5

Detectives of Change: Indicator Saturation

Abstract Structural changes are pervasive from innovations affecting many disciplines. These can shift distributions, altering relationships and causing forecast failure. Many empirical models also have outliers: both can distort inference. When the dates of shifts are not known, they need to be detected to be handled, usually by creating an indicator variable that matches the event. The basic example is an impulse indicator equal to unity for the date of an outlier and zero elsewhere. We discuss an approach to finding multiple outliers and shifts called saturation estimation. For finding outliers, an impulse indicator is created for every observation and the computer program searches to see which, if any, match an outlier. Similarly for location shifts: a step indicator equal to unity till time t is created for every t and searched over. We explain how and why this approach works.

Shifting distributions are indicative of structural change, but that can take many forms, from sudden location shifts, changes in trend rates of growth, or in estimated parameters reflecting changes over time in relationships between variables. Further, outliers that could be attributed to specific events, but are not modelled, can lead to seemingly fat-tailed distributions when in fact the underlying process generating the data is thin tailed. Incorrect or changing distributions pose severe problems for modelling any phenomena, and need to be correctly dealt with for viable estimation and inference on parameters of interest. Empirical modelling that does not account for shifts in the distributions of the variables under analysis risks reaching potentially misleading conclusions by wrongly attributing explanations from such contamination to chance correlations with other included variables, as well as having non-constant parameters.

While the dates of some major events like the Great Depression, oil and financial crises, and major wars are known *ex post*, those of many other events are not. Moreover, the durations and magnitudes of the impacts on economies of shifts are almost never known. Consequently, it behoves any investigator of economic (and indeed many other) time series to find and neutralize the impacts of all the in-sample outliers and shifts on the estimates of their parameters of interest. Shifts come at unanticipated times with many different shapes, durations and magnitudes, so general methods to detect them are needed. 'Ocular' approaches to spotting outliers in a model are insufficient: an apparent outlier may be captured by one of the explanatory variables, and the absence of any obvious outliers does not entail that large residuals will not appear after fitting.

It may be thought that the considerable number of tests required to check for outliers and shifts everywhere in a sample might itself be distorting, and hence adversely affect statistical inference. In particular, will one find too many non-existent perturbations by chance? That worry may be exacerbated by the notion of using an indicator saturation approach, where an indicator for a possible outlier or shift at every observation is included in the set of explanatory variables to be searched over. Even if there are just 100 observations, there will be a hundred indicators plus variables, so there are many trillions of combinations of models created by including or omitting each variable and every indicator, be they for outliers or for shifts starting and ending at different times.

Despite the apparent problems, indicator saturation methods can address all of these forms of mis-specification. First developed to detect unknown numbers of outliers of unknown magnitudes at unknown points in the sample, including at the beginning and end of a sample, the method can be generalized to detect all forms of deterministic structural change. We begin by outlining the method of impulse-indicator saturation (IIS) to detect outliers, before demonstrating how the approach can be generalized to include step, trend, multiplicative and designer saturation. We then briefly discuss how to distinguish between non-linearity and structural change.

Saturation methods can detect multiple breaks, and have the additional benefit that they can be undertaken conjointly with all other aspects of model selection. Explanatory variables, dynamics and non-linearities can be selected jointly with indicators for unknown breaks and outliers. Such a 'portmanteau' approach to detecting breaks while also selecting over many candidate variables is essential when the underlying DGP is unknown and has to be discovered from the available evidence. Most other break detection methods rely on assuming the model is somehow correctly specified other than the breaks, and such methods can lack power to detect breaks if the model is far from 'correct', an event that will occur with high probability in non-stationary time series.

5.1   Impulse-Indicator Saturation

IIS creates a complete set of indicator variables. Each indicator takes the value 1 for a single observation, and 0 for all other observations. As many indicators as there are observations are created, each with a different observation corresponding to the value 1. So for a sample of T observations, T indicators are then included in the set of candidate variables. However, all those indicators are most certainly **not** included together in the regression, as otherwise a perfect fit would always result and nothing would be learned. Although saturation creates T additional variables when there are T observations, *Autometrics* provides an expanding and contracting block search algorithm to undertake model selection when there are more variables than observations, as discussed in the model selection primer

in Chapter 2. To aid exposition, we shall outline the 'split-half' approach analyzed in Hendry et al. (2008), which is just the simplest way to explain and analyze IIS, so bear in mind that such an approach can be generalized to a larger number of possibly unequal 'splits', and that the software explores many paths.

Defining Indicators

Impulse indicators are defined as $\{1_{\{j=t\}}\}$ where $1_{\{j=t\}}$ is equal to unity when $j = t$ and equal to zero otherwise for $j = 1, \ldots, T$.

Including an impulse indicator for a particular observation in a static regression delivers the same estimate of the model's parameters as if that observation had been left out. Consequently, the coefficient of that indicator is equal to the residual of the associated observation when predicted from a model based on the other observations. In dynamic relations, omitting an observation can distort autocorrelations, but an impulse indicator will simply deliver a zero residual at that observation. Thus, in both cases, including $T/2$ indicators provides estimates of the model based on the other half of the observations. Moreover, we get an estimate of any discrepancies in that half of the observations relative to the other half. Those indicators can then be tested for significance using the estimated error variance from the other half as the baseline, and any significant indicators are recorded. Importantly, under the null, each half's estimates of parameters and error variance are unbiased.

To understand the 'split-half' approach, consider a linear regression that only includes an intercept, to which we add the first $T/2$ impulse indicators, although there are in fact no outliers. Doing so has the same effect as dummying out the first half of the observations such that unbiased estimates of the mean and variance are obtained from the remaining data. Any observations in the first half that are discrepant relative to those estimates at the chosen significance level, α, say 1%, will result in selected indicators. The locations of any significant indicators are recorded, then the first $T/2$ indicators are replaced by the second $T/2$, and the procedure repeated. The two sets of sub-sample significant indicators (if any) are added to the model for selection of the finally significant indicators. This step is not superfluous: when there is a location shift, for example, some
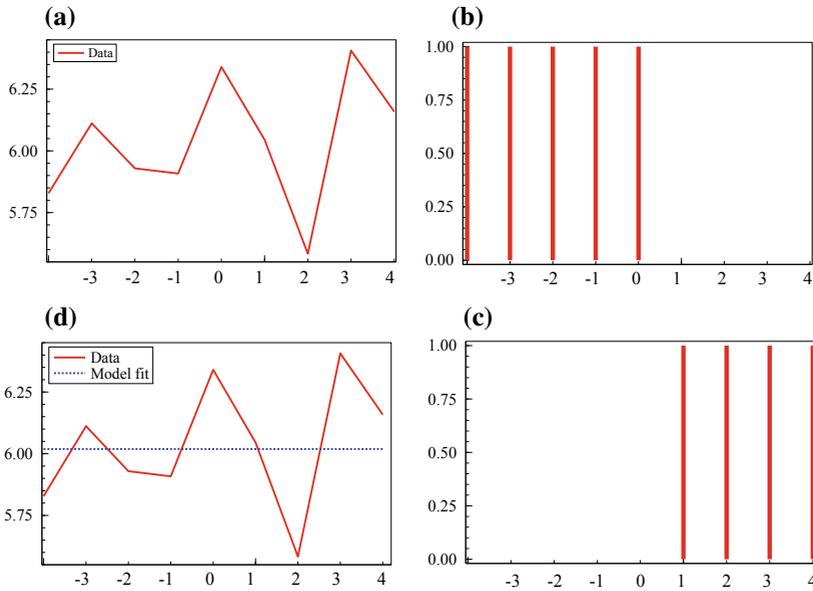
(a)

(b)

(d)

(c)

Fig. 5.1 'Split-half' IIS search under null. **(a)** The data time series; **(b)** the first 5 impulse indicators included; **(c)** the other set of impulse indicators; **(d)** the outcome, as no indicators are selected

indicators may be significant as approximations to the shift, but become insignificant when the correct indicators are included.

Figure 5.1 illustrates the 'split-half' approach when $T = 9$ for an independent, identically distributed (IID) Normal random variable with a mean of 6.0 and a variance of 0.33. Impulse indicators will be selected at the significance level $\alpha = 0.05$.

> **Computer Generated Data**
> The IID Normal variable is denoted by $y_t \sim \text{IN}[\mu, \sigma_y^2]$, where μ is the mean and σ_y^2 is the variance. A random number generator on a computer creates an $\text{IN}[0, 1]$ series which is then scaled appropriately.

Figure 5.1(a) shows the data time series, where the dating relates to periods before and after a shift described below. Then panels (b) and (c) record which of the 9 impulse indicators were included in turn, then panel

(d) shows the outcome, where the fitted model is just a constant as no indicators are selected. Since $\alpha T = 0.05 \times 9 = 0.45$, that is the average null retention rate, where α is called the theoretical gauge, which measures a key property of the procedure. This implies that we expect about one irrelevant indicator to be retained every *second* time IIS is applied to $T = 9$ observations using $\alpha = 0.05$ when the null is true, so finding none is not a surprise.

Hendry et al. (2008) establish a feasible algorithm for IIS, and derive its null distribution for an IID process. Johansen and Nielsen (2009) extend those findings to general dynamic regression models (possibly with trends or unit roots), and show that the distributions of regression parameter estimates remain almost unaltered, despite investigating the potential relevance of T additional indicators, with a small efficiency loss under the null of no breaks when αT is small. For a stationary process, with a correct null of no outliers and a symmetric error distribution, under relatively weak assumptions, the limiting distribution of the estimators of the regression parameters of interest converges to the population parameters at the usual rate (namely $\sqrt{T}$) despite using IIS. Moreover, that is still a Normal distribution, where the variance is somewhat larger than the conventional form, determined by the stringency of the significance level used for retaining impulse indicators. For example, using a 1% significance level, the estimator variance will be around 1% larger.

If the significance level is set to the inverse of the sample size, $1/T$, only one irrelevant indicator will be retained on average by chance, entailing that just one observation will be 'dummied out'. Think of it: IIS allows us to examine T impulse indicators for their significance almost costlessly when they are not needed. Yet IIS has also checked for the possibility of an unknown number of outliers, of unknown magnitudes and unknown signs, not knowing in advance where in the data set they occurred!

The empirical gauge g is the fraction of incorrectly retained variables, so here is the number of indicators retained under the null divided by T. More generally, if on average one irrelevant variable in a hundred is adventitiously retained in the final selection, the empirical gauge is $g = 0.01$. Johansen and Nielsen (2016) derive its distribution, and show g is close to α for small α. IIS has a close affinity to robust statistics, which is not surprising as it seeks to prevent outliers from contaminating estimates of parameters

of interest. Thus, they also demonstrate that IIS is a member of the class of robust estimators, being a special case of a 1-step Huber-skip estimator when the model specification is known.

Illustrating IIS for an Outlier
We generate an outlier of size λ at observation k by
$y_t = \mu + \lambda \mathbf{1}_{\{t=k\}} + \varepsilon_t$ where $\varepsilon_t \sim \text{IN}\left[0, \sigma_\varepsilon^2\right]$ and $\lambda \neq 0$.

To illustrate 'split-half' IIS search under the alternative (i.e., when there is an outlier as in the box), Fig. 5.2 records the behaviour of IIS for an outlier of $\lambda = -1.0$ at observation $k = 1$, so earlier dates are shown as negative. Selecting at $\alpha = 0.05$, no first-half indicators are retained (Fig. 5.2 panel (b)), as the discrepancy between the first-half and second-half means is not large relative to the resulting variance. When those indicators are
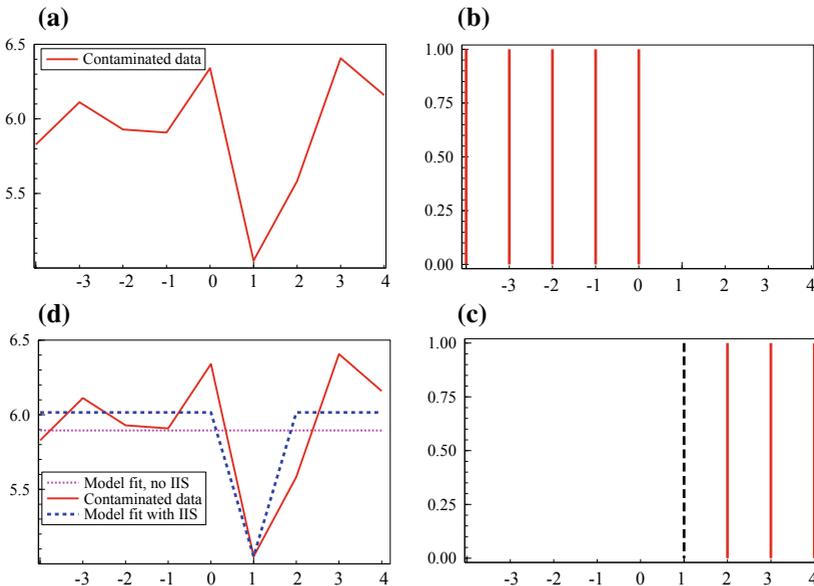

Fig. 5.2 **(a)** Perturbed data time series; **(b)** the first 5 impulse indicators included; **(c)** the other set of impulse indicators where the dashed line indicates retained; **(d)** the outcome with and without the selected indicator

dropped and the second set entered, the first for the period after the outlier is now retained: note that the first-half variance is very small.

Here the combined set is also just the second selection. When the null of no outliers or breaks is true, any indicator that is significant on a sub-sample would remain so overall, but for many alternatives, sub-sample significance can be transient, due to an unmodelled feature that occurs elsewhere in the data set.

Despite its apparently arcane formulation involving more variables plus indicators than available observations, the properties of which we discussed above, IIS is closely related to a number of other well-known statistical approaches. First, consider recursive estimation, where a model is fitted to a small initial subset of the data, say $K > N$ values when there are N variables, then observations are added one at a time to check for changes in parameter estimates. In IIS terms, this is equivalent to starting with impulse indicators for the last $T - K$ observations, then dropping those indicators one at a time as each next observation is included in the recursion.

Second, rolling regressions, where a fixed sample length is used, so earlier observations are dropped as later ones are added, is a further special case, equivalent to sequentially adding impulse indicators to eliminate earlier observations and dropping those for later.

Third, investigators sometimes drop observations or truncate their sample for what they view as discrepant periods such as wars. Again, this is a special case of IIS, namely including impulse indicators for the observations to be eliminated, precisely as we discussed above for modelling US food demand from 1929 to 1952. A key lack in all these methods is not inspecting the indicators for their significance or information content. However, because the variation in such apparently 'discrepant' periods can be invaluable in breaking collinearities and enhancing estimation precision, much can be learned by applying IIS instead, and checking which, if any, observations are actually problematic, perhaps using archival research to find out why.

Fourth, the Chow test for parameter constancy can be implemented by adding impulse indicators for the subsample to be tested, clearly a special case of IIS. Thus, IIS nests all of these settings. There is a large literature on testing for a known number of breaks, but indicator saturation is

applicable when there is an unknown number of outliers or shifts, and can be implemented jointly with selecting over other regressors. Instrumental variables variants follow naturally, with the added possibility of checking the instrument equations for outliers and shifts, leading to being able to test the specification of the equation of interest for invariance to shifts in the instruments.

IIS is **designed** to detect outliers rather than location shifts, but split-half can also be used to illustrate indicator saturation when there is a single location shift which lies entirely within one of the halves. For a single location shift, Hendry and Santos (2010) show that the detection power, or potency, of IIS is determined by the magnitude of the shift; the length of the break interval, which determines how many indicators need to be found; the error variance of the equation; and the significance level, α, as a Normal-distribution critical value, c_α, is used by the IIS selection algorithm. Castle et al. (2012) establish the ability of IIS in *Autometrics* to detect multiple location shifts and outliers, including breaks close to the start and end of the sample, as well as correcting for non-Normality. Nevertheless, we next consider step-indicator saturation, which is explicitly designed for detecting location shifts.

5.2   Step-Indicator Saturation

A step shift is just a block of contiguous impulses of the same signs and magnitudes. Although IIS is applicable to detecting these, then the retained indicators could be combined into one dummy variable taking the average value of the shift over the break period and 0 elsewhere, perhaps after conducting a joint F-test on the *ex post* equality of the retained IIS coefficients, there is a more efficient method for detecting step shifts. We can instead generate a saturating set of $T - 1$ step-shift indicators which take the value 1 from the beginning of the sample up to a given observation, and 0 thereafter, with each step switching from 1 to 0 at a different observation. Step indicators are the cumulation of impulse indicators up to each next observation. The 'T'th step would just be the intercept. The $T - 1$ steps are included in the set of candidate regressors.

The split-half algorithm is conducted in exactly the same way, but there are some differences.

Defining Step Indicators
Step indicators are defined by $\mathbf{1}_{\{t \leq j\}}$, $j = 1, \ldots, T$, where $\mathbf{1}_{\{t \leq j\}} = 1$ for observations up to j, and zero otherwise.

First, while impulse indicators are mutually orthogonal, step indicators overlap increasingly as their second index increases. Second, for a location shift that is not at either end, say from T_1 to T_2, two indicators are required to characterize it: $\mathbf{1}_{\{t \leq T_2\}} - \mathbf{1}_{\{t < T_1\}}$. Third, for a split-half analysis, the ease of detection is affected by whether or not T_1 and T_2 lie in the same split, and whether location shifts occur in both halves with similar signs and magnitudes. Castle et al. (2015) derive the null retention frequency of SIS and demonstrate the improved potency relative to IIS for longer location shifts.

We now consider 'split-sample' SIS for the same data as used for IIS above. As it happens, the second half coincides with the break period, so rather than use the first and second halves, we illustrate 'half-sample' SIS, where some indicators are chosen from each half as shown in Fig. 5.3 under the null. As *Autometrics* software uses multi-path block searches, this choice is potentially one of many paths explored, so has no specific advantage, but hopefully avoids the impression that the method is successful because the shift neatly coincides with the second half.

Figure 5.3 panel (a) records the time series; panels (b) and (c) the first and second choices of the 9 step indicators where now solid, dotted, dashed and long dashed clarify the steps, and panel (d) reports the same outcome as for IIS, as no indicators are selected.

Illustrating SIS for a Location Shift
Here we generate a location shift of magnitude λ at observation k by $y_t = \mu + \lambda \mathbf{1}_{\{t \geq k\}} + \varepsilon_t$ where $\varepsilon_t \sim \mathsf{IN}\left[0, \sigma_\varepsilon^2\right]$ and $\lambda \neq 0$.

Next, we modify the process that generated an outlier to instead generate a location shift of $\lambda = -1$ at $k = 0$, but with the same half selections of step indicators. Figure 5.4 illustrates the outcome. Panel (a) records the shifted data, (b) shows the first selection of step indicators and (c) the remainder
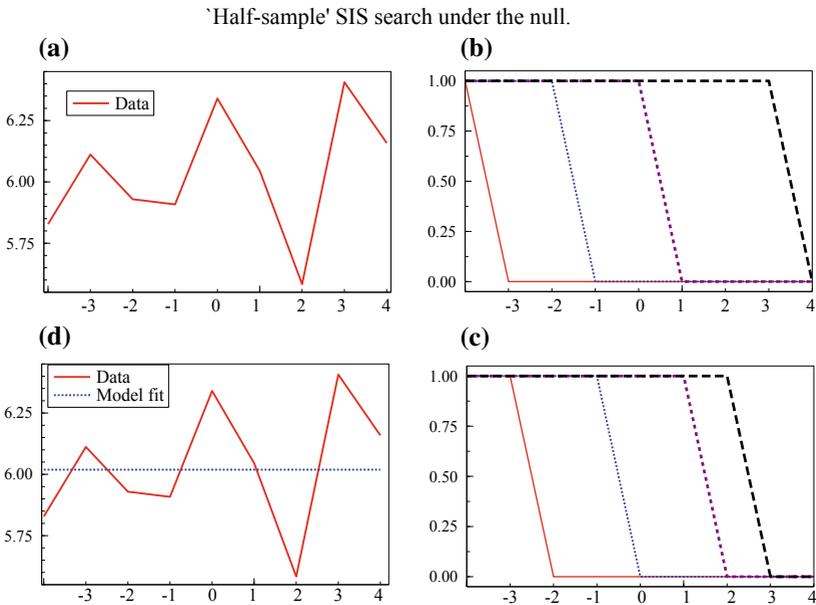
Fig. 5.3 **(a)** The data time series; **(b)** the 4 step indicators included; **(c)** the other set of step indicators; **(d)** the outcome as no indicators are selected

where now the thick solid line denotes the selected indicator, with (d) showing the outcome with and without that selected step indicator.

Notice how the fit without handling the shift produces 'spurious' residual autocorrelation, as all the residuals are first positive, then all become negative after observation 1. 'Treating' the residual autocorrelation by a conventional recipe would not be a good solution (see Mizon 1995) as the location shift is not correctly modelled. Finally, a more parsimonious and less 'overfitted' outcome results than would be found using IIS which would produce a perfect fit to the last 4 data points.

Figure 4.6 for the growth of real wages was used to illustrate co-breaking between wage growth and inflation, both of which experienced myriad shifts. However, the graph hides that the latter half of the twentieth century had a substantively higher mean real-wage growth at 1.8% p.a. post-1945 versus 0.7% p.a. pre, and 1.3% overall. Real wages would have increased 16-fold at 1.8% p.a. from 1860, rather than just threefold at 0.7% p.a.,
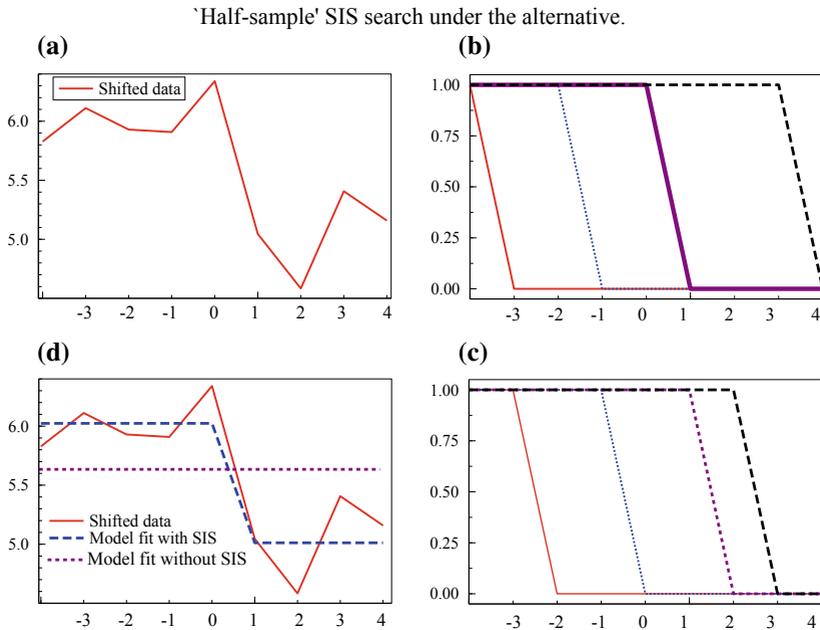
Fig. 5.4 **(a)** The shifted time series; **(b)** the first 4 step indicators included where the thick solid line denotes selected; **(c)** the other 4 step indicators; **(d)** the outcome with the selected step indicator

and sevenfold in practice: 'small' changes in growth rates can dramatically alter living standards. The location shifts shown on the graph were selected by SIS at $\alpha = 0.005$, and were not noticed, or included, in earlier models, but helped clarify the many influences on real wages (see Castle and Hendry 2014).

5.3   Designing Indicator Saturation

But why stop at step-indicator saturation? A location shift in the growth rate of a variable must imply that there is a change in the trend of the variable itself.

5.3.1 Trend-Indicator Saturation

Thus, one way of capturing a trend break would be to saturate the model with a series of trend indicators, which generate a trend up to a given observation and 0 thereafter for every observation. However, trend breaks can be difficult to detect as small changes in trends can take time to accumulate, even if they eventually lead to very substantial differences.

Defining Trend Indicators
Trend indicators are defined as $\mathsf{T}_{jt} = t - j + 1$ for $t \geq j, \; j = 1, \ldots, T$ and 0 otherwise.

Figure 5.5 also illustrated the issue that although the long-run effect of the step shift detected by SIS starting in 1945 was dramatic, that would not have been clear at the time. The average growth of 1.4% p.a. over the first 15 years, 1945–1960, after SIS detects the shift, is little different from the 1.2% p.a. near the start of data period over the 15 years 1864–1879. Indeed, fitting SIS to the sample up to 1960, it finds a location shift from 1944 of 1.1% which could be the end of a World War II effect rather than the start of the prolonged higher growth to come.
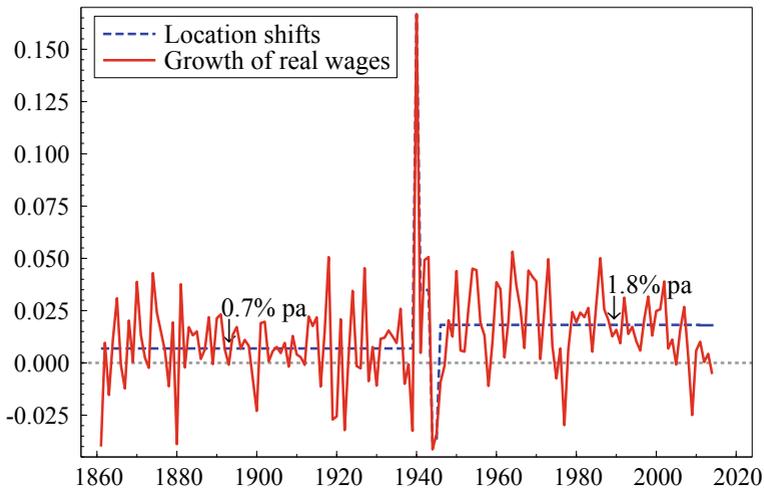

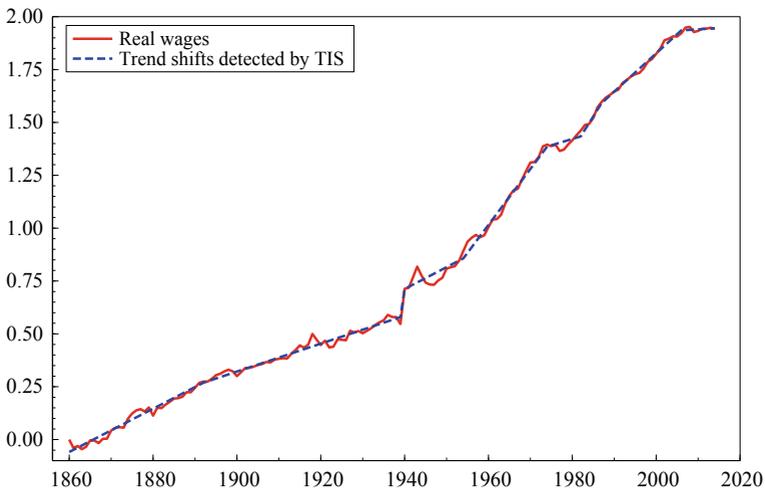Fig. 5.5 A location shift in the growth of UK real wages

Fig. 5.6 Several trend breaks in UK real wages detected by TIS

We illustrate trend-indicator saturation (TIS) for the level of real wages as shown in Fig. 5.6. Selection was undertaken at $\alpha = 0.001$, using such a tight significance level because the variable is $I(1)$ with shifts, so considerable residual serial correlation seemed likely. An overall trend was retained without selection, so deviations therefrom were being detected. Even at such a tight significance level, nine trend indicators were retained, several acting for short periods, as with the jump between 1939 and 1940 (matching the spike in Fig. 5.5), and the flattening over 1973–1981, and again at the end of the period.

5.3.2 Multiplicative-Indicator Saturation

Ericsson (2012) considered a wide range of possible indicator saturation methods, including combining IIS and SIS (super saturation) and multiplicative-indicator saturation (MIS) where every variable in a candidate set is multiplied by every step indicator. For example, with 100 observations and four regressor variables there will be 400 candidates to select from. Kitov and Tabor (2015) have investigated the properties of

MIS by simulation, and found it can detect shifts in regression parameters despite the huge number of candidate variables. This prompted Castle et al. (2017) to apply the approach to successfully detect induced shifts in estimated models following a policy intervention. They offer an explanation for the surprisingly good performance of MIS as follows. Imagine knowing where a shift occurred, so you split your data sample at that point and fit the now correctly specified model separately to the two sub-samples. You would be deservedly surprised if those appropriate sub-sample estimates did not reflect the parameter shifts. Choosing the split by MIS will add variability, but the correct indicator, or one close to it, should be selected as that is where the parameters changed. Of course, as ever with model selection, 'unlucky' draws from the error distribution may make the shift appear to happen slightly earlier or later than actually occurred. We consider an application of MIS in the next Chapter.

5.3.3 Designed-Break Indicator Saturation

If the breaks under investigation have a relatively regular shape, saturation techniques can be 'designed' appropriately, denoted DIS. This idea has been used by Pretis et al. (2016) to detect the impacts of volcanic eruptions on temperature records. When a volcano erupts, it spews material into the atmosphere and above, which can 'block' sunlight, or more accurately, reduce received solar radiation. The larger the eruption, the more solar radiation is reduced. Thus, the eruption of Tambora in 1816 created the 'year without a summer' in the Northern Hemisphere, adding to the difficulties people confronted just after the end of the Napoleonic wars. More generally, atmospheric temperatures drop rapidly during and immediately after an eruption, then as the ejected material is removed from the atmosphere, temperature slowly recovers, like a 'v'. Thus, a saturating set of indicators with such a shape can be created and applied to the relevant time series, selecting rather like we described above for SIS. The follow up in Schneider et al. (2017) demonstrates the success of DIS for detecting the impacts of volcanic eruptions to improve dendrochronological temperature reconstructions.

5.4   Outliers and Non-linearity

The methods discussed above were designed to detect unknown outliers (IIS), location shifts (SIS), trend breaks (TIS), parameter changes (MIS) and volcanic eruptions (DIS) that actually happened, at a pre-set significance level. An alternative explanation for what appears to be structural change is that the data generating process is non-linear. Possible examples include Markov switching models (see e.g., Hamilton 1989), threshold (see e.g., Priestley 1981) and smooth transition models (see e.g., Granger Teräsvirta 1993), where the non-linearity is 'regular' in some way. Distinguishing between the two explanations can be difficult. Indeed, non-linearities and deterministic structural breaks can often be closely similar. But a key advantage of *Autometrics* is that it operates as a variable selection algorithm, allowing selection over non-linear functions as well as potential outliers and breaks, so both explanations can be tested jointly, and both explanations could well play a role in explaining the phenomena of interest.

The *Autometrics*-based approach in Castle and Hendry (2014) creates a class of non-linear functions from transformations of the original data variables to approximate a wide range of potential non-linearities in a low-dimensional way. The problem with including, say, a general cubic function of all the (non-indicator) candidate variables is the explosion in the number of terms that need to be considered. For example, with 20 candidates, there are 1539 cubic terms. However, their simplification adds only 60 terms, at the possible risk of not capturing all the non-linearity in some settings. When an investigator has a specific non-linear function as a preferred explanation, that can be tested against the selected model by encompassing to see if (a) the proposed function is significant, and if so (b) whether it eliminates all the other non-linear terms.

References

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012). Model selection when there are multiple breaks. *Journal of Econometrics*, **169**, 239–246.

Castle, J. L., Doornik, J. A., Hendry, D. F., and Pretis, F. (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, **3(2)**, 240–264. http://www.mdpi.com/2225-1146/3/2/240.

Castle, J. L., and Hendry, D. F. (2014). Semi-automatic non-linear model selection. In Haldrup, N., Meitz, M., and Saikkonen, P. (eds.), *Essays in Nonlinear Time Series Econometrics*, pp. 163–197. Oxford: Oxford University Press.

Castle, J. L., Hendry, D. F., and Martinez, A. B. (2017). Evaluating forecasts, narratives and policy using a test of invariance. *Econometrics*, **5(39)**. https://doi.org/10.3390/econometrics5030039.

Ericsson, N. R. (2012). Detecting crises, jumps, and changes in regime. Working paper, Federal Reserve Board of Governors, Washington, DC.

Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Hamilton, J. D. (1989). A new approach to economic analysis of nonstationary time series. *Econometrica*, **57**, 357–384.

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

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

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

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

Kitov, O. I., and Tabor, M. N. (2015). Detecting structural changes in linear models: A variable selection approach using multiplicative indicator saturation. Unpublished paper, University of Oxford.

Mizon, G. E. (1995). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics*, **69**, 267–288.

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.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. New York: Academic Press.

Schneider, L., Smerdon, J. E., Pretis, F., Hartl-Meier, C., and Esper, J. (2017). A new archive of large volcanic events over the past millennium derived from reconstructed summer temperatures. *Environmental Research Letters*, **12(9)**. https://doi.org/10.1088/1748-9326/aa7a1b/meta.