimes where $\Delta\text{RATE}_i$ and $\Delta\text{DEIND}_i$ fluctuate around these long-run trends within the regimes.

The reduced form evidence is a two variable analysis that showcases a correlation between two time-series, while suggesting co-integration. We will in turn develop a model in which terms of trade shocks act as exogenous shocks that affect both variables. Then, we will use the catalyst method to establish the causal transmission of the terms of trade shocks from $\text{DEIND}_i$ to $\text{RATE}_i$.

## 3.3 Theory

This section presents a theoretical model that motivates the observation that deindustrialisation, which we define as a decline in the manufacturing share of output, is associated with a lower real rate. Our model is based upon the trade shock that follows the industrialisation of emerging economies. While a number of emerging

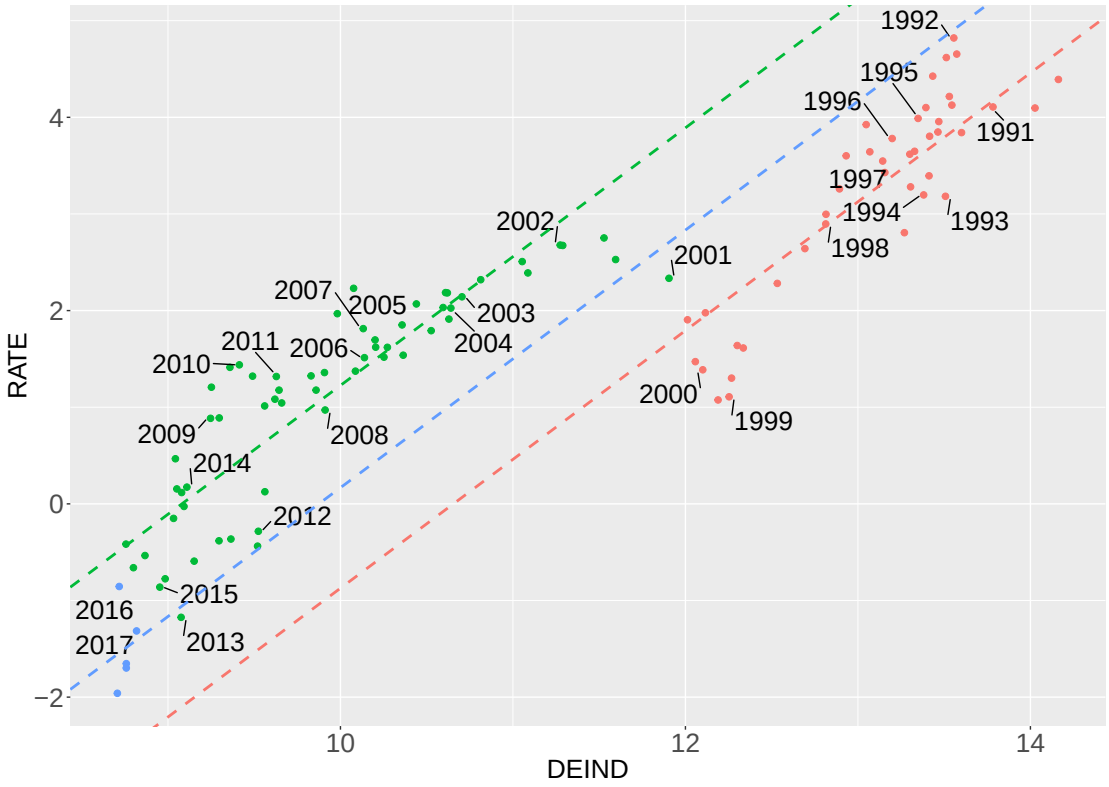



Figure 3.3: A scatter-plot of the real interest rate and deindustrialisation from 1990:3 to 2017:1. Both axis are in percent. The year labels refer to the first quarter of the year. The dotted lines refer to the three conjectured long-run trends in the co-integrating relationship which are identified by Step-indicator Saturation.

economies have industrialised since the 1990s, including among others Korea, Taiwan and China, it is the size of the Chinese economy that dominates the other emerging economies. We will therefore call the trade shock, the China trade shock. We discuss our model in three steps. First, we describe the general setting of our model. Second, we solve the model and show that the trade shock can simultaneously change the rate of industrialisation and the real rate. Third, we illustrate the channel in which the model works.

Observing both the steady decline in manufacturing output and employment of the advanced economies in the past decades, the academic literature offers three different explanations for these phenomena (Rowthorn and Ramaswamy (1997), Rowthorn and Ramaswamy (1998), Ngai and Pissarides (2004), Buera and Kaboski

(2009), Foellmi and Zweimüller (2008), Nickell et al. (2008), Beenstock (1982), Wood (1995)). The first two are domestic factors which relate to the supply and demand for manufactures. On the demand side, deindustrialisation is explained through an adopted Engel's Law argument: the proportion of income spend on manufactures declines as per capita income rises. It is therefore the high income elasticity of services that gives rise to the changing pattern of expenditures from manufactures to services. On the supply side, deindustrialisation is explained through the faster rate of productivity growth in the manufacturing relative to other sectors. Faster productivity growth translates into fewer factor inputs to achieve the same production and lowers relative prices of manufactures. As long as the elasticity of substitution is sufficiently small, this results in a decline in the share of manufacturing employment. To complicate matters, the supply side explanation depends on whether or not the economy is open to trade (Matsuyama (2009)). In an open economy, faster productivity gains need not be correlated with a more rapid decline in manufacturing employment across countries, even if productivity change is globally responsible for manufacturing employments' decline. This is because productivity gains can cause a shift in comparative advantage towards manufacturing that offsets its domestic effect. The third explanation for deindustrialisation is rooted in the trade between advanced and developing countries, which it labels the 'North-South trade'. According to this explanation, the manufactured imports from the developing countries are highly labour intensive, and displace many times more workers in the advanced economies than their dollar value would suggest. This argument implies that North-South trade could result in large job losses for unskilled workers in advanced economies even when the import penetration ratio is low, and trade between the two groups is balanced.

In our exposition, we follow the idea of the well-established literature on the effect of North-South trade on deindustrialisation (Beenstock (1982), Wood (1995)). However, we do this with an emphasis on manufacturing output and the real rate.

### 3.3.1 The setting of the 1-2-3 model

In this section, we present the setting of our general equilibrium model. We model the effect of the China trade shock on the United Kingdom by considering its effect

on the international price of manufactured goods and thus the price of the British imports and exports. Due to the comparative advantage of China in the production of manufactured goods and its abundance of cheap labour, the integration of China into the world markets reduced the international price for manufactured goods. The lower international prices for manufactured goods benefit the United Kingdom through lower prices for manufacturing imports and at the same time reduce the prices that domestic manufacturers can get for their exports. We then make the assumption that the lower price of manufactured goods affected the price of British exports more than the price of imports. Therefore, we assume that the Chinese exports drive a wedge between the price of imports and exports and result in a deterioration in the terms of trade in Britain.

We study the consequences of the China trade shock using a 1-2-3 model which was introduced in Devarajan et al. (1990) and Devarajan et al. (1993). It contains one country, two domestic productive sectors and three goods. Each of the two domestic sectors produce a distinct good. The export sector ($E$) produces a manufacturing good which is sold to foreigners and not consumed domestically. The service sector ($S$) produces a non-traded service good which is only consumed domestically. The third good, which is the imported manufacturing good ($I$), is produced overseas and consumed domestically.[5] The country is a small open economy, taking world prices for exports $p_E$ and imports $p_I = 1$ (the numeraire) as given. Only the price of services $p_S$ is determined endogenously in general equilibrium. In our 1-2-3 model we make a further simplifying assumption, which takes the wedge between export and import prices to an extreme: we assume that the China trade shock only affects export prices and that it leaves import prices untouched.[6]

In line with the literature in trade economics, such as Neary (1978), we assume sector specificity of one factor input. Yet, we deviate from this literature in that

[5]This model does not consider intermediate goods and the global value chain (GVC) literature. Over the past decades, the fragmentation of production across firms and countries has changed the nature of international trade. It is up to future research to study the implications of GVCs on the financial markets.

[6]The data for the United Kingdom over the past three decades show intermittent periods of a deterioration of the terms of trade for goods. There have been three episodes of a deterioration of the terms of trade by more than five percent which are from 1993 to 1995, 1998 to 2000 and 2005 to 2006. Over the long run, however, the terms of trade for goods has been relatively stable.

we discuss a model with labour as the specific factor. Capital is assumed to flow freely between the export and the service sector. The interest rate is perfectly flexible and not subject the distortions. This ensures that capital is always fully utilized. The sector specificity of labour can be justified by the existence of two types of labour: skilled and unskilled. Say, the manufacturing plants of the export sector only require unskilled workers while the service sector with its banks and law firms only require skilled workers. In considering labour specificity we modify the model of Devarajan et al. (1990). We also note that our 1-2-3 model is a purely real model and that it does not consider monetary effects. The prices we determine are relative prices, with the import good being the numeraire. Finally, the national output equals expenditure, so trade is balanced.

### 3.3.2 The algebra of the 1-2-3 model

We next discuss the key equations of our 1-2-3 model and solve it for the variables of interest. The service and export goods are supplied in quantities $Q_S^S$ and $Q_E^S$. The domestic consumer demand for services and imports are $Q_S^D$ and $Q_I^D$. We denote the return on capital as the interest rate $r$. For the sector specific return on labour we write the wage $w_S$ and $w_E$. Both factors labour and capital are fully utilized. We follow the notation of Jones (1965) to write,

$$a_{KS}Q_S^S + a_{KE}Q_E^S = K, \quad a_{LS}Q_S^S = L_S, \quad a_{LE}Q_E^S = L_E, \tag{3.3}$$

where $a_{ij}$ is the factor $i$ input for each output of good $j$. We can totally differentiate the full factor utilisation equations and express them as rates of change ($\wedge$):

$$0 = \lambda_{KS}(\hat{a}_{KS} - \hat{a}_{LS}) + \lambda_{KE}(\hat{a}_{KE} - \hat{a}_{LE}), \tag{3.4}$$

where $\lambda_{Kj}$ is the proportion of capital used in sector $j$. Note that $\lambda_{Kj} > 0$ and that $\lambda_{KS} + \lambda_{KE} = 1$. The two expressions in the parenthesis of equation (3.4) can be related to the change in the relative factor prices using the elasticity of substitution between labour and capital as

$$\hat{a}_{Kj} - \hat{a}_{Lj} = -\sigma_j(\hat{r} - \hat{w}_j), \quad j = \{S, E\}. \tag{3.5}$$

We assume a perfectly competitive economy, such that the price equals its factor costs. We then have,

$$\hat{p}_S = \theta_{KS}\hat{r} + \theta_{LS}\hat{w}_S, \tag{3.6}$$

$$\hat{p}_E = \theta_{KE}\hat{r} + \theta_{LE}\hat{w}_E, \tag{3.7}$$

where $\theta_{ij}$ is the share of factor $i$ in the value of output $j$, such that

$$\theta_{Kj} = \frac{a_{Kj}r}{p_j}, \quad \theta_{Lj} = \frac{a_{Lj}w_j}{p_j}. \tag{3.8}$$

Note that $\theta_{KE} > \theta_{KS}$ and $\theta_{LS} > \theta_{LE}$, as we assume the export sector to be more capital intensive and the service sector to be more labour intensive. Moreover, we have $\theta_{Lj} + \theta_{Kj} = 1$. Note that the only endogenous price is $p_S$, as $\hat{p}_E$ expresses the magnitude of the China trade shock which is determined exogenously. We combine equation (3.5) to (3.7) with the full factor utilisation equation (3.4) to get our first equation that relates the interest rate to the price of services and the China trade shock

$$\hat{r} = \frac{\xi_S}{\xi_S + \xi_E}\hat{p}_S + \frac{\xi_E}{\xi_S + \xi_E}\hat{p}_E, \tag{3.9}$$

where $\xi_S + \xi_E$ is the interest rate elasticity of the aggregate demand for capital and where

$$\xi_S = \frac{\lambda_{KS}\sigma_S}{\theta_{LS}}, \quad \xi_E = \frac{\lambda_{KE}\sigma_E}{\theta_{LE}}. \tag{3.10}$$

Next, we model the demand and supply side of the economy. Our demand equation is based on the elasticity of demand between the imports and the services and with $\hat{p}_I = 0$, we get

$$\hat{Q}_S^D - \hat{Q}_I^D = -\sigma_D\hat{p}_S. \tag{3.11}$$

For the supply side of the domestic production sectors, we use the property of Cobb-Douglas production functions, that the ratio of the relative change in the quantity of the output in one sector to the input of the quantity of the flexible factor into that sector is equal to the share of the flexible factor in the value of the output,

$$\hat{Q}_j^S = \theta_{Kj}\hat{K}_S, \quad j = \{S, E\}. \tag{3.12}$$

After a few algebraic manipulations, we get the supply equations for the service and export sector as

$$\hat{Q}_S^S = \phi_S(\hat{p}_S - \hat{r}), \tag{3.13}$$

$$\hat{Q}_E^S = \phi_E(\hat{p}_E - \hat{r}), \tag{3.14}$$

where

$$\phi_S = \frac{\theta_{KS}\sigma_S}{\theta_{LS}}, \quad \phi_E = \frac{\theta_{KE}\sigma_E}{\theta_{LE}}. \tag{3.15}$$

We then use the market clearing equilibrium conditions. Supply must equal demand for the service sector with $\hat{Q}_S^S = \hat{Q}_S^D$. Use the balance-of-trade constraint

$$\hat{p}_E + \hat{Q}_E^S = \hat{Q}_I^D, \tag{3.16}$$

to get an explicit expression for the demand for imports

$$\hat{Q}_I^D = \hat{p}_E(1 + \phi_E) - \hat{r}\phi_E. \tag{3.17}$$

We plug the equations for $\hat{Q}_I^D$ and $\hat{Q}_S^D$ into (3.11) to get the second equation that relates the interest rate to the price of services and the China shock as

$$\hat{r}(\phi_S - \phi_E) = \hat{p}_S(\sigma_D + \phi_S) - \hat{p}_E(1 + \phi_E). \tag{3.18}$$

We are now ready to solve the model for the variables of interest. We first solve for the interest rate and the relative price of services using (3.9) and (3.18) to get

$$\hat{r} = \hat{p}_E\psi_r \text{ and } \hat{p}_S = \hat{p}_E\psi_p, \tag{3.19}$$

with $\psi_r > 0$ and $\psi_p > 0$ and

$$\psi_r = \frac{\xi_S(1 + \phi_E) + \xi_E(\sigma_D + \phi_S)}{\xi_S(\sigma_D + \phi_E) + \xi_E(\sigma_D + \phi_S)}, \quad \psi_p = \frac{\xi_S(1 + \phi_E) + \xi_E(1 + \phi_S)}{\xi_S(\sigma_D + \phi_E) + \xi_E(\sigma_D + \phi_S)}. \tag{3.20}$$

This shows that following a reduction in the price of exports, both the real rate and the price of services fall. The intuition for the fall in the real rate is as follows. The real rate falls as the results of a combination of two forces. First, the lower price of manufacturing goods reduces the value marginal product of capital in the export sector. This can be thought of a negative productivity shock in the export sector, which pushes capital out of the export sector into the domestic service

sector. Second, in equilibrium, the price of services falls, which in turn reduces the value marginal product of capital in the service sector, which further reduces the real rate. In the next section, we further illustrate the transmission channel of the trade shock within the model using two different illustrations.

We then turn to the decline of the export sector, which is the domestic manufacturing sector. To solve for the change in output by sector, we can use the supply equations and plug in the expressions for the interest rate and the price of services to get

$$\hat{Q}_S = \hat{p}_E \frac{\xi_E \phi_S (1 - \sigma_D)}{\xi_S(\sigma_D + \phi_E) + \xi_E(\sigma_D + \phi_S)}, \tag{3.21}$$

$$\hat{Q}_E = \hat{p}_E \frac{\xi_S \phi_E (\sigma_D - 1)}{\xi_S(\sigma_D + \phi_E) + \xi_E(\sigma_D + \phi_S)}. \tag{3.22}$$

The sign of the relative output change therefore depends on the elasticity of substitution $\sigma_D$ of the demand for imports and services. If $\sigma_D > 1$, the China trade shock will result in deindustrialisation and an expansion of the services sector. However, if demand is relatively inelastic, then the opposite is true: the export sector will expand and the service sector contracts. There is also the chance that $\sigma_D = 1$ in which case the output in the two sectors does not change.[7]

The changes in the size of the sectors due to the lower price of exports can be explained by considering the income and substitution effect. The income effect increases the size of the export sector, as the China shock reduces real domestic income and services have a large and positive income elasticity. The substitution effect increases the size of the service sector, as domestic goods become relatively cheaper. When $\sigma_D < 1$, the income effect dominates and vice versa.

The empirical literature provides some, albeit limited, support for the elasticity of substitution between traded and nontradable goods $\sigma_D$ to be larger than one. Cashin and McDermott (2003) use the generalized method of moments using annual time series and report for the United Kingdom from 1970 to 1994 an estimate of $\hat{\sigma}_D = 1.7$. However, using a single equation regression model on a cross

[7]If $\sigma_D > 1$, then the $Q_E$ in the new equilibrium is smaller. At the same time, the share of $Q_E$ to total output $Q = Q_E + Q_S$ is also smaller. The 1-2-3 model can therefore not reproduce the situation in which the value of $Q_E$ raises while the share of $Q_E$ to $Q$ falls. But the trade shock, with $\sigma_D > 1$, will result in a lower manufacturing share of output and therefore deindustrialisation.

sectional dataset of 13 industrialized countries, Stockman and Tesar (1995) report an estimate of $\hat{\sigma}_D = 0.44$. As Cashin and McDermott (2003) report substantial differences in their estimates across countries, while the results in Stockman and Tesar (1995) are an average of the elasticities across a cross section of industrialized countries, we tend to put more weight on the former study, which suggests that traded and non-traded goods are substitutes.

### 3.3.3 A graphical illustration of the 1-2-3 model

Figure 3.4, which we adapt from Devarajan et al. (1997), uses four quadrants to illustrate the effect of the China trade shock in our 1-2-3 model. In the north-east quadrant, we show the balance-of-trade constraint. This constraint changes with the falling price of exports, which corresponds to a clockwise rotation. In the north-west quadrant, we plot the consumption possibility frontier, which are all the bundles of imported goods and domestic services that can be supplied. The change in the terms of trade shift the consumption possibility frontier inwards, as less imports become affordable. The utility function that we model as the demand equation, determine the equilibrium demand for services and imports, as well as the relative price of $p_S$ and $p_I$. The depicted situation corresponds to $\sigma_D > 1$, as the consequence of the China trade shock is a higher consumption of domestic services. The consumption on imports falls at the new equilibrium. This is reflected from the shift of the original equilibrium in $C$ to the new equilibrium in $C'$.

In the south-eastern quadrant, we find the production possibility frontier of the bundles of domestic services and export goods that can be produced given the production function. The equilibrium from the north-western quadrant maps to the equilibrium in the productive sectors and determines the relative price of $p_S$ and $p_E$. The deteriorated terms of trade shift the equilibrium in the productive sectors from $P$ to $P'$. This corresponds to a lower production of the export good and a higher production of the domestic services.

If the deterioration in the terms of trade is sufficiently big, then the consumption of imports would fall for any $\sigma_D$. However, the lower demand for imports has two different effects on the supply of exports, depending on $\sigma_D$. With $\sigma_D < 1$,

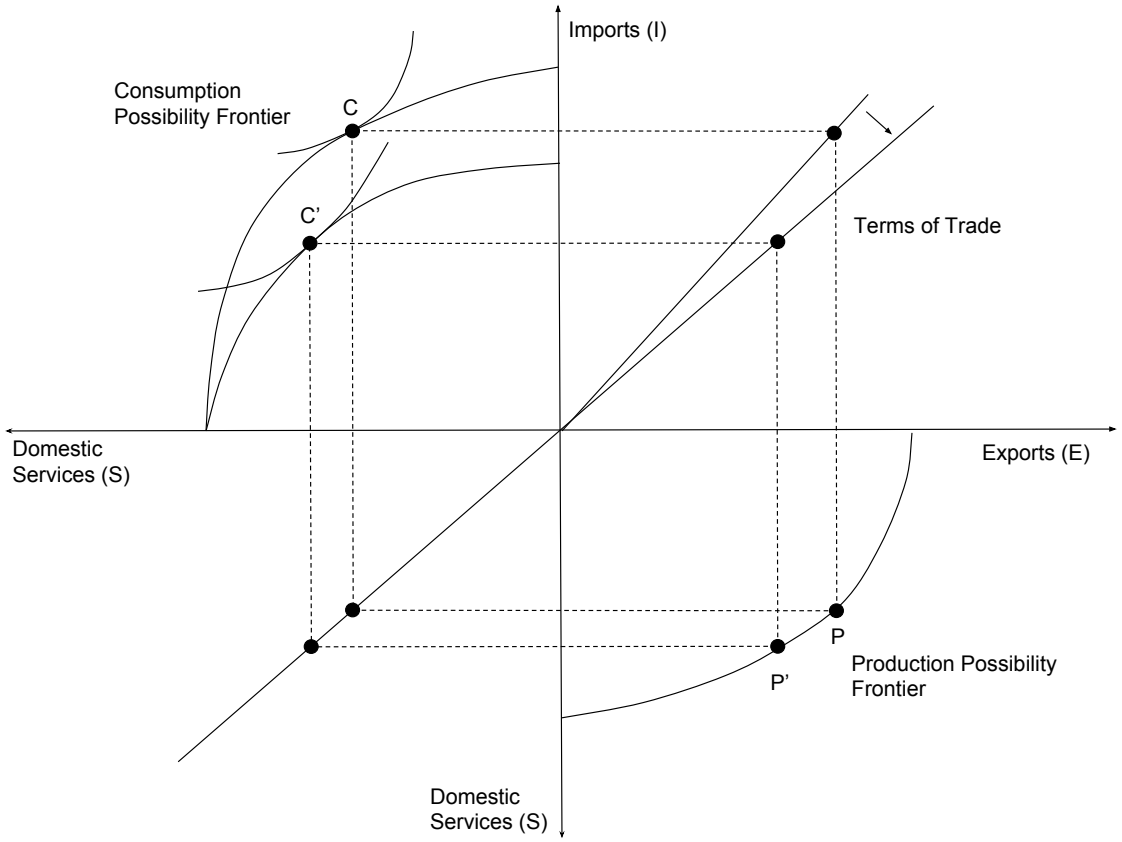



Figure 3.4: The effect of the China shock on output and consumption

the output of exports increases despite the fall in the demand for imports. This reflects the case in which the increased production of exports compensates for the higher relative price of imports.

Figure 3.5 illustrates the effect of the China trade shock on the capital market. This figure follows the ideas in Neary (1978). The interest rate is measured on the vertical axis and the total supply of capital is given by the horizontal axis $O_S O_E$. Capital input into services is measure by the distance from $O_S$ while the distance from $O_E$ measures the input into the export sector. Assuming that the interest rate adjusts to ensure full employment of capital at all times, the initial interest rate and capital allocation is determined by the intersection of the two value marginal products of the capital schedule, $V_S$ and $V_E$, at $A_0$ in the upper part of figure 2. The fall of the price of exports with the China trade shock shifts $V_E$ to $V_E'$. However, as we have shown earlier, the deterioration in the terms of

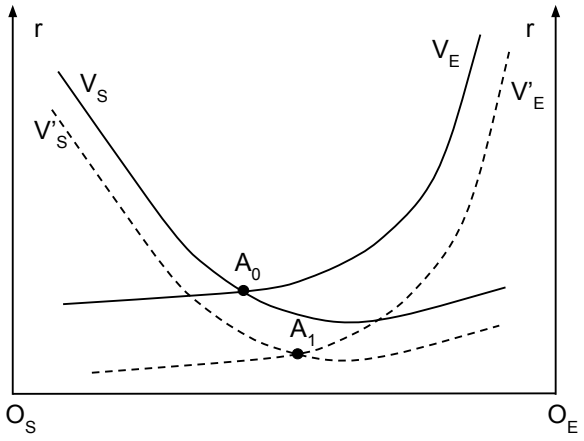



Figure 3.5: The effect of the China shock on the domestic capital market

trade also reduce the price of services. Depending of the magnitude of this effect (which depends on $\sigma_D$), the shift of $V_S$ to $V_S'$ can result in more or less capital being allocated to the export sector. The graph depicts the case of $\sigma_D > 1$. At the new equilibrium $A_1$, the capital allocated to the export sector falls, which corresponds to lower output in this sector.

We have shown that the China trade shock can give rise to the concurrent decline of the share of manufacturing output and the real rate. In our analysis, we found that the elasticity of substitution between services and imports is a critical parameter that determines whether or not we observe deindustrialisation. The decline of the real rate does not depend on any parameter. It is a consequence of both capital flowing out the export sector after the lower price on exports has decreased its value marginal product of capital and the decrease in the relative price of services.

To sum up, we have developed a theoretical model that motivates the positive correlation between the deindustrialisation and the real rate using terms of trade shocks that hit the British economy from the shift of the location of the manufacturing supply chains from high income countries to industrialising emerging economies. The theoretical model can only explain the positive correlation between the deindustrialisation and the real rate. It does not introduce a causal structure between the two variables. In the empirical section we will go a step further and study the causal direction in which the terms of trade shocks transmit

between the two variables of interest, by establishing the causal transmission of the terms of trade shocks from deindustrialisation to the real rate.

## 3.4 Empirical strategy

In the traditional literature in econometrics, there exist broadly speaking two empirical methods to show causality. First, an experimental or quasi-experimental setting delivers exogenous variation of the regressors that is then interpreted as the causal effect of a treatment on the outcome variable. This is considered to be the gold standard for studying causal effects. If experiments cannot be implement due to ethical or practical limitations, the method of instrumental variables can be used instead. Here the exogenous variation of the regressor derives from an instrument variable which is correlated with the regressor, but otherwise plausibly unrelated with the dependent variable.

In this section, we discuss a third way to attempt a study of causal relationships between economic variables. Our empirical strategy to identify the causal effect of deindustrialisation on the real rate exploits the causal transmission of catalysts. The method was first introduced in Bazinas and Nielsen (2015). It is based on the transmission of catalysts, large exogenous shocks, between the two endogenous variables of interest.

The transmission of catalysts reveals a direction for the empirical relationship between the endogenous variables. The otherwise non-directional correlation of these two variables, combined with the direction of influence identified by the catalysts, gives rise to a causal interpretation of their otherwise non-directional correlation. In other words, it is the combination of correlation and a direction that gives rise to causality.

As catalysts we use a collection of step-indicators which are constructed from the British terms of trade. We construct the catalysts in three stages. First, we search for location shifts in the terms of trade variable. Then, we identify all subsets of the step-indicators that model these location shifts that satisfy the conditions for causal transmission. In the third step, we pick the catalyst using a loss function that evaluates all subsets of step-indicators that satisfy the conditions for causal transmission.

The idea to use exogenous shocks for identification has also been considered in Magnusson and Mavroeidis (2014). They develop new limited-information methods of inference that exploit the identifying information contained in the structural breaks of the data.

We explain the identification strategy in two steps. First, we discuss the idea behind the causal transmission of shocks and explain how to use hypothesis tests to establish causal transmission. Second, we discuss the construction of the catalysts that we use in our empirical analysis. The final part of this section discusses the impulse indicator saturation algorithm which we use to detect outliers in each of the regression models that we estimate.

### 3.4.1 The structure of causal transmission

Causal transmission sets out a unique causal ordering based on the transmission of large exogenous shocks through the endogenous variables of interest. It gives rise to a causal interpretation of an otherwise non-directional correlation. Based upon the concept of super exogeneity in Engle et al. (1983) and Santos and Hendry (2010), the concept of causal transmission was first introduced in Bazinas and Nielsen (2015). In this section, we motivate the use of causal transmission and then explain the statistical hypothesis tests which are used to establish causal transmission.

We motivate the use of causal transmission in a bivariate statistical framework. We are interested in the relationship of the real rate $\Delta$RATE and our measure of deindustrialisation $\Delta$DEIND, given a catalyst $C$ which represents a collection of large exogenous shocks that hit the economy. If we assume that the two variables are jointly normal distributed, we can write their conditional distribution as

$$\begin{pmatrix} \Delta\text{RATE} \\ \Delta\text{DEIND} \end{pmatrix} = \begin{pmatrix} \mu_{rc} \\ \mu_{dc} \end{pmatrix} C + \begin{pmatrix} \varepsilon_r \\ \varepsilon_d \end{pmatrix}, \tag{3.23}$$

where the innovations are normally distributed, with positive definite variance:

$$\begin{pmatrix} \varepsilon_r \\ \varepsilon_d \end{pmatrix} \overset{\mathsf{D}}{=} \mathsf{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{rr} & \sigma_{rd} \\ \sigma_{dr} & \sigma_{dd} \end{pmatrix} \right\}. \tag{3.24}$$

The catalyst $C$ does not yet introduce any direction, because there are two different ways of ordering $\Delta$RATE and $\Delta$DEIND, corresponding to two Cholesky

decompositions. We could either condition $\Delta$RATE on $\Delta$DEIND as

$$\Delta\text{RATE} = \gamma_{rd}\Delta\text{DEIND} + \gamma_{rc\cdot d}C + \varepsilon_{r\cdot d}, \tag{3.25}$$

$$\Delta\text{DEIND} = \mu_{dc}C + \varepsilon_{d}, \tag{3.26}$$

where the innovations are independent and normal, or we could condition $\Delta$DEIND on $\Delta$RATE as

$$\Delta\text{DEIND} = \gamma_{dr}\Delta\text{RATE} + \gamma_{dc\cdot r}C + \varepsilon_{d\cdot r}, \tag{3.27}$$

$$\Delta\text{RATE} = \mu_{rc}C + \varepsilon_{r}. \tag{3.28}$$

However, we obtain a unique direction through a set of parameter restrictions, namely the restriction $\gamma_{rc\cdot d} = 0$ and $\mu_{dc} \neq 0$. The equations (3.25) and (3.26) then reduce to

$$\Delta\text{RATE} = \gamma_{rd}\Delta\text{DEIND} + \varepsilon_{r\cdot d}, \tag{3.29}$$

$$\Delta\text{DEIND} = \mu_{dc}C + \varepsilon_{d}. \tag{3.30}$$

The restrictions imply that the catalyst first affects the degree of industrialisation which in turn causes the real rate to change. Importantly, the restrictions rule out the opposite direction of causal transmission. If the given restrictions hold, it cannot be that $\gamma_{dc\cdot r} = 0$. This result is proven in Theorem 2.2 of Bazinas and Nielsen (2015). The uniqueness of the direction with which the catalyst transmits between the two variables then gives rise to the causal interpretation of $\gamma_{rd}$.

We statistically test the restrictions for the causal transmission in a system of three regression equations. In these regressions, we add additional lags to the variables of interest to capture time series dynamics. The first condition is the rejection of the null hypothesis of $\mathsf{H}_d : \mu_{dc} = 0$, with $i = 1, ..., n$, in

$$\begin{aligned}\Delta\text{DEIND}_i = \mu_d + \mu_{dc}C_i + \beta_{dd}\Delta\text{DEIND}_{i-1} + \beta_{dr}\Delta\text{RATE}_{i-1} \\ + \gamma_{dd}\text{DEIND}_{i-1} + \gamma_{dr}\text{RATE}_{i-1} + \varepsilon_{di}.\end{aligned} \tag{3.31}$$

We call this equation the marginal equation on $\text{DEIND}_i$. Intuitively, we require that the catalyst significantly moves our measure of deindustrialisation.

The second condition is the rejection of the null hypothesis of $\mathsf{H}_r : \mu_{rc} = 0$ in

$$\begin{aligned}\Delta\text{RATE}_i = \mu_r + \mu_{rc}C_i + \beta_{rd}\Delta\text{DEIND}_{i-1} + \beta_{rr}\Delta\text{RATE}_{i-1} \\ + \gamma_{dd}\text{DEIND}_{i-1} + \gamma_{dr}\text{RATE}_{i-1} + \varepsilon_{ri}.\end{aligned} \tag{3.32}$$

We call this equation the marginal equation on $\text{RATE}_i$. The condition requires that the catalyst significantly moves the real rate as well.

The third, and most important condition, are based on the joint hypothesis $\mathsf{H}_{r\cdot d,1} : \phi = 0$ and $\mathsf{H}_{r\cdot d,2} : \mu_{rc\cdot d} = 0$ in

$$\begin{aligned}\Delta\text{RATE}_i = \mu_{r\cdot d} + \mu_{rc\cdot d}C_i + \phi\Delta\text{DEIND}_i + \beta_{rd\cdot d}\Delta\text{DEIND}_{i-1} \\ + \beta_{rr\cdot d}\Delta\text{RATE}_{i-1} + \gamma_{dd}\text{DEIND}_{i-1} + \gamma_{dr}\text{RATE}_{i-1} + \varepsilon_{r\cdot d,i}.\end{aligned} \tag{3.33}$$

We require that we reject $\mathsf{H}_{r\cdot d,1}$ and that we fail to reject $\mathsf{H}_{r\cdot d,2}$. When all three conditions are satisfied, we say that we have established causal transmission of the catalyst $C$ to $\Delta$RATE from $\Delta$DEIND.

### 3.4.2 Our catalyst

A catalyst is a strictly exogenous shock which hits the two endogenous variables of interest. The order in which it channels through the two variables then gives rise to the causal transmission, which helps us to identify the causal effect of deindustrialisation on the real rates. In this section, we discuss our choice of the catalyst and argue that the assumption of the strict exogeneity of our catalyst is satisfied.

#### 3.4.2.1 Catalysts from step-indicators

We follow the approach by Bazinas and Nielsen (2015) and use indicator variables as catalysts. In their paper, the authors summed up different impulse-indicators to construct the catalysts. Impulse indicators are vectors of length equal to the sample size of time-series that are zero except in one period, say period $\ell$: $\iota_\ell = (0, ..., 0, 1, 0, ..., 0)$. The sum of impulse-indicators would be non-zero for multiple time periods. For example, the authors constructed a catalyst for fiscal expansions which was non-zero in 1972:4-1973:1 and 1979:2. A second catalyst

that the authors considered was based on oil price shocks with non-zero entries in 1973:3-4 and 1979:3.

In our empirical analysis, we use step-indicators instead of impulse-indicators. Step-indicators have been illustrated in 2.1 and they are dummy variables that are one, for the say $\ell - 1$ first observations, before turning and staying at zero from the $\ell$'s observation onwards: $\rho_\ell = (0, ..., 0, 1, 1, ..., 1)$. If the catalyst represents not a single shock, but a collection of them, we can model each shock with a different step-indicator. We think of the collection of step-indicators as a single catalyst. The hypothesis to test for causal transmission, $\mathsf{H}_d$, $\mathsf{H}_r$ and $\mathsf{H}_{rc \cdot d,2}$ are then tested not as individual hypothesis tests on the individual regressors, but as joint hypothesis tests on the collection of step-indicator variables that jointly form the catalyst. For example, the hypothesis $\mathsf{H}_d : \mu_{dc} = 0$ changes from a hypothesis test for a single coefficient $\mu_{dc}$ to a test of joint hypotheses $\mathsf{H}_d : \mu_{dc,1} = \mu_{dc,2} = ... = \mu_{dc,K} = 0$, in

$$\begin{aligned}\Delta\text{DEIND}_i = \mu_d + \sum_{k=1}^{K} \mu_{dc,k} C_{i,k} + \beta_{dd} \Delta\text{DEIND}_{i-1} + \beta_{dr} \Delta\text{RATE}_{i-1} \\ + \gamma_{dd} \text{DEIND}_{i-1} + \gamma_{dr} \text{RATE}_{i-1} + \varepsilon_{di},\end{aligned}$$

where $K$ is the number of step-indicators contained in the catalyst. The joint hypothesis $\mathsf{H}_d$ is evaluated using the $F$-ratio which is distributed as $F(1, n - K)$.

#### 3.4.2.2 The construction of the terms of trade catalyst

We choose our catalyst based on the discussion of the theory section of this chapter. In our 1-2-3 model, we find that a deterioration in the terms of trade can result in deindustrialisation and a lower real rate. We therefore use the British terms of trade variable to construct the catalyst. The British term of trade are illustrated in Figure 3.6. As before, we fit the time-varying mean using Step-indicator Saturation at a 1% level to highlight the many location shifts to this time series.

We construct our catalysts in three stages. In the first stage, we collect the location shifts in the terms of trade variable, as highlighted in Figure 3.6. Because the exact step-indicators being retained by Step-indicator Saturation can vary by the specification of the regression model, we create a more comprehensive set of step-indicators, by running the marginal model of $\text{TOT}_i$ with a different number

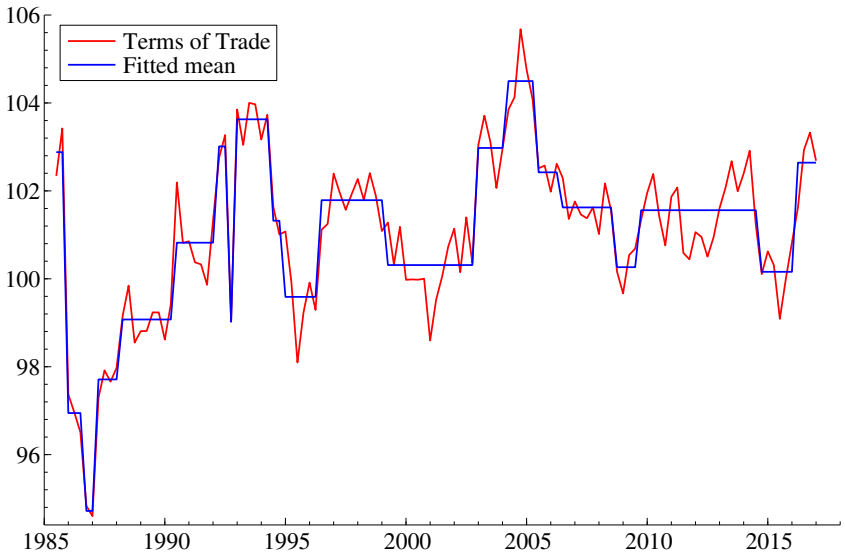



Figure 3.6: The British terms of trade from 1985:3 to 2017:1. The time-varying mean is fitted with Step-indicator Saturation at a 1% significance level.

of autoregressive lags (up to four). For example, the regression equation with the maximum number of four autoregressive lags would be

$$\text{ToT}_i = \mu + \sum_{j=1}^{4} \beta_j \text{ToT}_{i-j} + u_i.$$

We then collect all the step-indicators that are identified in each of the four regression equations. In total, we identify 20 step-indicators that form the space of variables from which we construct the catalyst variable.

Second, we search over the space of step-indicators that we identified in the first stage, to see which subset of them satisfy all of the four hypothesis for causal transmission, namely $\mathsf{H}_d$, $\mathsf{H}_r$, $\mathsf{H}_{r\cdot d,1}$ and $\mathsf{H}_{r\cdot d,2}$ discussed in Section 3.4.1. We accept all those collections of step-indicators that satisfy the four hypothesis at a five percent level. Each catalyst can only contain a limited number of step-indicators and we decide to limit our search at the maximum number of eight step-indicators. This threshold is also motivated by computational restrictions as the number of combinations of step-indicators grows exponentially with the number of maximum number of step-indicators we allow in the catalyst. In our case, we search around

263k combinations of step-indicators.[8] Our search has resulted in 75 candidate catalyst variables for which we cannot reject all four hypothesis for causal transmissions.

In the final third step we pick our preferred catalyst. This problem is complicated, because it depends on four separate hypothesis tests. It therefore relates to a multi-dimensional objective function. The search for the optimal catalyst could be conducted with an Autometrics type tree-search algorithm, yet the problem that Autometrics tackles is simpler, because it is limited to decide which variables to select and retain and which variables to drop. For Autometrics, the decision about each variable is based on a single hypothesis test. In future work, we would be interested to investigate the properties of a Autometrics type algorithm to pick the optimal catalyst. Yet for this chapter, we use an ad-hoc rule to choose our preferred catalyst. We first compute the p-value for each of the four hypothesis tests of causal transmission. Then, for each hypothesis, we rank the p-value for all the 75 previously identified candidate catalyst. If a hypothesis would be satisfied for a small p-value, we give the catalyst with the smallest p-value the smallest rank. In contrast, if the hypothesis would be satisfied for large p-values, we give the catalyst with the largest p-value the smallest rank. The loss function defined over each catalyst is then defined as the sum of the ranks for all four hypothesis that we consider:

$$\text{loss} = \text{rank}_{\mathsf{H}_d} + \text{rank}_{\mathsf{H}_r} + \text{rank}_{\mathsf{H}_{r\cdot d,1}} + \text{rank}_{\mathsf{H}_{r\cdot d,2}}.$$

The preferred specification is then the catalyst with the smallest loss. We call this catalyst the terms of trade catalyst $\mathrm{C}_{\mathrm{TOT,i}}$. It contains the following five step-indicators: 1995:1, 2004:1, 2004:4, 2008:3, 2009:2, 2011:2. Each of the step-indicators in our terms of trade catalyst relates to a structural break in the British terms of trade. Two of these breaks coincide with the period of the Great Recession in 2008 to 2009.

The assumption of strict exogeneity of the catalyst implies that each error term is uncorrelated with all past, present and future values of the catalyst. Following our theoretical model, we think of the terms of trade changes as exogenous shocks,

[8] A variation of Autometrics could be developed to reduce the complexity of the search-tree over our step-indicators. In this chapter, we do not consider a more intelligent search algorithm.

which are driven by the industrialisation of China, South Korea and Taiwan among other emerging economies. Because the British real rates or the state of its manufacturing sector had limited effects on the timing of the industrialisation of East Asian economies, the exogeneity of the terms of trade shocks is plausible. Moreover, the United Kingdom is usually considered to be a small open economy, which cannot influence the global prices of goods and services.

## 3.5 Empirical Results

In this section, we present and discuss our preferred specification that satisfies the conditions for the causal transmission of our terms of trade catalyst from our measure of deindustrialisation $\text{DEIND}_i$ to the real rates $\text{RATE}_i$. This specification establishes a causal ordering in which our catalyst $\text{C}_{\text{TOT,i}}$ first causes deindustrialisation, which in turn causes the real rates to adjust. The unique direction of the ordering gives rise to a causal interpretation of the coefficient on $\Delta\text{DEIND}_i$ in the conditional equation of $\Delta\text{RATE}_i$.

As discussed in section 3.4.1, the regression equations use the first differences of the two endogenous variables. From Figure 3.2, their first-differences exhibit broadly the patterns of a stationary time series. A concern may be the outliers and location shifts. We use Impulse-indicator Saturation and Step-indicator Saturation at the one percent significance level to alleviate the concern about outliers and location shifts.

For the terms of trade catalyst $\text{C}_{\text{TOT,i}}$, the p-value corresponds to the joint exclusion restriction for the six step-indicators in the catalyst. We omit the parameter estimates for the individual step-indicators.

Table 3.2 shows the results of the three regression models. We start with a discussion of the hypothesis $\mathsf{H}_d : \mu_{dc} = 0$. Here we look at the marginal model of $\Delta\text{DEIND}_i$. We find that we can reject $\mathsf{H}_d$ at a one percent significance level. Next, we consider the hypothesis $\mathsf{H}_r : \mu_{rc} = 0$. Looking at the marginal model of $\Delta\text{RATE}_i$ we find that we can reject $\mathsf{H}_r$ at a 4% significance level. Finally, the hypothesis $\mathsf{H}_{r\cdot d,1} : \phi = 0$ and $\mathsf{H}_{r\cdot d,2} : \mu_{rc\cdot d} = 0$ relate to the conditional model of $\Delta\text{RATE}_i$. We can reject $\mathsf{H}_{r\cdot d,1}$ at a 1% significance level, while we fail to reject $\mathsf{H}_{r\cdot d,2}$ at the 5% significance level. Therefore, all conditions for causal transmission are

Table 3.2: Regression table with the test for causal transmission using the ToT catalyst. Sample from 1990:3 to 2017:1. T-statistics in reported parenthesis.

| | **Marginal Model** | | **Conditional Model** |
|---|---|---|---|
| | $\Delta$RATE$_i$ | $\Delta$DEIND$_i$ | $\Delta$RATE$_i$\|$\Delta$DEIND$_i$ |
| $\Delta$DEIND$_i$ | | | 0.79* |
| | | | (0.01) |
| RATE$_{i-1}$ | -0.27** | 0.05** | -0.30** |
| | (0.00) | (0.01) | (0.00) |
| DEIND$_{i-1}$ | 0.27** | -0.16** | 0.39** |
| | (0.01) | (0.00) | (0.00) |
| $C_{TOT,i}$ | -* | -** | - |
| | (0.04) | (0.00) | (0.06) |
| SIS dummies: | 2000:4 | 2000:4 | 2000:4 |
| | 2015:4 | 2015:4 | 2015:4 |
| IIS dummies: | 1993:1 | 1993:1 | 1993:1 |
| | 2009:2 | 2009:2 | 2009:2 |
| | 2013:2 | 2013:2 | 2013:2 |
| **Diagnostic Tests (p-value)** | | | |
| AR 1-5 test: | 0.56 | 0.04* | 0.67 |
| ARCH 1-4 test: | 0.45 | 0.88 | 0.37 |
| Normality test: | 0.15 | 0.94 | 0.33 |
| Hetero test: | 0.77 | 0.96 | 0.40 |
| RESET23 test: | 0.20 | 0.40 | 0.04* |
| Adj. $R^2$: | 0.33 | 0.33 | 0.37 |
| Log-Likelihood: | -1.72 | 121.19 | 1.73 |
| Short-run fluctuations: | 2 lags | 2 lags | 2 lags |
| Constant: | Yes | Yes | Yes |

** significance at 1%, * significance at 5%
For the terms of trade catalyst $C_{TOT,i}$, the p-value corresponds to the joint exclusion restriction for the six step-indicators in the catalyst. We omit the parameter estimates for the individual step-indicators.

satisfied. However, the decisions regarding the catalyst in the conditional model are somewhat marginal and this reduces the strength of our empirical evidence. Ideally, we would have preferred the conditional variable $\Delta\text{DEIND}_i$ to pick up more of the significance level of $\text{C}_{\text{TOT,i}}$. Particularly in the light of p-value of $\text{C}_{\text{TOT,i}}$ in the marginal model of $\Delta\text{RATE}_i$ being already at 4% the increase in the p-value of $\Delta\text{RATE}_i$ after the inclusion of $\Delta\text{DEIND}_i$ is very limited.

The coefficient of $\Delta\text{DEIND}_i$ in the conditional equation of the $\Delta\text{RATE}_i$ is relatively large. A one percent reduction in the share of manufacturing output results in a 79 basis point reduction of the real rate. From 1990 to 2017, $\Delta\text{DEIND}_i$ dropped from 14.1 percent to 8.8 percent. Therefore, over the time frame of our study, deindustrialisation accounts for a 419 basis points reduction of the real rate. This represents nearly 70% of the 605 basis points reduction of the real rate and it suggests that the deindustrialisation is a major explanation for lower real rates.

In the analysis of the causal transmission of catalysts, the conditioning set of variables needs to stay the same, to be able to attribute the insignificance of the catalyst to the conditioning variable. Because of this, we add the step-indicators and impulse-indicators that our saturation methods find to all three models.

The diagnostic tests show that the models are reasonably well specified, despite containing only a small number of regressors. These diagnostic tests include the $\mathsf{F}_{ar}$ tests for $5^{th}$-order autocorrelation (Godfrey (1978) and Nielsen (2006)), $\mathsf{F}_{arch}$ tests for $4^{th}$-order autoregressive conditional heteroskedasticity (Engle (1982)), the $\chi^2(2)$ tests for non-normality (Doornik and Hansen (2008)), the $\mathsf{F}_{het}$ tests for heteroskedasticity (White (1980)) and the $\mathsf{F}_{reset}$ tests for functional-form misspecification (Ramsey (1969)). The adjusted $R^2$ incorporates a penalty for the number of regressors.

Overall, we have used the 1-2-3 model to show that terms of trade shocks can explain why deindustrialisation is associated with a lower the real rate. We use this idea in the empirical section, where we use the terms of trade variable to construct our term of trade catalyst. We show that terms of trade catalysts transmit with a unique direction through the two endogenous variables of interest: they first result in deindustrialisation which in turn cause lower the real rates.

## 3.6 Conclusion

In this chapter, we highlighted an important channel towards lower real rates, which economists have not yet studied: the decline of the manufacturing sector. We studied this problem from two perspectives. On the theoretical front, this chapter built on the existing literature of trade shocks and developed a theoretical framework that motivates why deindustrialisation is coupled with lower real rates. Empirically, we then established the causal transmission of the terms of trade shock from the manufacturing sector into the real interest rate.

This chapter raised a number of important questions for future research. First, we are interested to develop a valid statistical test for co-integration when we simultaneously search for location shifts with Step-indicator Saturation. This would allow us to go beyond speculating about the existence of a co-integrating relationship between the real rate and deindustrialisation. Second, we like to have a more sophisticated procedure to select the optimal catalyst from the pool of candidate catalysts. Ideally, this would involve a study of the properties of an algorithm that is similar to Autometrics, but that can also handle multi-dimensional objective functions. Third, in regard to theory models, we want to formulate a model that considers the ordering in which shocks transmit through the economic variables. This would help to match the theoretical models closer to the causal transmission of catalysts. Indeed, this chapter is a forward looking one which outlines a research agenda for the next years.

In the context of this thesis, the goal of the third chapter has been to present new empirical methods that are based on Step-indicator Saturation. We showed that Step-indicator Saturation is essential to uncover the otherwise hidden relationship between deindustrialisation and the real rate. Further, we used Step-indicator Saturation to construct the terms of trade catalyst. Step-indicator Saturation is an important contribution to how we build time-series models. It promises to be a central ingredient for future discoveries in the economic sciences.

# Conclusion

This thesis investigated the gauge of the split-half algorithm with Step-indicator Saturation and used this algorithm to study the relationship between deindustrialisation and the real interest rate. Step-indicator Saturation is needed if time-series data contains location shifts which are changes in the first moment of the data over time.

The mathematical centre piece of this thesis are the empirical process results for differenced residuals. Based on the structure of Johansen and Nielsen (2016*a*) we started with an exponential martingale inequality and used it to bound tail probabilities of a family of martingales. We used this bound to handle the terms that arise in the chaining argument that we constructed to show uniform convergence of our empirical process. An advantage of this proof architecture is that it can handle regressors with any form of time-series correlation, as long as the regressors are either stationary, random walk or trend stationary.

The empirical process results were used to study the gauge of the split-half algorithm with Step-indicator Saturation. We used the insight that the empirical gauge has an empirical process representation which allowed us to use our empirical process results. We showed that the empirical gauge converges to the population gauge which depends only on the distribution of the error terms. This results allows us to choose the only tuning parameter of Step-indicator Saturation to be set according to the tolerance level of the researcher for false positives. As false positives can reduce our confidence that identified location shifts are also in the data generating process, we recommended to target a low population gauge below 1%. We also derived asymptotic distribution results for the empirical gauge. These results describe the variation of the draws of the empirical gauge around the population gauge. We show that the asymptotic distribution of the empirical gauge

with Step-indicator Saturation is larger than the same measure with Impulse-indicator Saturation.

The Step-indicator Saturation estimator was applied to study the relationship between deindustrialisation and real interest rates. We advanced the hypothesis that the relocation of global supply chains to low labour cost countries in Asia has accelerated the deindustrialisation in the United Kingdom which in turn has lowered the real interest rate. Yet, without Step-indicator Saturation, their relationship is insignificant in a regression model of the real rate. The inclusion of only two step-indicators, however, that we identified with Step-indicator Saturation, has been sufficient to uncover a highly significant relationship between deindustrialisation and the real rate. We then studied the transmission of terms of trade shocks from the manufacturing sector to the real rate using the catalyst method. In the application of the catalyst method we used Step-indicator Saturation in the catalyst construction process.

Still there remain a number of open problems to solve, including those mentioned at the end of the individual chapters. Moreover, in the field of indicator saturation, alternatives have emerged from Multiplicative-indicator Saturation to Designed Break-indicator Saturation in Pretis, Schneider, Smerdon and Hendry (2016). We hope to continue to contribute to these methods by developing asymptotic theory and finding relevant applications.

# Bibliography


Ahumada, H. and Cornejo, M. (2016), Forecasting food prices: The case of corn, soybeans and wheat, *International Journal of Forecasting* **32**, 838–848.

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

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

Bai, C.-E., Hsieh, C.-T. and Qian, Y. (2006), The return to capital in China, *Brookings Papers on Economic Activity* **37**, 61–102.

Bazinas, V. and Nielsen, B. (2015), Causal transmission in reduced-form models, Discussion paper 7, Nuffield College.

Beenstock, M. (1982), *The World Economy in Transition*, Routledge, Oxford.

Beltran, D. O., Kretchmer, M., Marquez, J. and Thomas, C. P. (2013), Foreign holdings of US treasuries and US treasury yields, *Journal of International Money and Finance* **32**, 1120–1143.

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

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

Berenguer-Rico, V., Johansen, S. and Nielsen, B. (2018), Some new results on marked and weighted empirical processes of residuals, Working paper, University of Oxford.

Berenguer-Rico, V. and Nielsen, B. (2017), Marked and weighted empirical processes of residuals with applications to robust regressions, Working paper 841, Department of Economics, University of Oxford.

Bernanke, B. S. (2005), The global saving glut and the US current account deficit, Sandridge lecture, Virginia Association of Economics, Richmond, Va.

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

Billingsley, P. (1979), *Probability and Measure*, John Wiley & Sons, New York.

Brockwell, P. and Davis, R. (1991), *Time series: Data analysis and Theory*, Springer, New York.

Buera, F. J. and Kaboski, J. P. (2009), Can traditional theories of structural change fit the data?, *Journal of the European Economic Association* **7**, 469–477.

Cashin, P. and McDermott, C. J. (2003), Intertemporal substitution and terms-of-trade shocks, *Review of International Economics* **11**, 604–618.

Castle, J., Doornik, J., Hendry, D. F. and Pretis, F. (2015), Detecting location shifts during model selection by step-indicator saturation, *Econometrics* **3**, 240–264.

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

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

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

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

Devarajan, S., Go, D. S., Lewis, J. D., Robinson, S. and Sinko, P. (1997), Simple general equilibrium modeling, *in* J. F. Francois and K. A. Reinert, eds, 'Applied Methods for Trade Policy Analysis: a Handbook', Cambridge University Press, Cambridge.

Devarajan, S., Lewis, J. D. and Robinson, S. (1990), Policy lessons from trade-focused, two-sector models, *Journal of Policy Modeling* **12**, 625–657.

Devarajan, S., Lewis, J. D. and Robinson, S. (1993), External shocks, purchasing power parity, and the equilibrium real exchange rate, *The World Bank Economic Review* **7**, 45–63.

Doornik, J. A. (2009), Autometrics, *in* Castle and Shephard (2009).

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

Engle, R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987–1007.

Engle, R. F., Hendry, D. F. and Richard, J.-F. (1983), Exogeneity, *Econometrica* **51**, 277–304.

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

Ericsson, N. R. (2016), Eliciting GDP forecasts from the FOMC's minutes around the financial crisis, *International Journal of Forecasting* **32**, 571–583.

Ericsson, N. R. (2017), How biased are US government forecasts of the federal debt?, *International Journal of Forecasting* **33**, 543–559.

Evans, M. D. D. (1998), Real rates, expected inflation, and inflation risk premia, *Journal of Finance* **53**, 187–218.

Foellmi, R. and Zweimüller, J. (2008), Structural change, Engel's consumption cycles and Kaldor's facts of economic growth, *Journal of Monetary Economics* **55**, 1317–1328.

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

Giraitis, L., Koul, H. L. and Surgailis, D. (2012), *Large sample inference for long memory processes*, Imperial College Press, London.

Giuseppe, F., Gross, M. and Neri, S. (2017), On secular stagnation and low interest rates: demography matters, Working paper 2088, European Central Bank.

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

Gürkaynak, R. S., Sack, B. and Wright, J. H. (2007), The US treasury yield curve: 1961 to the present, *Journal of Monetary Economics* **54**, 2291–2304.

Hadi, A. S. (1992), Identifying multiple outliers in multivariate data, *Journal of the Royal Statistical Society, Series B* **54**, 761–771.

Hansen, A. H. (1939), Economic progress and declining population growth, *American Economic Review* **29**, 1–15.

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

Hendry, D. F. (2000), On detectable and non-detectable structural change, *Structural Change and Economic Dynamics* **11**, 45–65.

Hendry, D. F. and Doornik, J. A. (2014), *Empirical model discovery and theory evaluation: automatic selection methods in econometrics*, MIT Press, Boston.

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

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

Holston, K., Laubach, T. and Williams, J. C. (2017), Measuring the natural rate of interest: International trends and determinants, *Journal of International Economics* **108**, 59–75.

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

Huber, P. J. (1964), Robust estimation of a location parameter, *Annals of Mathematical Statistics* **35**, 73–101.

Jiao, X. and Nielsen, B. (2017), Asymptotic analysis of iterated 1-step Huber-skip M-estimators with varying cut-offs, *in* J. Antoch, J. Jurečková, M. Maciak and M. Pešta, eds, 'Analytic Methods in Statistics', Vol. 193 of *Springer Proceedings in Mathematics & Statistics*, Springer, New York.

Johansen, S. and Nielsen, B. (2009), Saturation by indicators in regression models, *in* Castle and Shephard (2009).

Johansen, S. and Nielsen, B. (2013), Outlier detection in regression using an iterated one-step approximation to the Huber-skip estimator, *Econometrics* **1**, 53–70.

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

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

Jones, R. W. (1965), The structure of simple general equilibrium models, *Journal of Political Economy* **73**, 557–572.

King, M. and Low, D. (2014), Measuring the "world" real interest rate, Working paper 19887, National Bureau of Economic Research.

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

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

Lawrence, R. Z., Edwards, L. et al. (2013), US employment deindustrialization: insights from history and the international experience, Policy brief 27, Peterson Institute for International Economics.

Magnusson, L. M. and Mavroeidis, S. (2014), Identification using stability restrictions, *Econometrica* **82**, 1799–1851.

Matsuyama, K. (2009), Structural change in an interdependent world: A global view of manufacturing decline, *Journal of the European Economic Association* **7**, 478–486.

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

Neary, J. P. (1978), Short-run capital specificity and the pure theory of international trade, *Economic Journal* **88**, 488–510.

Nelson, C. R. and Plosser, C. R. (1982), Trends and random walks in macroeconmic time series: some evidence and implications, *Journal of Monetary Economics* **10**, 139–162.

Ngai, L. R. and Pissarides, C. (2004), Balanced growth with structural change, Working paper 627, Centre for Economic Performance, London School of Economics and Political Science.

Nickell, S., Redding, S. and Swaffield, J. (2008), The uneven pace of deindustrialisation in the OECD, *World Economy* **31**, 1154–1184.

Nielsen, B. (2006), Correlograms for non-stationary autoregressions, *Journal of the Royal Statistical Society, Series B* **68**, 707–720.

OECD (2017), 'Value added by activity (indicator), doi: 10.1787/a8b2bd2b-en, (accessed on 20 October 2017)'.

Perron, P. (1989), The Great Crash, the oil price shock, and the unit root hypothesis, *Econometrica* **57**, 1361–1401.

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

Pretis, F., Mann, M. L. and Kaufmann, R. K. (2015), Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection, *Climatic Change* **131**, 705–718.

Pretis, F., Reade, J. and Sucarrat, G. (2016), General-to-specific (gets) modelling and indicator saturation with the R package gets (version 0.7).

Pretis, F., Schneider, L., Smerdon, J. E. and Hendry, D. F. (2016), Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation, *Journal of Economic Surveys* **30**, 403–429.

Ramsey, J. B. (1969), Tests for specification errors in classical linear least-squares regression analysis, *Journal of the Royal Statistical Society, Series B* **31**, 350–371.

Rodrik, D. (2016), Premature deindustrialization, *Journal of Economic Growth* **21**, 1–33.

Rowthorn, R. E. and Ramaswamy, R. (1997), Deindustrialization: causes and implications, Working paper 42, International Monetary Fund.

Rowthorn, R. E. and Ramaswamy, R. (1998), Growth, trade, and deindustrialization, Working paper 60, International Monetary Fund.

Sajedi, R. and Thwaites, G. (2016), Why are real interest rates so low? the role of the relative price of investment goods, *IMF Economic Review* **64**, 635–659.

Santos, C. and Hendry, D. (2010), An automatic test of super exogeneity, *in* M. W. Watson, T. Bollerslev and J. R. Russell, eds, 'Volatility and time series econometrics: essays in honor of Robert F. Engle', Oxford University Press, Oxford.

Santos, C., Hendry, D. F. and Johansen, S. (2008), Automatic selection of indicators in a fully saturated regression, *Computational Statistics* **23**, 317–335.

Stockman, A. C. and Tesar, L. L. (1995), Tastes and technology in a two-country model of the business cycle: Explaining international comovements, *American Economic Review* **85**, 168–185.

Summers, L. H. (2014), US economic prospects: secular stagnation, hysteresis, and the zero lower bound, *Business Economics* **49**, 65–73.

Thomas, R. and Dimsdale, N. (2017), 'A millennium of UK data, Bank of England OBRA dataset'.

Warnock, F. E. and Warnock, V. C. (2009), International capital flows and US interest rates, *Journal of International Money and Finance* **28**, 903–919.

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

Wood, A. (1995), *North-South trade, employment, and inequality: Changing fortunes in a skill-driven world*, Oxford University Press, Oxford.