CEIS Centre for Economic and International Studies
TOR VERGATA

CEIS Tor Vergata

RESEARCH PAPER SERIES

Vol. 12, Issue 9, No. 325 – August 2014

Outlier Detection in Structural Time Series Models: the Indicator Saturation Approach

Martyna Marczak and Tommaso Proietti

Outlier Detection in Structural Time Series Models: the Indicator Saturation Approach

Martyna Marczak*
University of Hohenheim, Germany

Tommaso Proietti
Università di Roma "Tor Vergata", Italy
and CREATES, Denmark

Abstract

Structural change affects the estimation of economic signals, like the underlying growth rate or the seasonally adjusted series. An important issue, which has attracted a great deal of attention also in the seasonal adjustment literature, is its detection by an expert procedure. The general–to–specific approach to the detection of structural change, currently implemented in Autometrics via indicator saturation, has proven to be both practical and effective in the context of stationary dynamic regression models and unit–root autoregressions. By focusing on impulse– and step–indicator saturation, we investigate via Monte Carlo simulations how this approach performs for detecting additive outliers and level shifts in the analysis of nonstationary seasonal time series. The reference model is the basic structural model, featuring a local linear trend, possibly integrated of order two, stochastic seasonality and a stationary component. Further, we apply both kinds of indicator saturation to detect additive outliers and level shifts in the industrial production series in five European countries.

JEL Classification: C22, C51, C53

Keywords: Indicator saturation, seasonal adjustment, structural time series model, outliers, structural change, general–to–specific approach, state space model

*Corresponding author: University of Hohenheim, Department of Economics, Schloss, Museumsfluegel, D-70593 Stuttgart, Germany, e-mail: marczak@uni-hohenheim.de

1 Introduction

Structural change affects the estimation of economic signals, like the underlying growth rate or the seasonally adjusted series. An important issue is its detection by an expert procedure. Automatic outlier detection is already implemented in official seasonal adjustment procedures, like TRAMO–SEATS (Gómez and Maravall, 1996) and X–12 ARIMA (and its enhanced version X–13 ARIMA–SEATS). Both procedures consist of two main stages. First, the observed time series is modeled by means of a seasonal ARIMA (SARIMA) model with possible regression effects, which may include outlier effects. In the subsequent step, based on the identified model, the series is decomposed into different components, e.g. trend or seasonal component, according to the so–called canonical decomposition (TRAMO–SEATS) or by using a cascade filter (X–12 ARIMA). Outlier detection is carried out in the first stage and follows a specific–to–general approach based on sequential addition (potential outliers are identified one after the other), followed by backward deletion.

In this paper, we take a new look at the detection of structural change in seasonal economic time series. In particular, we consider the structural time series approach proposed by Harvey (1989) and West and Harrison (1997), according to which a parametric model for the series is directly formulated in terms of unobserved components. The reference model for the adjustment purpose is the basic structural model (BSM), proposed by Harvey and Todd (1983) for univariate time series, and extended by Harvey (1989) to the multivariate case. The BSM postulates an additive decomposition of the series into a trend, a seasonal and an irregular component. Though this model is relatively simple, it is flexible and provides a satisfactory fit to a wide range of seasonal time series. The model can be represented in state space form, which enables the use of efficient algorithms, such as the Kalman filter and smoother, for likelihood evaluation, prediction and the estimation of the unobserved components. We refer to Durbin and Koopman (2012) for a comprehensive and up–to–date treatment of state space methods.

Seasonal adjustment using structural time series models is well established and can be performed by the specialized software STAMP 8 (Koopman et al., 2009). However, in contrast to the officially used software packages for seasonal adjustment, the latter offers only a basic facility for automatic treatment of outliers. This aspect justifies the necessity

for investigation of different approaches to outlier detection in this particular framework.

We follow here the indicator saturation (IS) approach which is a new, yet very promising strand of research on outlier detection. It has been proposed by Hendry (1999) and constitutes a general–to–specific approach. In his seminal work, Hendry (1999) introduced the impulse–indicator saturation (IIS) as a test for an unknown number of breaks, occurring at unknown times, with unknown duration and magnitude. The procedure relies on adding a pulse dummy as an intervention at every observation in the sample. Significant dummies at individual points in time indicate additive outliers. Properties of this method have been studied by Johansen and Nielsen (2009), Hendry et al. (2008) and Castle et al. (2012). Economic applications of IIS have been provided by, e.g., Hendry and Mizon (2011), Hendry and Pretis (2011) and Ericsson and Reisman (2012).

Recently, also other types of indicator saturation have been discussed in the literature. They are related to different types of intervention functions representing level shifts, slope changes etc. Considering different indicator functions should aid finding the most appropriate types of a structural change; see, for example, Doornik et al. (2013). From the computational point of view, IIS and its extensions pose a problem of having more regressors than observations, which can be solved by dividing all dummies into blocks and selecting over blocks; see, e.g., Hendry and Krolzig (2004). A more elaborate search algorithm, also accounting for collinearity between indicators, is provided by Autometrics (Doornik, 2009c) which is an integral part of PcGive (Doornik and Hendry, 2013). Even though indicator saturation has proven to be both practical and effective in the context of the stationary dynamic regression model, its performance in the structural time series models framework has not been examined yet.

This paper contributes to the literature in that it for the first time combines seasonal adjustment using BSM with the general–to–specific approach to outlier detection. The method presented here substantially differs from the procedures in TRAMO–SEATS and X–12 ARIMA in both the modeling and the outlier detection strategy. In the first step, we assess the performance of indicator saturation via Monte Carlo simulations. After that, we provide an empirical application of the considered method to raw industrial production series in France, Germany, Italy, Spain and UK in the time span 1991.M1 – 2014.M1. In our analysis, we apply impulse–indicator saturation (IIS) and step–indicator saturation

(SIS). The reason for this specific choice is twofold. Pulse and step dummies are the most simple and at the same time the most flexible way of modeling structural changes. Moreover, in the empirical exercise our greatest interest lies in the question whether the procedure is capable of identifying a potential level shift corresponding to the economic and financial crisis starting in Europe around the end of 2008.

The remainder of the article is organized as follows. In Section 2, we describe the framework for modeling seasonal time series with outlying observations and location shifts. In particular, in Section 2.1 we set out the basic structural model with calendar effects, whereas in Section 2.2 we present the concept of indicator saturation and explain how it is integrated in the current framework. Section 3 summarizes findings on the detection power of IIS and SIS, obtained by Monte Carlo simulations using differing settings for the model and outlier detection. In Section 4, IIS and SIS are applied to real data to detect outliers and level shifts. Section 5 concludes.

2 Modeling framework

2.1 The basic structural time series model

The BSM postulates an additive and orthogonal decomposition of a time series into unobserved components representing the trend, seasonality and the irregular component. If y_t denotes a time series observed at $t = 1, 2, \ldots, T$, the decomposition can be written as follows:

$$y_t = \mu_t + \gamma_t + \sum_{k=1}^{K} \delta_{xk} x_{kt} + \epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where μ_t is the trend component, γ_t is the seasonal component, the x_{kt}'s are appropriate regressors that account for any known interventions as well as calendar effects, namely trading days, moving festivals (Easter) and the length of the month, and $\epsilon_t \sim \text{IID } \text{N}(0, \sigma_\epsilon^2)$ is the irregular component.

The trend component has a local linear representation:

$$\begin{array}{rcll} \mu_{t+1} & = & \mu_t + & \beta_t + \eta_t \\ \beta_{t+1} & = & & \beta_t + \zeta_t \end{array} \tag{2}$$

where η_t and ζ_t are mutually and serially uncorrelated normally distributed random shocks with zero mean and variance σ_η^2 and σ_ζ^2, respectively.

The seasonal component can be modeled as a combination of six stochastic cycles whose common variance is σ_ω^2. The single stochastic cycles have a trigonometric representation and are defined at the seasonal frequencies $\lambda_j = 2\pi j/12$, $j = 1, \ldots, 6$. The parameter λ_1 denotes the fundamental frequency (corresponding to a period of 12 monthly observations) and the remaining ones represent the five harmonics (corresponding to periods of 6 months, i.e. two cycles in a year, 4 months, i.e. three cycles in a year, 3 months, i.e. four cycles in a year, 2.4, i.e. five cycles in a year, and 2 months):

$$\gamma_t = \sum_{j=1}^{6} \gamma_{jt}, \qquad \begin{bmatrix} \gamma_{j,t+1} \\ \gamma^*_{j,t+1} \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t} \\ \gamma^*_{j,t} \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega^*_{j,t} \end{bmatrix}, j = 1, \ldots, 5, \quad (3)$$

and $\gamma_{6,t+1} = -\gamma_{6t} + \omega_{6t}$. The disturbances ω_{jt} and ω^*_{jt} are normally and independently distributed with common variance σ_ω^2 for $j = 1, \ldots, 5$, whereas $\mathrm{Var}(\omega_{6t}) = 0.5\sigma_\omega^2$.

Calendar effects are treated by adding regression effects in the model equation for y_t. Trading day (working day) effects occur when the level of activity varies with the day of the week, e.g. it is lower on Saturdays and Sundays. Letting D_{jt} denote the number of days of type j, $j = 1, \ldots, 7$, occurring in month t and assuming that the effect of a particular day is constant, the differential trading day effect for series i is given by:

$$TD_{it} = \sum_{j=1}^{6} \delta_{ij} \left(D_{jt} - D_{7t} \right) \quad (4)$$

The regressors are the differential number of days of type j, $j = 1 \ldots, 6$, compared to the number of Sundays, to which type 7 is conventionally assigned. The Sunday effect on the i-th series is then obtained as $\left(-\sum_{j=1}^{6} \delta_{ij} \right)$. This expedient ensures that the trading day effect is zero over a period corresponding to multiples of the weekly cycle.

As far as moving festivals are concerned, the Easter effect is modeled as $E_t = \delta h_t$ where h_t is the proportion of 7 days before Easter that fall in month t. Subtracting the long run average, computed over the first 400 years of the Gregorian calendar (1583-1982), from h_t yields the regressor $h^*_t = h_t - \bar{h}_t$, where $\bar{h}_t$ takes the values 0.354 and 0.646 in March and April, respectively, and zero otherwise. Finally, the length–of–month regressor

results from subtracting from the number of days in each month, $\sum_j D_{jt}$, its long run average, which is 365.25/12.

2.2 Indicator saturation

Indicator saturation is a general–to–specific approach according to which for every observation an indicator of a specific type is included in the set of candidate regressors. This means that, if T is the number of observations, T indicator variables are added. In this article, we consider two types of indicator saturation: IIS and SIS.

IIS has been the first approach extensively discussed in the indicator saturation literature. If $I_t(\tau)$ denotes an indicator variable, then $I_t(\tau)$ is in the IIS case a pulse dummy taking value 1 for $t = \tau$, and 0 otherwise. Hendry et al. (2008) analyze the distributional properties of IIS when the observations are generated according to the model $y_t = \mu + \varepsilon_t, t = 1, \ldots, T$, where ε_t is normally and independently distributed with mean zero and variance σ_ε^2. For that purpose, they integrate IIS into the model for y_t using the so–called split–half approach. More specifically, in the first step $[T/2]$ indicators for the first half of the sample are added to the model, where $[\cdot]$ denotes integer division, i.e.:

$$y_t = \mu + \sum_{k=1}^{[T/2]} \delta_{Ik} I_t(k) + \varepsilon_t, \quad t = 1, \ldots, T$$

Once the indicators have been selected at the significance level α, the second $T - [T/2]$ indicators replace the first ones, and the selection procedure is repeated. Finally, both sets of significant dummies are combined to determine the terminal model. On average, in the absence of any outlier, αT indicators are expected to be retained by chance in the final stage, so that setting $\alpha = 1/T$ leads to the misclassification of only one observation on average. Hendry et al. (2008) also show that the different number of splits or unequal splits do not affect the retention rate. Johansen and Nielsen (2009) generalize the analysis to stationary and nonstationary autoregressions.

SIS can be seen as an extension of IIS to the case when $I_t(\tau)$ represents a step variable taking value 0 for $t < \tau$, and 1 for $t \geq \tau$. SIS has been evaluated by Doornik et al. (2013) in view of its ability to deal with level shifts. Their study is based on a comprehensive set of Monte Carlo simulations within a simple static framework. While selecting significant

indicators, they apply the standard split–half approach as well as split–half with sequential selection. The latter relies on the iterative elimination of the least significant indicators in each split, until only the significant ones are retained. The finding is that sequential selection considerably improves the power of SIS in detecting location shifts.

In situations when a single set of indicators constitutes the only set of regressors in the model, like in the references previously mentioned, split–half is always a feasible approach. It is, however, possible that the total number of regressors exceeds the number of the available observations, for example if additional regressor variables are included in the model, or different types of indicator saturation are considered at the same time. A simple method to deal with this problem is the cross–block algorithm proposed by Hendry and Krolzig (2004). After partitioning all the indicators into m blocks and performing the initial selection, cross–pairings are formed for which the selection algorithm is run again. This leads in total to $m(m-1)/2$ runs of the selection algorithm. A disadvantage of the cross–block algorithm is that it does not make use of learning and can be thus very slow. A more elaborate method offering a more progressive search is the Autometrics block–search algorithm consisting of expansion and reduction steps (see Doornik, 2009a). Moreover, in cases when different indicator saturation types are used, block–search with an appropriate partitioning of indicators can solve the problem of perfect collinearity. Doornik (2009b) demonstrates that Autometrics block–search is not only faster, but also more successful in finding breaks than the cross–block algorithm.

The indicator saturation approach is integrated in the BSM in the following way. If m denotes the number of blocks into which indicators are split, assuming that the blocks are of equal size and that T is a multiple of m, then in the first stage eq. (1) is extended to:

$$y_t = \mu_t + \gamma_t + \sum_{k=1}^{K} \delta_{xk} x_{kt} + \sum_{k=(T/m)(i-1)+1}^{(T/m)i} \delta_{Ik} I_t(k) + \epsilon_t, \quad t = 1, \ldots, T, \;\; i = 1, \ldots, m \tag{5}$$

where $I_t(k)$ represents an impulse or a step indicator, depending on whether IIS or SIS is considered. Eq. (5) along with models (2) and (3) is put into state space form.

Estimation is carried out by maximum likelihood; the initial states and the regression effects are considered as diffuse and the likelihood is evaluated by the augmented Kalman filter (see de Jong, 1991), which also yields estimates of the intervention effects, $\tilde{\delta}_{xk}, k =$

$1, \ldots, K$, and $\tilde{\delta}_{Ik}, k = 1, \ldots, T$.[1] Once significant indicators are found for every block i, cross–block search is applied to find the terminal model.[2]

3 A Monte Carlo experiment

3.1 Design of the experiment

We investigate the performance of the indicator saturation approach to outlier detection by means of an extensive Monte Carlo experiment.[3] For that purpose, we generate time series from a BSM given in eq. (1) and include an additive outlier (outliers) or a level shift (shifts), thereafter abbreviated by AO and LS, respectively. For simplicity, calendar effects are omitted in the simulations. First, we design a benchmark specification for the data generating process (DGP) and the outlier detection procedure. We subsequently check the robustness of the procedure by considering alternative settings. They are obtained by modifying a single attribute of the DGP and/or the outlier detection procedure, keeping the remaining ones fixed. Every single experiment is based on $M = 1000$ replications.

As regards the simulation settings, we consider a reference GDP with the following specifications:

- The variance parameters are set equal to $\sigma_\epsilon^2 = 1, \sigma_\eta^2 = 0.08, \sigma_\zeta^2 = 0.0001, \sigma_\omega^2 = 0.05$. There is no loss of generality in setting the irregular variance equal to 1; the remaining parameters are thus interpreted as signal to noise ratios. The benchmark GDP is chosen on the basis of our experience in fitting the BSM to industrial production and turnover time series.

- $T = 144$ observations (12 years of monthly data).

- A single additive (AO) or level shift (LS) outlier is located at 0.5 of the sample (observation number 72).

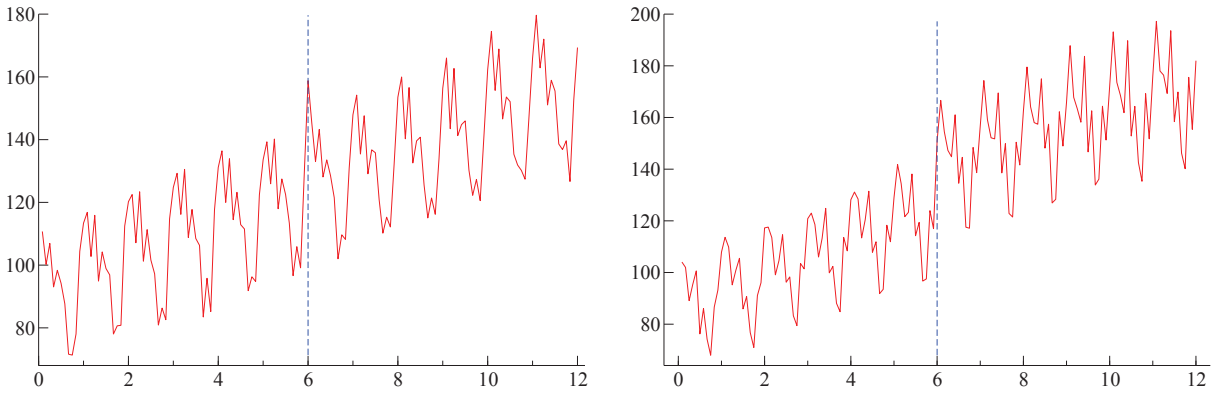
[1] In the SIS case, $I_t(1)$ is left out as it is perfectly collinear with the initial level effect.

[2] Since there does not exist any evidence on indicator saturation within structural time series models at all so far, we want to concentrate on a search algorithm which is easier to implement. Applying Autometrics block–search in context of structural time series models might be, however, an attractive line of future research.

[3] All computations are performed with Ox 6.2 (64-bit version), see Doornik (2008).

- The magnitude of the AO/LS is 7 times the prediction error standard deviation (PESD). The prediction error standard deviation is thereby obtained from the innovations form of the model in the steady state.

Examples of benchmark–based simulated series with an AO and LS are given in Figure 1a and Figure 1b, respectively.



(a) Series with an additive outlier **(b)** Series with a level shift

Figure 1: Examples of series simulated using the benchmark specification

As regards outlier detection with the IIS and SIS, we also specify a benchmark setting:

- the indicator variables are split into 2 blocks,
- the variance parameters are not re-estimated when the split–half indicators are added to the model.

Computation of the vector of regressor effects is thus based on the parameter values estimated with the model excluding indicators. Holding the parameter values fixed can introduce bias in the estimated parameter values, but in our framework it seems to be of crucial importance for the feasibility of indicator saturation from the computational standpoint. Each iteration step often requires several passes of the Kalman filter. This involves several computations of a large vector of regressor effects which introduces instability problems when applying a maximization algorithm. In an alternative setting, we allow for one re-estimation using a single iteration to keep the computational burden at a minimum level and still be able to reduce the bias in the variance estimates. The

same argument has been put forward by Atkinson et al. (1997) who propose, albeit in a different framework, score–based one–step estimates of intervention effects.

We move away from the benchmark scenario in several directions. As regards the variance parameters regulating the DGP, we consider four alternative DGPs: a stable trend – stable seasonal setup (labelled sT – sS), such that the level and seasonal variances are small compared to the irregular variance; an unstable trend–stable seasonal DGP (uT–sS) where the level evolution variance is 0.8; a stable trend–unstable seasonal DGP (sT–uS), such that $\sigma_\omega^2 = 0.5$; and finally we formulate a DGP with unstable trend and seasonality (uT–uS). As for the sample size, we consider shorter time series ($T = 72$, corresponding to 6 years of monthly observations) and longer time series ($T = 288$, i.e. 24 years of monthly data). We also consider different locations for a single outlier and different magnitudes. Concerning the outlier detection settings, we use two alternative numbers of blocks, 3 and 4, respectively, so as to assess the role of further splits in the performance of the outlier detection procedure. Finally, we also examine the role of re-estimation of the parameters within the blocks. A summary of all different settings is provided in Table A.1.

The ability of indicator saturation to detect multiple outliers is also evaluated. Table A.2 provides details on the number of additive outliers and temporary level shifts and their location as a fraction of the total sample size. A temporary level shift occurs when two level shifts have the same magnitude but opposite sign.

For illustrative purposes, examples of series simulated with different settings are presented in Appendix B in Figures B.1 – B.3 for the AO case, and in Figures B.4 – B.7 for the LS case. Independently of the setup, the significance level for retention of respective indicators is always equal to $\alpha = 1/T$.

3.2 Assessing the performance of indicator saturation

In the next subsection, the performance of IIS is evaluated in the presence of additive outliers (in the benchmark as well as alternative setups). Section 3.4 reports the corresponding results of applying SIS to the series with level shifts. The last subsection shows how the performance of IIS compares to SIS in presence of additive outliers or level shifts.

The effectiveness of the procedure is throughout the current section assessed using the

concepts of *potency* and *gauge*. The former is the fraction of relevant indicator variables that are retained in the final model, whereas the latter is the fraction of irrelevant variables in the final model. More formally, let M denote the number of Monte Carlo replications and let n be the number of relevant indicators, i.e. true outliers in any particular time series of length T (e.g. in the benchmark case $n = 1$). Moreover, let $\mathcal{I}_n$ and $\mathcal{I}_{T-n}$ be sets of time indices corresponding to relevant and irrelevant indicators, respectively. Then, potency and gauge are calculated based on the *retention rate*, denoted by $\tilde{p}_k, k = 1, ..., T$, as follows:

$$\begin{aligned}
\tilde{p}_k &= \frac{1}{M}\sum_{i=1}^{M} 1[\tilde{\delta}_{Ik} \neq 0], \quad k = 1, \ldots, T \\
\text{potency} &= \frac{1}{n}\sum_{k} \tilde{p}_k, \quad k \in \mathcal{I}_n \\
\text{gauge} &= \frac{1}{T-n}\sum_{k} \tilde{p}_k, \quad k \in \mathcal{I}_{T-n}
\end{aligned}$$

where $\tilde{\delta}_{Ik}$ denotes the estimated coefficient on the impulse or step indicator, $I_t(k)$, in replication i, if $I_t(k)$ is selected (0 otherwise); $1[\tilde{\delta}_{Ik} \neq 0]$ is variable taking value 1, if the argument in brackets is true, and 0 otherwise.

Potency and gauge as well as their links with concepts commonly used in the multiple testing literature can be illustrated by means of the following confusion matrix summarizing the outcome of a single Monte Carlo experiment:

	Decision		
Actual	No outlier	Outlier	Total
No outlier	A	B	$M(T-n)$
Outlier	C	D	Mn
Total	A+B	B+D	MT

A and D denote numbers of correct decisions in the cases of no outlier and in the cases of an outlier (at a particular observation), respectively. B and C, on the other hand, summarize all false decisions when no outlier is present, and in situations when there is an outlier (at a particular observation), respectively. Potency is then defined as the ratio $D/(Mn)$, which is the true positive rate (also called hit rate, recall or sensitivity) in the classification literature. Gauge is given by the ratio $B/[M(T-n)]$, the so–called

false positive rate (or false alarm rate). The misclassification rate is $(B + C)/(Mn)$, $B/(B + D)$ is the false discovery proportion, and $P(B > 0)$ denoting probability of at least one false retention is the family–wise error rate.

Using the benchmark specification for simulations and outlier detection, we also examine an effectiveness measure which we call *probability of first detection*. More specifically, this probability is defined as the rate at which the true outlier is for the first time spotted. Since it is crucial to detect potential structural breaks as quickly as possible, this property is particularly important to assess the application of indicator saturation for forecasting purposes if the break is close to the forecast origin. For that reason, we consider a situation in which the AO/LS is placed at the end of the benchmark sample (observation number 144). Using simulated series of length 155 observations, the model is estimated recursively along with the outlier detection. Probability of first detection is then computed for the 12 observations starting with the occurrence of the change.

3.3 Additive outliers and impulse–indicator saturation

The simulation results for the benchmark specification featuring a single AO are reported in the first column of Table 1. It can be seen that IIS is capable of identifying the outlier in nearly 100% of cases with a small error rate only. As the other columns show, different variance combinations do not change the potency of the procedure. Gauge remains at a low level, except for the case of a stable trend and unstable seasonal component (fourth column).

Table 1: IIS and AO in the benchmark setup and in alternative setups with different parameter values

	Benchmark	(sT–sS)	(uT–sS)	(sT–uS)	(uT–uS)
Potency in %	99.9	99.3	99.9	99.5	99.5
Gauge in %	0.03	0.04	0.00	0.15	0.06

Comparison of results related to different simulation and outlier detection settings for a single AO are summarized in Tables A.3 – A.6. Similarly as with different parameter values, potency does not change much if different numbers of observations are considered (see Table A.3). Gauge, however, seems to decrease with series length. Potency of the

procedure is considerably affected by the location of the outlier – it decreases towards the ends of the sample (see Table A.4). The lowest gauge values can also be observed against the ends of the series. Moreover, the pattern displays symmetry as the potency and gauge values for outliers located in the same distance from the middle are very similar. As regards the magnitude of the outlier, the effectiveness of IIS increases with outlier size up to some point and then deteriorates (see Table A.5). Using 3 or 4 blocks instead of 2, while keeping parameter values fixed, does not have any impact on potency. In contrast, re-estimation of the variance parameters, when the respective blocks of indicator variables are included, leads to a slightly lower potency of 96.3% (see Table A.6).

As regards the probability of first detection (see Table A.7), it is conspicuous that the performance of IIS strongly depends on the outlier magnitude. For the benchmark outlier size, the probability of first detection of 35% at the time point of the AO is very low. Moreover, small positive probabilities are still observed at the remaining 11 observations. When the size of the AO is doubled, the chance of immediate detection of the AO increases to almost 73%.

Table A.8 summarizes the results for multiple outliers. The findings suggest that it is easier to detect outliers if they are placed in the same sample half, irrespective of whether 2 outliers (first column) or 4 outliers (fourth column) are considered. This finding can be explained by the fact that using the indicators set covering the same half in which all the outliers are present allows for immediate outlier detection.

3.4 Level shifts and step–indicator saturation

An important factor in the detection of LS using SIS is the sequential or non-sequential nature of the outlier detection procedure in each block. As has been mentioned in Section 2.2, sequential selection is supposed to have beneficial effects on the efficiency of SIS.

The results for the benchmark case are presented in the first column of Table 2. Even though potency is smaller than in the benchmark case of detecting AO with IIS, a value of about 90% for both non-sequential and sequential selection is still satisfactory, especially when coupled with the low rates of false retentions. Examination of different combinations of parameter values leads to three observations:

1. Potency is smaller when both components are stable. It increases as the variance of the trend or the seasonal component increases, and it eventually attains the highest value when both components variances are high.

2. Gauge is at its lowest level when trend and seasonal variances are high.

3. Sequential selection improves the detection performance of SIS. This, however, comes at a computational cost. For example, in the benchmark setting, the total simulations time for the non-sequential selection amounts to about 43 minutes whereas for the sequential selection it extends to 3 hours 9 minutes.

Table 2: SIS and LS in the benchmark setup and in alternative setups with different parameter values

		Benchmark	(sT–sS)	(uT–sS)	(sT–uS)	(uT–uS)
Potency	non-seq.	89.3	72.3	74.3	95.2	98.4
in %	seq.	90.7	79.4	82.0	96.8	98.8
Gauge	non-seq.	0.04	0.09	0.02	0.02	0.01
in %	seq.	0.01	0.02	0.03	0.01	0.00

The results corresponding to a single LS and alternative settings are given in the Tables A.9 – A.12 of the Appendix. The length of the series seems to matter more for the effectiveness of the outlier detection procedure than in the case of a single AO (see Table A.9). After doubling the number of observations, potency increases to 99% with a concurrent decrease in gauge to 0%, for both non-sequential and sequential selection.

The location of the shift has similar implications as for a single AO (see Table A.10). However, even though the general pattern of decreasing potency for shift locations more distant from the middle of the sample is maintained, the location symmetry is not existent anymore. A shift location in the second half of the sample allows for higher detectability compared to its mirror location in the first half. Moreover, it becomes apparent that sequential selection plays a crucial role if SIS is applied to identify level shifts, as it raises the chance of spotting the true shift once its location is moved away from the middle.

Sequential selection can also help detect shifts of smaller magnitude whereas there is no gain of applying this procedure when the size is bigger than in the benchmark case (see Table A.11). As can be seen in Table A.12, using more blocks improves the accuracy of

the detection for both considered selection procedures. In contrast, this precision becomes very poor if re-estimation of the model with each block of indicators is performed.

Next, we evaluate probability of first detection (see Table A.13). Similar observations emerge to those made for IIS, as far as SIS is performed with non-sequential selection. In this case, SIS is not reliable enough to detect the shift immediately if the shift is of the benchmark size. Sequential selection, however, has again a beneficial effect for the SIS performance. It is to be noted that in 98.5% of the cases, the shift can be spotted after one observation at the latest. When the size of the shift is doubled, these discrepancies between non-sequential and sequential selection vanish, and they both serve the purpose of timely identification of level shifts.

In addition to a single LS, we also analyze multiple LS. In particular, we focus on temporary LS, by which we mean level shifts that are reversed after some time, so that the initial level is restored. Hence, modeling a temporary LS requires two step indicators having countervailing effects on a series. Table A.14 reports results of the simulation exercise dealing with 1 and 2 temporary shifts. As for 1 shift, it is more demanding to identify it using non-sequential selection when the shift occurs close to the beginning of the series (first column). Interestingly, a temporary LS spanning both halves of the sample can be detected with high probability. Potency corresponding to non-sequential selection generally decreases when 2 temporary shifts are present, especially when they are distributed over both sample halves. The same observation has been made in the context of multiple AO spread across both sample halves. Sequential selection essentially improves performance of SIS, irrespective of the number or position of temporary LS.

3.5 Comparison of impulse– and step–indicator saturation

So far we have investigated the effectiveness of indicator saturation when the intervention (pulse or step dummy) coincides with the indicator type used by the procedure (IIS and SIS, respectively). In practice, however, it is usually not known which type of structural change occurs. It is therefore relevant to assess the performance of SIS when an AO is present, as well as that of IIS in the case of a temporary LS. This entails the necessity to redefine the concepts of potency and gauge. A single AO can in fact be modeled by two adjacent step indicators, whose effects have the same magnitude but opposite signs.

As a result, for a single AO to be identified by SIS, both relevant step indicators have to be retained, which implies that, instead of a single relevant impulse indicator, two relevant step indicators are the reference in the computation of potency and gauge. As for a temporary LS, it can be represented by impulse indicators covering the whole span of the shift and having effects of the same magnitude.[4] Therefore, retaining all indicators in this time span would be required to detect a temporary LS. However, as this condition is very restrictive, we follow Doornik et al. (2013) and measure the effectiveness of IIS in the case of a LS using the so–called *proportional potency*, defined as the average percentage of the level shift captured by the impulse indicators.

The Monte Carlo results for a single AO with two different magnitudes and located at three different fractions of the sample are summarized in Table 3. The results for IIS are also provided for comparison. When the AO is located at 0.25 and 0.4 of the sample, SIS applied with non-sequential selection performs manifestly worse than IIS, but a substantial improvement can be gained by applying sequential selection. Gauge is low for both implementations of SIS. The overall conclusion is that SIS can successfully identify the true outlier. When the AO is located in the middle of the sample, results are less satisfactory, as a consequence of the application of the split–half approach. In the best of the considered scenarios, potency reaches up to only 6.5% at the cost of 0.57% gauge. This implies that, in contrast to IIS, SIS fails at finding the correct AO at the border between two blocks with indicators.

The performance of SIS and IIS in the presence of a temporary LS at different locations and with different magnitudes is presented in Table 4. It is apparent that, compared to SIS, the proportional potency of IIS is very low and gauge is relatively large, except for the first considered shift location. However, some care has to be taken when interpreting these results. As a matter of fact, to get a better insight into the results it is necessary to examine which indicators are retained in the individual simulations.[5] Detailed examination reveals that there are essentially two scenarios that account for the overall poor potency values of IIS. In the first, corresponding to the time span between 0.25 and

[4]Castle et al. (2012) examines the ability of IIS to detect multiple level shifts and outliers. Hendry and Santos (2010) show in context of a single level shift that the detection power of IIS depends on the magnitude of the shift, sample size, the duration of the shift, the error variance and the significance level.

[5]Due to large simulation output, additional results are not presented in the article. They can, however, be made available upon request.

Table 3: Comparison of IIS and SIS in presence of AO at different locations and with different magnitudes

Location[a)]			0.25		0.4		0.5	
Magnitude[b)]			7	14	7	14	7	14
Potency in %	IIS		86.40	98.20	98.70	98.90	99.30	98.70
	SIS	non-seq.	33.60	51.00	40.30	53.75	0.25	0.05
		seq.	68.95	66.20	70.45	64.90	4.40	6.50
Gauge in %	IIS		0.03	0.09	0.01	0.08	0.02	0.10
	SIS	non-seq.	0.01	0.01	0.00	0.00	0.33	0.25
		seq.	0.02	0.02	0.01	0.02	0.50	0.57

[a)] Location is given as a share of the sample length T.
[b)] Magnitude is given as a factor to be multiplied with the prediction error standard deviation (PESD).

0.35 of the sample, only a small fraction of impulse indicators from the relevant range is retained, and the gauge is zero, so that no false positive outlier is found. In the second scenario, which corresponds to the remaining locations, IIS predominantly identifies clusters of few adjacent indicators bordering the time span of the LS on both sides. As the estimated effects of these dummies are negative, the periods before and after the actual LS are treated as periods of negative LS. Although this can be considered as an equivalent way of modeling series with a temporary positive LS, the concepts of potency and gauge are not tailored to deal with this possibility, since they classify the retained indicators as false positives.[6] As a result, the performance of IIS is underestimated. It is worth noting that for a LS occurring in the middle of the sample, i.e. on the boundary of the indicator blocks, few dummies from the first block are retained only so that it is nearly impossible to detect such a shift by IIS. A similar conclusions was drawn for SIS in the context of detection of a single AO at the middle of the series.

4 Applications

In the statistical analysis of economic time series, the detection of structural change has important consequences for the purposes of signal extraction and forecasting. In this sec-

[6] At first sight, it seems difficult to distinguish between a single positive temporary LS and 2 negative temporary LS when only few indicators are kept on both sides of the true shift. All the same, the largest t–values relate to indicators in the direct neighborhood of the borders and thus help recognize a positive LS.

Table 4: SIS and IIS in presence of temporary LS at different locations and with different magnitudes

Location[a]			[0.25, 0.35]		[0.45, 0.55]		[0.5, 0.6]	
Magnitude[b]			7	14	7	14	7	14
Potency in %[c]	IIS		0.67	9.81	0.03	0.00	0.11	0.00
	SIS	non-seq.	74.10	89.45	89.00	97.55	92.85	98.40
		seq.	91.15	87.20	95.80	97.20	97.30	98.20
Gauge in %	IIS		0.00	0.00	1.68	3.49	1.15	2.03
	SIS	non-seq.	0.03	0.04	0.01	0.01	0.02	0.00
		seq.	0.00	0.01	0.00	0.01	0.00	0.00

[a] Location is given as a share of the sample length T.
[b] Magnitude is given as a factor to be multiplied with the prediction error standard deviation (PESD).
[c] For IIS, the numbers refer to proportional potency.

tion, we illustrate the application of indicator saturation to the monthly industrial production time series referring to the manufacturing sector of five European countries: Spain, France, Germany, Italy and the United Kingdom. More specifically, the series concern the monthly seasonally unadjusted volume index of production in manufacturing (according to the NACE Rev.2 classification). The data covers the time span 1991.M1 – 2014.M1 (277 observations) and is provided by Eurostat (download at: `http://epp.eurostat.ec.europa.eu/portal/page/portal/short_term_business_statistics/data/main_tables`).

The objective is to assess how the recent recessionary episode, triggered by the global financial crisis, is characterized by the application of IIS and SIS – whether it can be accommodated by the regular evolution of the stochastic components, or it represents a major structural change.

4.1 Outlier detection with indicator saturation

The reference modeling framework for application of IIS and SIS is the BSM with calendar effects, see Section 2.1. Selection of significant impulse or step indicators is governed by the significance level $1/T = 0.0036$. As far as the implementation of SIS is concerned, we consider both non-sequential and sequential selection. In the sequential procedure, we follow the strategy of splitting the indicators in two blocks. For IIS and non-sequential selection in the SIS case, the number of blocks is an important factor affecting the outcome in terms of detected AO or LS. Therefore, we have decided to take the results generated

with different numbers of blocks into consideration, and combine them suitably to obtain the final results. To that end, we separately identify significant indicators choosing a block number from the range between two and ten. Subsequently, we take the union of all the significant indicators and select the significant ones from this set. The choice of the maximum of ten blocks can be justified by the fact that this is a reasonably high number to reduce the risk of missing any important structural changes.[7]

The results for IIS are reported in Table 5. We can observe a similar pattern for all countries – the procedure retains a couple of dummies with negative effects on the series, starting from 2009.M1 for Spain, Germany and the UK, from 2008.M11 for France and from 2008.M12 for Italy. This finding points to a LS corresponding to the economic and financial crises and enables dating the inception of the recession. For France, Germany and Italy, the AO pattern is very articulate, whereas for Spain and the UK only three impulse indicators show a significant impact. Interestingly, for Spain, Germany and Italy, a positive AO is detected in 2008.M4. Moreover, after a positive AO in 2011.M5, a negative AO is identified in France and Germany in the next month.

Table 5: Outliers detected in five European countries using IIS[a),b)]

ES		FR		GER		IT		UK	
2008.M4	(4.77)	2008.M11	(−3.09)	2008.M4	(3.45)	2008.M4	(3.36)	2002.M6	(−6.20)
2008.M7	(3.08)	2008.M12	(−3.68)	2008.M6	(3.63)	2008.M12	(−3.98)	2005.M3	(−4.50)
2009.M1	(−3.55)	2009.M1	(−5.03)	2008.M9	(3.18)	2009.M1	(−3.85)	2009.M1	(−3.72)
2009.M3	(−3.87)	2009.M2	(−4.92)	2009.M1	(−4.83)	2009.M2	(−5.05)	2009.M2	(−3.17)
2009.M5	(−2.92)	2009.M3	(−5.90)	2009.M2	(−4.75)	2009.M3	(−6.48)	2009.M3	(−3.41)
		2009.M4	(−4.33)	2009.M3	(−4.48)	2009.M4	(−4.39)		
		2009.M5	(−4.19)	2009.M4	(−4.79)	2009.M5	(−6.03)		
		2009.M6	(−3.64)	2009.M5	(−3.40)	2009.M6	(−5.29)		
		2009.M7	(−3.15)	2009.M6	(−3.77)	2009.M7	(−4.77)		
		2011.M5	(6.28)	2009.M7	(−2.92)				
		2011.M6	(−3.00)	2011.M5	(5.03)				
				2011.M6	(−3.25)				

[a)] ES: Spain, FR: France, GER: Germany, IT: Italy, UK: United Kingdom
[b)] t–values of the indicator effects are reported in parentheses.

The results for SIS are presented in Table 6, separately for the non-sequential and sequential implementations. For all countries except Spain, the non-sequential procedure detects a LS in 2008.M11 (France, UK) or 2008.M12 (Germany, Italy), associated with the beginning of the global recession. In the case of Spain, a LS is, however, detected

[7] In fact, increasing the number of blocks over ten did not lead to the detection of any additional AO or LS in the examined series.

by the sequential selection already in 2008.M10. It is worth noting that SIS is capable of detecting most of the AOs identified by IIS, such as those in 2011.M5 in France and Germany, or in 2002.M6 and 2005.M3 in the UK. The comparison of the results obtained with non-sequential and sequential procedure shows particularly striking differences for Spain and Italy. In the case of Spain, non-sequential selection leads to a more generous specification, whereas for Italy a richer specification is chosen by sequential selection. An interesting common pattern emerges from these two cases: every year starting from 2009, a positive LS detected in August is followed by a negative level shift in September. This systematic pattern may mimic a break in the seasonal component, associated with the month August, which possibly occurred in Spain and Italy in 2009.

Table 6: Outliers detected in five European countries using SIS[a),b)]

	ES		**FR**		**GER**		**IT**		**UK**	
non-seq.	2008.M3	(−3.18)	2008.M11	(−6.78)	2008.M12	(−8.48)	2008.M8	(−2.95)	1993.M6	(−3.62)
	2008.M4	(3.30)	2011.M5	(5.22)	2010.M3	(3.74)	2008.M12	(−6.94)	1998.M1	(4.10)
	2008.M5	(−3.48)	2011.M6	(−7.08)	2011.M5	(4.56)	2009.M8	(3.03)	2002.M6	(−5.93)
	2009.M8	(5.20)			2011.M6	(−5.91)			2002.M7	(5.13)
	2009.M9	(−4.46)			2011.M7	(3.13)			2005.M3	(−4.10)
	2010.M8	(4.94)							2005.M4	(4.05)
	2010.M9	(−5.03)							2008.M11	(−6.22)
	2011.M8	(4.32)								
	2011.M9	(−4.52)								
	2012.M8	(5.24)								
	2012.M9	(−5.89)								
	2013.M8	(5.97)								
	2013.M9	(−5.13)								
seq.	2008.M10	(−4.36)	2008.M11	(−6.78)	2008.M11	(−5.78)	2008.M12	(−7.50)	2002.M6	(−5.94)
			2011.M5	(5.22)	2009.M1	(−4.72)	2009.M8	(6.90)	2002.M7	(4.76)
			2011.M6	(−7.08)	2010.M3	(3.77)	2009.M9	(−5.37)	2005.M3	(−3.79)
					2011.M5	(4.88)	2010.M8	(4.99)	2005.M4	(3.94)
					2011.M6	(−5.57)	2010.M9	(−4.84)	2008.M11	(−6.41)
							2011.M8	(4.60)		
							2011.M9	(−6.16)		
							2012.M8	(5.48)		
							2012.M9	(−6.35)		
							2013.M8	(6.55)		
							2013.M9	(−6.54)		

[a)] ES: Spain, FR: France, GER: Germany, IT: Italy, UK: United Kingdom
[b)] t–values of the indicator effects are reported in parentheses.

The estimated trends resulting from the BSM model with the AO and LS identified by IIS and SIS are jointly displayed in Figure 2 for each country[8]. In particular, the plot represents the evolution of the underlying component μ_t, estimated by the Kalman filter

[8]For the sake of clarity, the pictures are restricted to the periods 2005.M1 – 2014.M1 since no outlier was detected before 2005, except in the UK case.

and smoother based on the entire sample. The vertical displacement reflects the location and magnitude of the identified level shifts. In general, SIS interprets the recession as a permanent level shift, whereas according to IIS the recession is a temporary shift taking place around the end of 2008 and affecting part of 2009. For France, the two SIS methods provide exactly the same results and the results are similar for Italy and the UK. For Germany, there is a sizable difference between the trends estimated by the two versions of SIS.

Figure 3 plots the sum of the estimated trend component and the outlier effects resulting from IIS and SIS. It is evident that the combined trends and outliers effects obtained with the IIS approach are more flexible and they adjust more closely to the observed data. IIS possibly leads to overfitting the data. A more parsimonious model could be obtained by SIS, which yields more steady trends, in particular when applied with the sequential procedure. The combined components are very similar across different indicator saturation versions, except for Spain and Italy. As it was mentioned before, SIS with non-sequential selection for Spain and with sequential selection for Italy leads to the identification of a seasonal cluster of additive outliers occurring every August after 2008, which may reflect the consequences of the global recession on the seasonal pattern.

To facilitate the comparison across models for different countries, we employ goodness–of–fit measures for the BSM model without any interventions as well as for different specifications following from indicator saturation. The goodness–of–fit measures include the log-likelihood, the coefficient of determination, R_S^2, suitable for series exhibiting trend and seasonal movements (constructed as the ratio of the innovations variance and the variance of the first differences around a seasonal drift), and the AIC and BIC information criteria. Additionally, we provide the results of the following diagnostic tests: the Ljung–Box autocorrelation test, the Durbin–Watson autocorrelation test, the Goldfeld–Quandt heteroscedasticity test, and the Bowman–Shenton normality test. The results are reported in Table A.15. The goodness–of–fit assessment is strongly in favor of the SIS specifications: SIS with non-sequential selection performs best for Spain and the UK, whereas SIS with sequential selection seems to be superior for Germany and Italy. SIS imparts the best fit also in the case of France. For Spain, Germany and the UK, the specifications associated with the best fit ensure that at least some of the model assumptions (no autocorrelation,

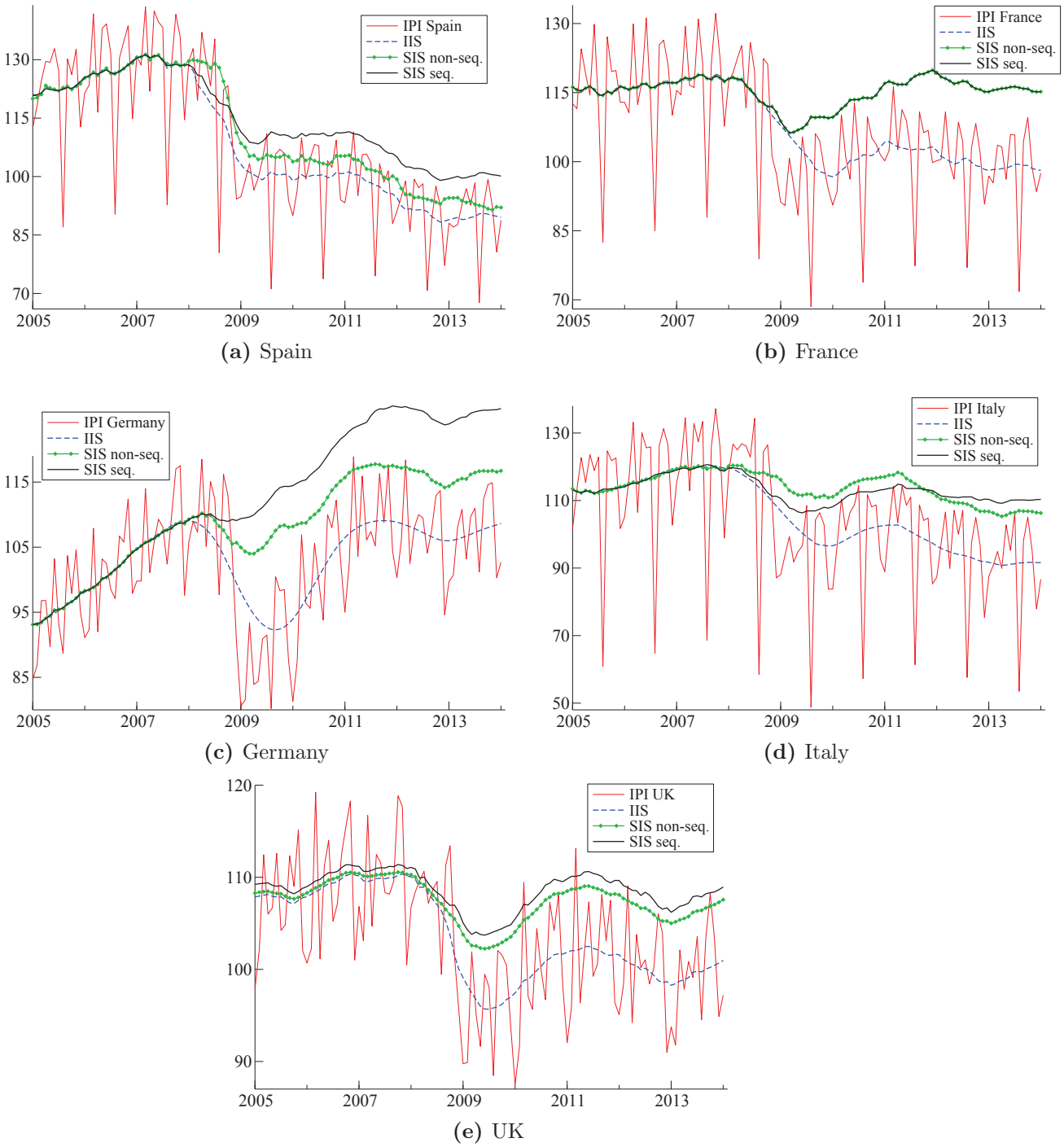


Figure 2: Trend components estimated using the BSM with IIS and SIS for five European countries

homoscedasticity, normality) cannot be rejected, or provide the smallest departures from them compared to other specifications. In contrast, in the case of Italy and France no clear

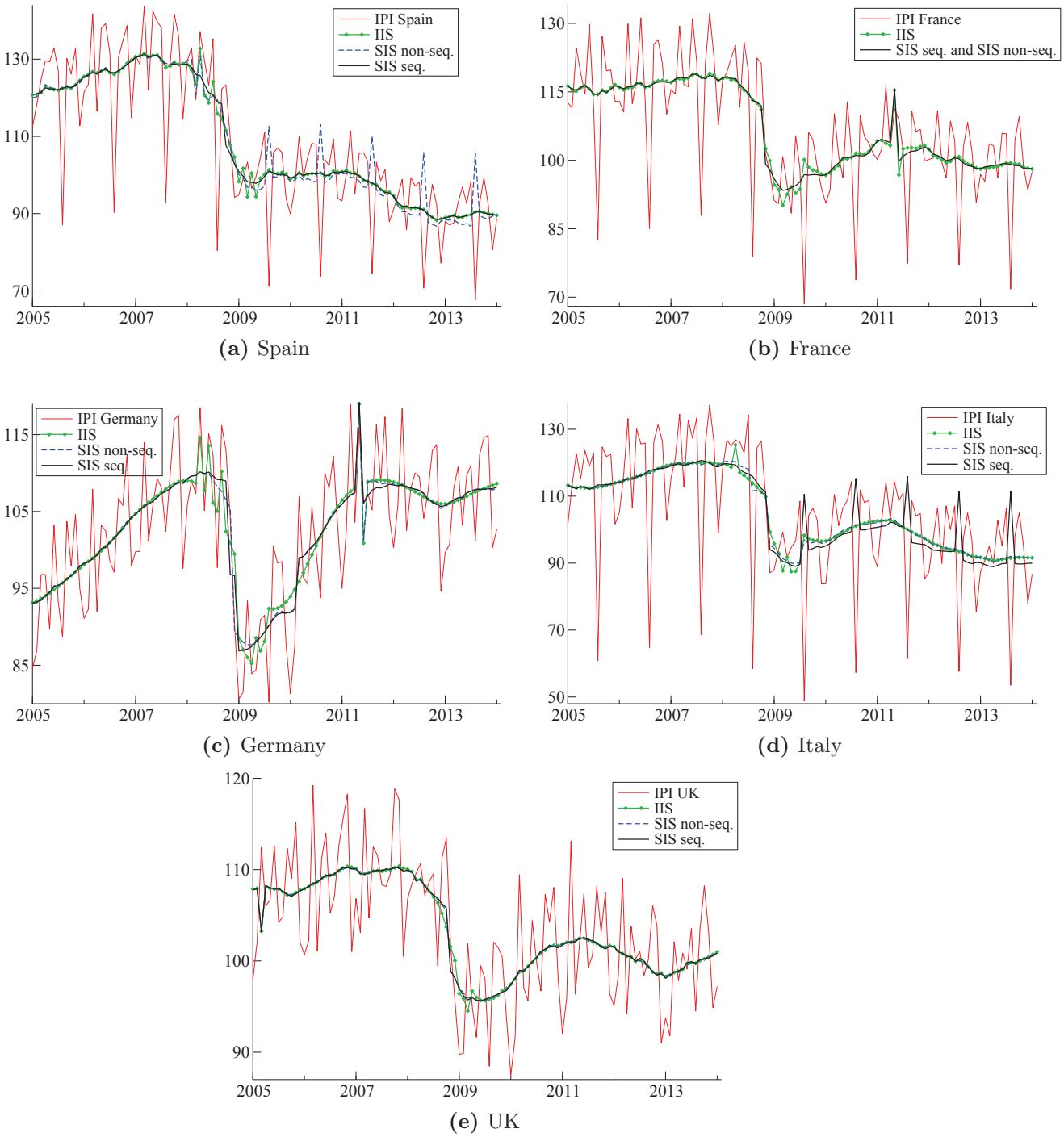


Figure 3: Trend components with outlier effects estimated using IIS and SIS for five European countries

improvement in the diagnostic test statistics relative to inferior models can be ascertained.

We can conjecture that the change in the behavior of European IPI series after the

end of 2008 cannot be fully attributed to the natural evolution of the stochastic trend. The conjecture is based on the fact that the best specifications suggest either a shift in the level of the trend (France, Germany, UK) or/and a change in the seasonal pattern (Italy, Spain). Though the model allows for stochastic evolution in the trend, this cannot fully explain the observed decline in the IPI series during the economic crisis.

4.2 Comparison with alternative outlier detection methods

In the following, we compare the results of the indicator saturation approach with those obtained with alternative methods of automatic outlier detection for seasonal time series. The considered methods are implemented in publicly available statistical software packages, namely TRAMO (see Gómez and Maravall, 1996), TSW: TRAMO–SEATS version for Windows (see Caporello and Maravall, 2004), X13–ARIMA (see U.S. Census Bureau, 2013), STAMP (see Koopman et al., 2009).

In TRAMO, TSW and X13–ARIMA, outlier detection is performed in the framework of seasonal ARIMA models for the underlying series, with models chosen automatically after a few user–predefined settings. Outlier detection is implemented as described by Tsay (1986), Chang et al. (1988), Chen and Liu (1993). In brief, this procedure searches for different types of outliers: additive outliers, level shifts, transitory changes, and innovation outliers (not considered in X13–ARIMA) and consists of two stages. The first one, forward addition, amounts to computing t–statistics for interventions referring to every outlier type at each observation and adding the most significant ones to the model. In the second one, backward deletion, the least significant interventions are eliminated.

In STAMP, the series are modeled in terms of unobserved components. For our analysis, we apply the BSM without any variance restrictions. Outlier detection in STAMP is based on the so–called auxiliary residuals which are the smoothed estimates of the disturbances driving the evolution of the components of the BSM (see Harvey and Koopman, 1992). Significant auxiliary residuals indicate outliers corresponding to particular components, like irregular, trend level, trend slope or seasonal in the case of the BSM.

The outliers identified by the aforementioned procedures are listed in Table A.16. For Spain, TRAMO, TSW and X13–ARIMA identify only one outlier, a LS in 2008.M12, while STAMP detects a number of AOs in addition to a LS in 2008.M12. Generally speaking,

these findings contrast with the indicator saturation outcomes. A LS related to the economic crisis could be detected only with the SIS sequential procedure, albeit already in 2008.M10. Further, none of these algorithms identifies a break in the seasonal pattern, as suggested by SIS with non-sequential selection. For France, all software packages find a LS in 2008.M11 and an AO in 2011.M5, also detected by IIS and SIS. However, TRAMO, TSW and X13–ARIMA additionally identify a LS in 2009.M1 and an AO in 2000.M5 (TRAMO, X13–ARIMA) or a transitory change in 2000.M6, not captured by the indicator saturation. As for Germany, a LS in 2008.M12 detected by TRAMO, TSW and X13–ARIMA, as well as an AO in 2011.M5, detected by TSW, X13–ARIMA and STAMP, are consistent with the outcome of the preferred SIS with sequential selection. In the case of Italy, all the above procedures identify a LS in 2008.M12, which accords with the corresponding findings for indicator saturation. It is worth noting that, except for STAMP, all softwares also find a LS in 2009.M8, which corresponds to one of the LS associated with possible seasonality change uncovered by SIS with sequential selection. As regards the UK, TRAMO, TSW and X13–ARIMA date the LS referring to the economic crisis, just as both SIS versions, at 2008.M11. Both AOs, in 2002.M6 and 2005.M3, detected by IIS and SIS, emerge also from the alternative methods considered. More specifically, the AO in 2002.M6 is also found by TRAMO, X13–ARIMA and STAMP, and the AO in 2005.M3 is also detected by TRAMO and TSW. To sum up, the comparison with different outlier detection procedures reveals that, whereas for some countries, like Germany or the UK, the discrepancies are small and mostly related to AO, in other cases, with Spain as the most distinct example, the mismatch is larger.

4.3 Forecasting

We have seen in Section 4.1 that for the IPI series under investigation the BSM without any interventions may not sufficiently explain potential structural changes, and that a much improved fit can be achieved by applying the indicator saturation approach to the BSM. In this situation, the major structural break identified by the procedures was relatively distant from the end of the sample. It should be recalled that in such a case outlier detection by both IIS and SIS is effective (i.e. has high potency), as was shown by Monte Carlo simulation.

However, for forecasting purposes, an essential property is the timely recognition of abrupt changes in the data occurring towards the end of the sample. Clements and Hendry (2011) show that an unanticipated location shift at the forecast origin can heavily impair forecast precision. The question also arises as to whether specifications resulting from indicator saturation can still prove to be superior to those without any interventions.

To address this question, we perform a recursive forecasting exercise aiming at testing the forecast ability of the BSM without interventions and the BSM with SIS. In other words, we investigate whether the detection of structural change is timely and whether it contributes positively to the accuracy of the predictions. We focus only on SIS as it proved superior in terms of goodness–of–fit, with particular reference to the information criteria computed on the full series. The series under consideration are the five IPI series; the training sample period is the pre-recessionary period ending in 2008.M9, and we use the subsequent observations as a test period.

For every specification (BSM with no interventions, BSM with SIS non-sequential selection, BSM with SIS sequential selection), starting with 2008.M9 as the first forecast origin, we compute 1– to 12–period–ahead recursive forecasts. Then the sample is extended by one month and again 1– to 12–period–ahead forecasts are calculated. These steps are repeated until 2009.M8, which is the last forecast origin.

This pseudo real–time forecasting exercise yields 12 forecasts at horizons from 1 to 12. We choose the 12–month interval between 2008.M9 and 2009.M8 for computing forecasts since the resulting predictions cover the periods shortly before, during and after the occurrence of the LS at the end of 2008. The forecasting performance is evaluated by the the root mean square errors (RMSE) for every specification and every forecast horizon between 1 and 12.

Several observations emerge from the comparison of RMSE values reported in Table 7. For Germany and Italy, SIS, both with non-sequential and sequential selection, by and large outperforms the specification without interventions (with the only exception represented by the 1–step–ahead forecast for Italy related to SIS with non-sequential selection). Interestingly, the gap between the RMSE values corresponding to the approach without SIS and with SIS increases with the forecast horizon. For France and the UK, SIS with non-sequential selection does not impart any improvement in the predictive accuracy. The

Table 7: Root mean square error (RMSE) of recursive forecasts of the industrial production index for five European countries[a),b)]

Forecast horizon			1	2	3	4	5	6	7	8	9	10	11	12
ES	no SIS		7.88	10.02	11.86	13.82	16.09	17.94	19.35	21.37	22.37	23.53	24.65	26.99
	SIS	non-seq.	8.47	10.42	12.08	14.03	16.29	18.29	19.69	21.64	22.67	23.78	24.71	27.61
		seq.	8.32	10.36	11.93	13.54	15.74	17.53	18.86	20.75	21.63	22.82	23.77	26.24
FR	no SIS		6.54	8.90	10.86	13.50	15.41	16.72	17.86	19.19	20.56	21.98	23.73	26.00
	SIS	non-seq.	6.81	9.02	10.89	13.50	15.40	16.70	17.86	19.18	20.63	22.00	23.79	26.24
		seq.	6.29	8.12	9.70	11.41	12.43	12.80	12.51	12.88	13.56	13.77	14.52	16.45
GER	no SIS		7.03	10.22	13.56	15.93	18.17	20.05	22.66	24.19	26.89	28.58	31.32	33.91
	SIS	non-seq.	6.55	9.77	12.32	14.65	16.17	17.75	19.47	20.31	21.70	23.47	25.54	27.78
		seq.	5.06	8.03	10.43	11.81	11.89	12.19	12.75	12.33	12.54	12.09	12.02	12.22
IT	no SIS		9.43	11.98	14.95	17.83	20.84	22.56	24.05	25.91	27.01	29.21	31.38	34.28
	SIS	non-seq.	9.51	11.17	12.91	15.30	16.60	16.38	16.06	16.56	14.94	15.39	15.72	17.37
		seq.	9.33	11.06	12.65	14.61	16.32	15.92	15.30	15.94	13.94	14.02	14.07	15.70
UK	no SIS		2.94	4.32	5.20	6.11	6.82	7.48	8.49	9.33	10.31	11.31	12.42	13.62
	SIS	non-seq.	2.69	4.08	5.14	6.15	6.97	7.68	8.80	9.68	10.75	11.79	13.01	14.22
		seq.	2.37	3.56	4.32	4.97	5.15	5.11	5.36	5.51	5.74	5.84	6.26	6.76

[a)] ES: Spain, FR: France, GER: Germany, IT: Italy, UK: United Kingdom

[b)] The reported RMSE values are computed for every forecast horizon with reference to the 12 1–step and multi–step forecast errors for the forecast lead times from 2008.M9 to 2009.M8.

accuracy improves considerably, however, when sequential selection is used. Similarly to the case of Germany and Italy, the RMSE progressively declines as the forecast horizon increases. As far as Spain is concerned, SIS with non-sequential selection performs worse than the approach without SIS. Sequential selection in general helps to improve the predictive accuracy, even though the RMSE values referring to the 1– to 3–step–ahead forecast are higher than in the case without any intervention. The remaining RMSE values are lower, but, unlike in the case of the other series, they do not differ much from those obtained with the BSM without interventions.

Summing up, SIS, particularly applied with sequential selection, proves to be suitable for forecasting purposes even when a structural break is close to the end of the sample. This conclusion is consistent with the simulation results discussed in Section 3.4, according to which, for large shifts, both non-sequential and sequential selection guarantee high probability of first detection right after the shift. For a smaller LS size, sequential selection is, though, of vital importance for timely outlier detection. An explanation for the different findings across countries is provided in Table A.17. The disappointing results obtained for Spain can be explained with the difficulty of identifying the LS associated with the economic crisis. If, on the other hand, the relevant LS is detected timely, like for Germany and Italy, SIS leads to models yielding substantially better results than the basic model.

5 Conclusions

This article has investigated the performance of the indicator saturation approach as a methodology for detecting additive outliers and location shifts when dealing with non-stationary seasonal series in a model based framework. While the currently available automatic outlier detection procedures follow a specific–to–general approach to uncover structural change, indicator saturation, as a general–to–specific approach, constitutes a relatively new concept in the literature.

Indicator saturation has proven very effective in a regression framework and is currently implemented in Autometrics. Its use for the class of structural time series models has not yet been investigated and this article aimed at filling the gap. The considered model–based framework is interesting as the time series model is directly formulated in terms of unobserved components that are evolving over time. Hence, stochastic change occurs with every new observation, as the components are driven by random disturbances. The issue is then to locate and quantify large economic shocks that configure a structural break differing from the regular endogenous variation of the dynamic system.

We have implemented both impulse–indicator and step–indicator saturation (IIS and SIS) in the framework of the basic structural time series model (BSM). IIS is customized to detect additive outliers (AO), whereas SIS, both with non-sequential as well as with sequential selection, is tailored to detect level shifts (LS). First, we have evaluated the effectiveness of IIS and SIS, by measuring their potency and gauge, in a comprehensive Monte Carlo simulation exercise. It has been shown that, for a reference data generating process and a baseline specification of the procedures, IIS and SIS are very effective methods for outlier detection, especially when SIS is combined with sequential selection.

We then explored several factors that can affect the performance of indicator saturation, and we concluded the following:

- The relative variability of the disturbances driving the evolution of the level and the seasonality does do not matter for the performance of the IIS procedure in detecting AOs. In the SIS case, on the other hand, the detection of LS is easier the higher the evolution error variance of the trend and seasonal.

- The time location of an AO strongly affects the performance of IIS, with potency

and gauge deteriorating when the AO occurs towards the beginning or the end of the sample. In the SIS case, similar considerations hold, but potency and gauge do not vary symmetrically with respect to the location of the LS (LS are easier to detect in the second half of the sample).

- The number of blocks considered in the implementation of the procedure are important drivers of its performance. For instance, if several AOs/LS are present, it is beneficial for both IIS and SIS if they are located in the same sample split.

- SIS with sequential selection provides systematically better results than SIS with non-sequential selection in all the alternative settings considered in the simulations.

- When SIS is used for AO detection, the success rate is satisfactory provided that the AO is not placed at the border between sample splits. IIS, in contrast, does not show acceptable properties when applied to identify LS.

In the last part of the article, we have applied indicator saturation to the monthly industrial production time series for five European countries, with the intent of investigating how the different methodologies characterized the global recessionary movements affecting the euro area economies towards the end of 2008. In general, SIS provided the best specification in terms of goodness–of–fit, capturing a LS in November or December 2008, depending on the series. The comparison with the currently available automatic outlier detection procedures showed a good degree of similarity of the results for Germany and the UK and some important differences for Spain.

Finally, we conducted a pseudo real–time recursive forecasting exercise comparing the out–of–sample performance of the BSM with and without indicator saturation, so as to investigate whether the timely detection of structural change leads to an improvement in the quality of the predictions. As a test sample, we considered the inception and the continuation of the global recession. SIS proved effective in detecting potential location shift close to the forecast origin. The overall conclusion is that the detection of structural change is necessary to obtain accurate forecasts. The sooner the relevant level shift is detected, like for Germany and Italy, the bigger is the improvement in the forecast precision. Sequential selection substantially helps to accomplish this goal.

Appendix

A Tables

Table A.1: Simulation and outlier detection specifications for series with a single additive outlier (AO) / single level shift (LS)

Attributes	Benchmark	Alternative settings
	Data generating process	
Parameter values	$\sigma_\epsilon^2 = 1$	1) (sT–sS) $\sigma_\eta^2 = 8 \cdot 10^{-5}$, $\sigma_\omega^2 = 5 \cdot 10^{-5}$
	$\sigma_\zeta^2 = 0.0001$	2) (uT–sS) $\sigma_\eta^2 = 0.8$, $\sigma_\omega^2 = 5 \cdot 10^{-5}$
	$\sigma_\eta^2 = 0.08$	3) (sT–uS) $\sigma_\eta^2 = 8 \cdot 10^{-5}$, $\sigma_\omega^2 = 0.5$
	$\sigma_\omega^2 = 0.05$	4) (uT–uS) $\sigma_\eta^2 = 0.8$, $\sigma_\omega^2 = 0.5$
Number of observations	144	1) 72, 2) 288
Outlier location[a)]	0.5	1) 0.05, 2) 0.1 3) 0.15 4) 0.25, 5) 0.4, 6) 60, 7) 0.75, 8) 0.85, 9) 0.9, 10) 0.95
Outlier magnitude	$7 \cdot PESD$,	$[2, 14] \cdot PESD$
	Outlier detection settings	
Blocks number	2	1) 3, 2) 4
Re-estimation in blocks	no	yes

[a)] Location is given as a share of the sample length T.

Table A.2: Outlier location for series with multiple additive outliers (AO) / multiple level shifts (LS)

Number of outliers	Location[a), b)]
	Additive outliers
2 outliers	1) 0.25, 0.35; 2) 0.3, 0.6
4 outliers	2) 0.2, 0.4, 0.6, 0.8; 2) 0.6, 0.7, 0.75, 0.9
	Temporary level shift[c)]
1 shift	1) [0.25, 0.35], 2) [0.45, 0.55], 3) [0.5, 0.6]
2 shifts	1) [0.2, 0.3], [0.35, 0.45]; 2) [0.25, 0.35], [0.65, 0.75]

[a)] Location is given as a share of the sample length T.

[b)] All other attributes used in simulations of series and outlier detection are as in the benchmark setup described in Table A.1.

[c)] Temporary level shift requires two level shifts of the same magnitude but opposite signs.

Table A.3: IIS and AO in the benchmark setup and in alternative setups with different numbers of observations

	Benchmark: 144	72	288
Potency in %	99.9	99.0	99.9
Gauge in %	0.03	0.06	0.01

Table A.4: IIS and AO in the benchmark setup and in alternative setups with different locations of the outlier[a)]

	Benchmark: 0.5	0.05	0.10	0.15	0.25	0.40	0.60	0.75	0.85	0.90	0.95
Potency in %	99.9	42.8	68.7	68.8	83.4	98.2	98.8	90.5	71.8	67.4	42.2
Gauge in %	0.03	0.01	0.02	0.03	0.04	0.01	0.01	0.02	0.04	0.02	0.01

[a)] Location is given as a share of the sample length T.

Table A.5: IIS and AO in the benchmark setup and in alternative setups with different magnitudes of the outlier[a)]

	Benchmark: 7	2	4	6	8	10	12	14
Potency in %	99.9	14.2	85.0	99.4	99.9	99.6	98.5	98.4
Gauge in %	0.03	0.01	0.03	0.02	0.01	0.02	0.02	0.08

[a)] Magnitude is given as a factor to be multiplied with the prediction error standard deviation (PESD).

Table A.6: IIS and AO in the benchmark setup and in alternative detection settings

	Benchmark: 2 blocks, no re-estimation	3 blocks, no re-estimation	4 blocks, no re-estimation	2 blocks, re-estimation[a)]
Potency in %	99.9	100	99.7	96.3
Gauge in %	0.03	0.01	0.02	0.01

[a)] Results are obtained after 1 iteration.

Table A.7: Probability of first detection of AO using IIS in the benchmark setup and an alternative setup[a)]

Obs. no.	144	145	146	147	148	149	150	151	152	153	154	155
Benchmark: $7 \cdot PESD$	35.2	4.6	4.2	2.7	2.6	1.6	1.1	1.3	0.7	1.0	0.5	0.6
$14 \cdot PESD$	72.8	11.8	7.6	1.4	1.7	1.3	1.0	0.7	0.2	0.1	0.1	0.0

[a)] Probability is expressed in %.

Table A.8: IIS in presence of multiple AO at different locations[a)]

	2 outliers		4 outliers	
	0.25, 0.35	0.3, 0.6	0.2, 0.4, 0.6, 0.8	0.6, 0.7, 0.75, 0.9
Potency in %	91.70	84.45	70.07	94.07
Gauge in %	0.05	0.00	0.03	0.27

[a)] Location is given as a share of the sample length T.

Table A.9: SIS and LS in the benchmark setup and in alternative setups with different numbers of observations

		Benchmark: 144	72	288
Potency in %	non-seq.	89.3	67.4	98.9
	seq.	90.7	77.6	98.9
Gauge in %	non-seq.	0.04	0.06	0.00
	seq.	0.01	0.10	0.00

Table A.10: SIS and LS in the benchmark setup and in alternative setups with different locations of the shift[a)]

		Benchmark: 0.5	0.05	0.1	0.15	0.25	0.40	0.60	0.75	0.85	0.90	0.95
Potency in %	non-seq.	89.3	34.0	53.7	53.9	64.5	77.1	92.4	83.3	62.0	63.8	44.4
	seq	90.7	61.8	92.6	93.1	90.4	90.8	97.8	96.6	97.8	96.9	96.5
Gauge in %	non-seq.	0.04	0.01	0.04	0.05	0.02	0.01	0.01	0.02	0.04	0.05	0.01
	seq.	0.01	0.36	0.00	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02

[a)] Location is given as a share of the sample length T.

Table A.11: SIS and LS in the benchmark setup and in alternative setups with different magnitudes of the shift[a)]

		Benchmark: 7	2	4	6	8	10	12	14
Potency in %	non-seq.	89.3	13.5	60.9	85.4	92.1	92.4	94.3	94.4
	seq.	90.7	70.5	87.4	90.8	92.8	92.1	93.8	94.4
Gauge in %	non-seq.	0.04	0.06	0.10	0.05	0.03	0.02	0.01	0.01
	seq.	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.00

[a)] Magnitude is given as a factor to be multiplied with the prediction error standard deviation (PESD).

Table A.12: SIS and LS in the benchmark setup and in alternative detection settings

		Benchmark: 2 blocks, no re-estimation	3 blocks, no re-estimation	4 blocks, no re-estimation	2 blocks, re-estimation[a)]
Potency in %	non-seq.	89.3	99.2	100	19.9
	seq.	90.7	99.9	100	–
Gauge in %	non-seq.	0.04	0.00	0.01	0.77
	seq.	0.01	0.01	0.02	–

[a)] Results are obtained after 1 iteration; sequential selection is not considered in the re-estimation due to high computational expense and high risk of estimation failures.

Table A.13: Probability of first detection of LS using SIS in the benchmark setup and an alternative setup[a)]

Obs. no.		144	145	146	147	148	149	150	151	152	153	154	155
Benchmark: $7 \cdot PESD$	non-seq.	42.5	9.0	2.9	0.8	0.3	0.2	0.1	0.1	0.0	0.1	0.1	0.2
	seq.	85.9	12.6	1.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$14 \cdot PESD$	non-seq.	96.6	2.7	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	seq.	90.1	9.7	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[a)] Probability is expressed in %.

Table A.14: SIS in presence of temporary LS at different locations[a)]

		1 shift			2 shifts	
		[0.25, 0.35]	[0.45, 0.55]	[0.5, 0.6]	[0.2, 0.3], [0.35, 0.45]	[0.25, 0.35], [0.65, 0.75]
Potency in %	non-seq.	71.95	87.50	81.80	72.05	53.93
	seq.	91.15	94.95	97.05	85.82	94.55
Gauge in %	non-seq.	0.01	0.01	0.02	0.03	0.00
	seq.	0.00	0.00	0.00	0.00	0.00

[a)] Location is given as a share of the sample length T.

Table A.15: Goodness–of–fit and diagnosis tests results for models without IS, with IIS and SIS for five European countries[a)]

			Goodness–of–fit[b)]				Diagnostics[c)]			
			Log-likelihood	R_S^2	AIC	BIC	$Q(24)$	DW	$H(93)$	BS
ES	no IS		−573.568	0.850	2.520	2.743	49.313*	1.996	2.727*	69.524*
	IIS		−550.238	0.869	2.504	2.897	46.382*	1.900	2.289*	27.091*
	SIS	non-seq.	−515.408	0.892	2.395	2.892	34.590	1.953	1.954*	7.765*
		seq.	−564.973	0.858	2.536	2.876	54.074*	1.995	2.497*	89.223*
FR	no IS		−533.839	0.815	2.195	2.417	44.007*	1.968	4.169*	94.560*
	IIS		−471.754	0.873	2.001	2.472	51.892*	1.859	1.697*	1.566
	SIS	non-seq.	−490.836	0.863	1.985	2.352	53.689*	1.879	2.324*	19.513*
		seq.	−490.836	0.863	1.985	2.352	53.689*	1.879	2.324*	19.513*
GER	no IS		−549.799	0.841	2.259	2.481	45.528*	2.191	1.148	41.605*
	IIS		−485.167	0.885	2.124	2.608	51.892*	1.859	1.697*	1.566
	SIS	non-seq.	−494.171	0.886	2.035	2.427	44.016*	2.269	1.585*	11.260*
		seq.	−490.111	0.889	2.010	2.403	44.761*	2.282	1.464	4.700
IT	no IS		−575.599	0.822	2.532	2.755	66.929*	1.998	2.304*	98.834*
	IIS		−537.196	0.854	2.491	2.936	60.213*	1.937	1.321	3.612
	SIS	non-seq.	−545.703	0.854	2.425	2.791	65.207*	1.983	1.566*	0.221
		seq.	−518.960	0.880	2.324	2.795	65.207*	1.956	1.185	10.841*
UK	no IS		−393.953	0.891	1.038	1.260	43.755*	2.002	1.145	40.822*
	IIS		−349.641	0.915	0.903	1.296	24.539	1.995	1.008	28.059*
	SIS	non-seq.	−325.007	0.930	0.740	1.159	29.742	2.061	1.071	2.727
		seq.	−336.676	0.925	0.796	1.189	23.852	2.046	0.909	2.801

[a)] ES: Spain, FR: France, GER: Germany, IT: Italy, UK: United Kingdom

[b)] R_S^2: coefficient of determination suitable for data displaying trend and seasonal movements; AIC and BIC: information criteria based on the prediction error variance

[b)] $Q(p)$: Ljung–Box statistic based on the first p standardised innovations; DW: Durbin–Watson statistic; $H(h)$: heteroscedasticity statistic based on the first h and the last h standardised innovations, with h being the closest integer to $T/3$; BS: Bowman–Shenton normality statistic; * indicates statistical significance at the 5% level.

Table A.16: Outliers detected for five European countries with different software packages[a),b),c)]

	TRAMO		TSW		X13–ARIMA		STAMP[d)]	
ES	2008.M12	(LS)	2008.M12	(LS)	2008.M12	(LS)	1997.M4	(AO)
							2002.M3	(AO)
							2002.M4	(AO)
							2005.M4	(AO)
							2008.M3	(AO)
							2008.M4	(AO)
							2008.M12	(LS)
FR	2000.M5	(AO)	2000.M6	(TC)	2000.M5	(AO)	2008.M12	(LS)
	2008.M11	(LS)	2008.M11	(LS)	2008.M11	(LS)	2011.M5	(AO)
	2009.M1	(LS)	2009.M1	(LS)	2009.M1	(LS)		
	2011.M5	(AO)	2011.M5	(AO)	2011.M5	(AO)		
			2001.M6	(AO)				
GER	2008.M12	(LS)	2000.M5	(AO)	2000.M5	(AO)	2000.M5	(AO)
			2008.M12	(LS)	2008.M12	(LS)	2009.M6	(SC)
			2009.M12	(TC)	2011.M5	(AO)	2011.M5	(AO)
			2011.M5	(AO)				
			2011.M6	(AO)				
IT	2008.M12	(LS)	1991.M4	(AO)	2008.M12	(LS)	2008.M4	(AO)
	2009.M8	(LS)	1998.M12	(TC)	2008.M12	(AO)	2008.M12	(LS)
			2002.M4	(AO)	2009.M1	(LS)		
			2008.M12	(LS)	2009.M8	(LS)		
			2009.M3	(LS)				
			2009.M8	(LS)				
UK	1998.M1	(LS)	1993.M6	(LS)	2002.M6	(AO)	2002.M6	(AO)
	2002.M6	(AO)	1998.M1	(LS)	2008.M11	(LS)	2008.M12	(LS)
	2005.M3	(AO)	2005.M3	(AO)				
	2008.M11	(LS)	2008.M11	(LS)				
	2009.M1	(LS)	2009.M1	(LS)				

[a)] ES: Spain, FR: France, GER: Germany, IT: Italy, UK: United Kingdom

[b)] Acronyms in the parentheses give the type of the outlier; AO: additive outlier, LS: level shift, TC: transitory change, SC: slope change

[c)] TRAMO: see Gómez and Maravall (1996); TSW: TRAMO–SEATS version for Windows, see Caporello and Maravall (2004); X13–ARIMA: see U.S. Census Bureau (2013); STAMP: see Koopman et al. (2009)

[d)] Time points of breaks in particular unobserved components of the BSM are translated to time points of changes in the observed series.

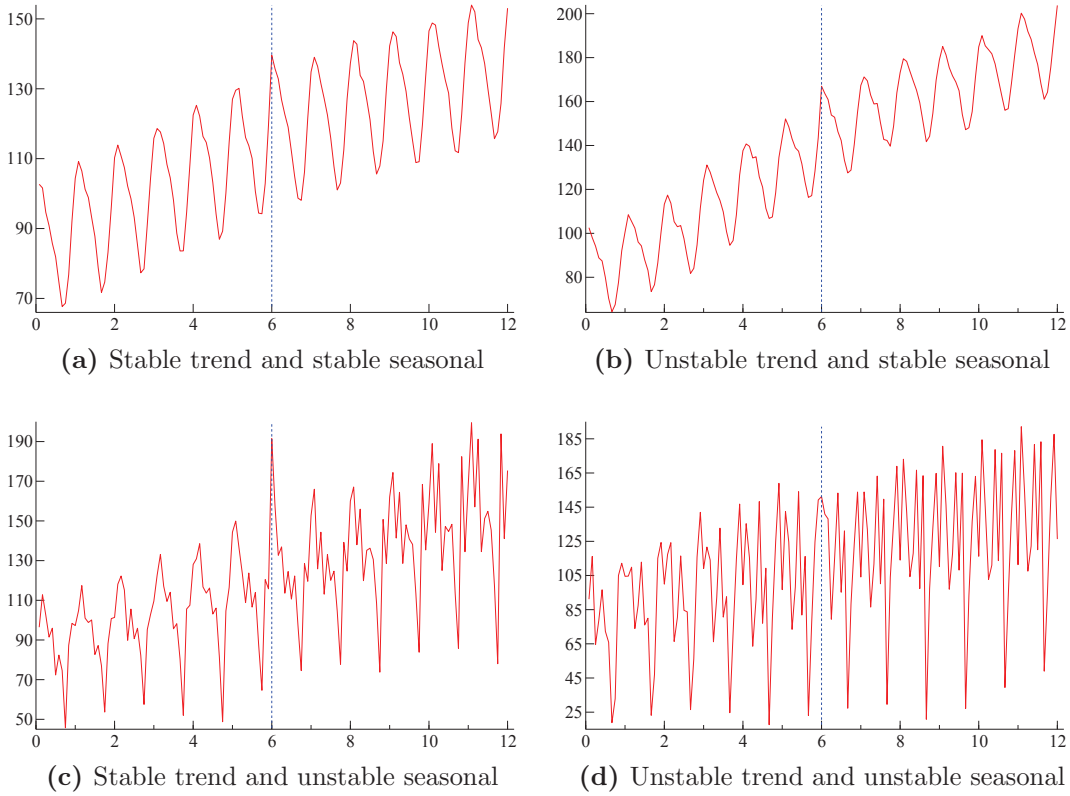
Table A.17: Periods of first detection of the relevant LS for five European countries[a),b)]

			Time point of LS	Time point of first detection
ES	SIS	non-seq.	–	–
		seq.	2008.M10	2009.M6
FR	SIS	non-seq.	–	–
		seq.	2008.M11	2008.M11
GER	SIS	non-seq.	2008.M12	2009.M2
		seq.	2008.M11	2008.M11
IT	SIS	non-seq.	2008.M12	2008.M12
		seq.	2008.M12	2008.M12
UK	SIS	non-seq.	–	–
		seq.	2008.M11	2008.M11

[a)] ES: Spain, FR: France, GER: Germany, IT: Italy, UK: United Kingdom

[b)] Relevant LS refers to the beginning of the economic crisis and and its time point, as detected by SIS, generally differs across countries. Detection is iteratively performed starting with the sample up to 2008.M9 and ending with the sample up to 2009.M8. Cases in which the relevant LS is not detected in the whole time span are indicated by –.

B Figures



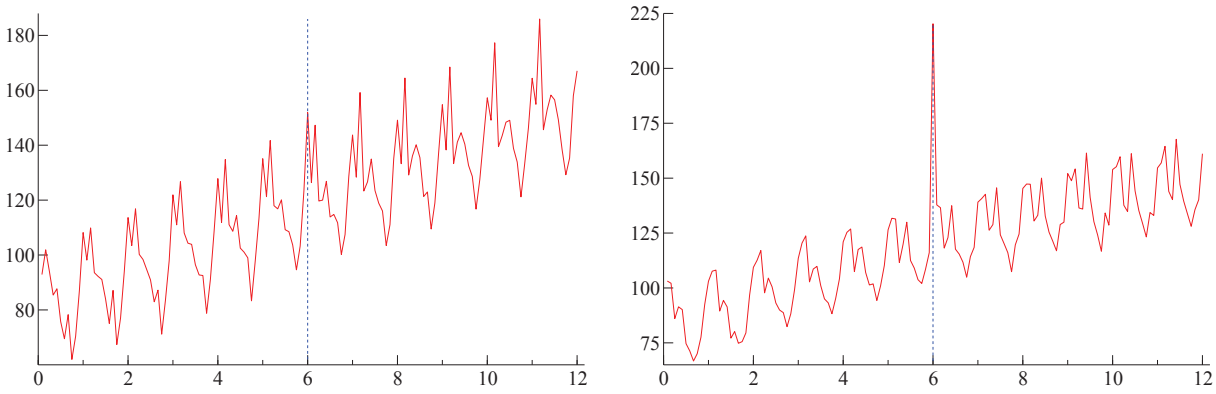
(a) Stable trend and stable seasonal

(b) Unstable trend and stable seasonal

(c) Stable trend and unstable seasonal

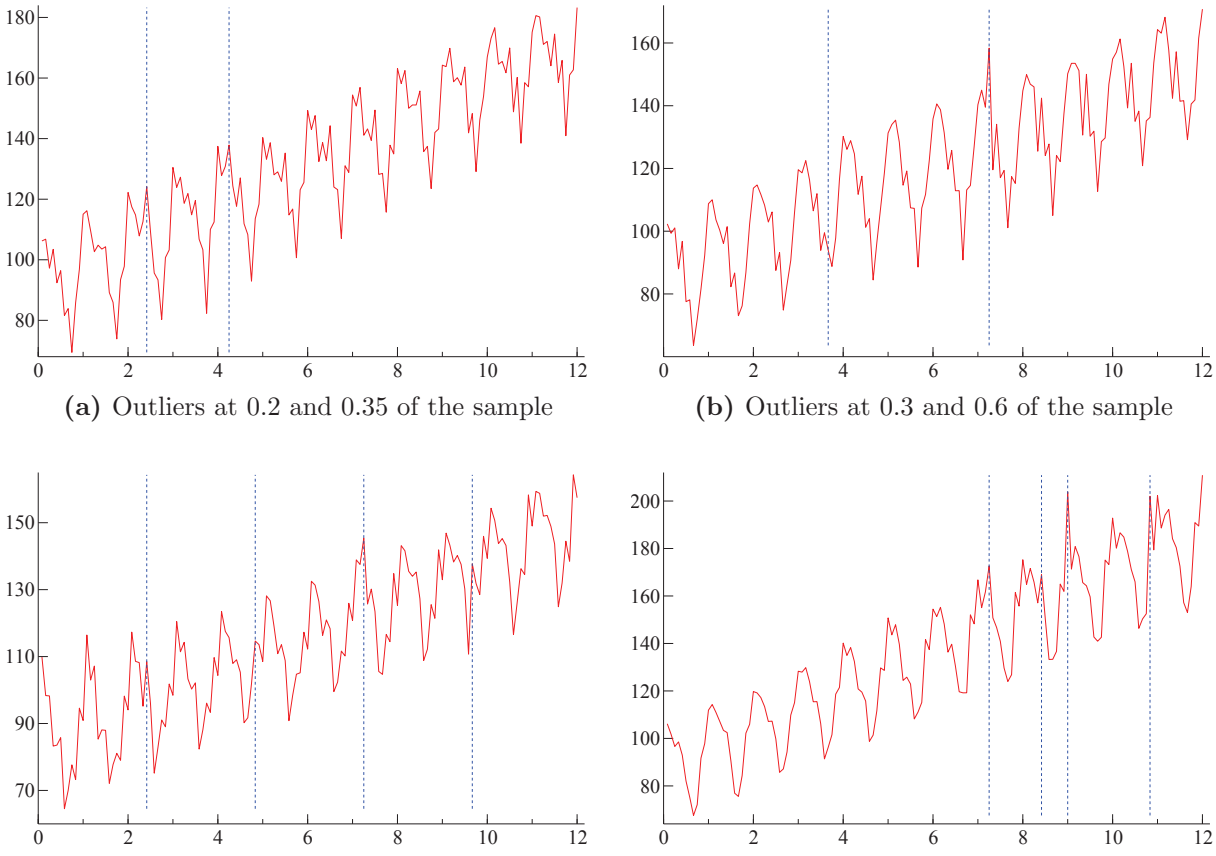
(d) Unstable trend and unstable seasonal

Figure B.1: Examples of series simulated with different variance parameters and an additive outlier at 0.5 of the sample



(a) Outlier magnitude given by $2 \cdot PESD$ **(b)** Outlier magnitude given by $14 \cdot PESD$

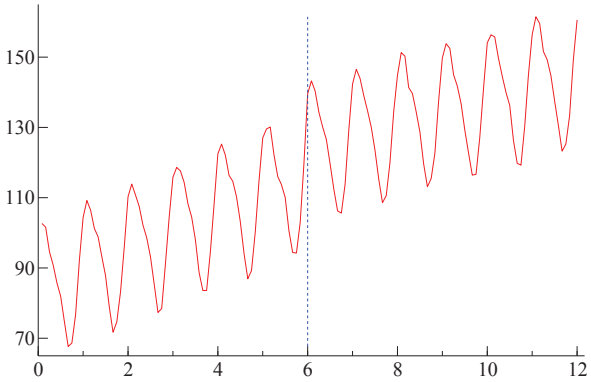
Figure B.2: Examples of series simulated with an additive outlier of two different magnitudes at 0.5 of the sample



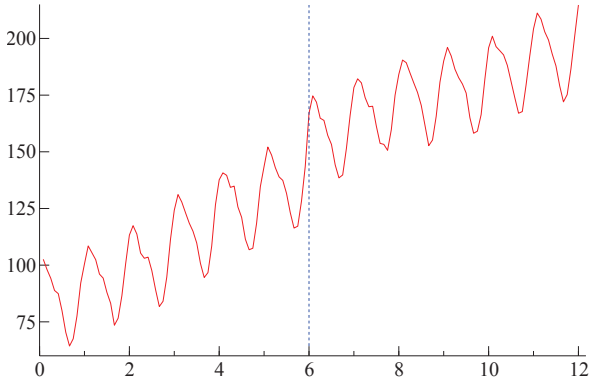
(a) Outliers at 0.2 and 0.35 of the sample **(b)** Outliers at 0.3 and 0.6 of the sample

(c) Outliers at 0.2, 0.4, 0.6 and 0.8 of the sample **(d)** Outliers at 0.6, 0.7, 0.75 and 0.9 of the sample

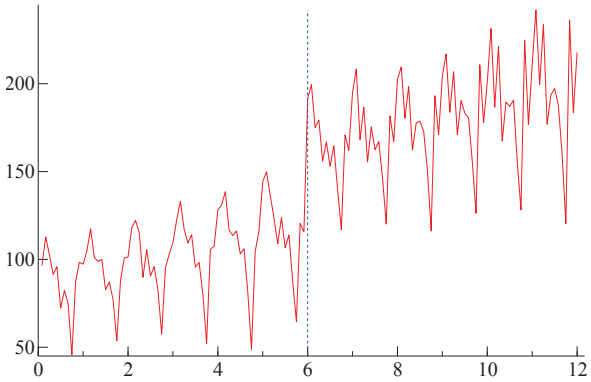
Figure B.3: Examples of series simulated with the benchmark specification and multiple outliers at different locations



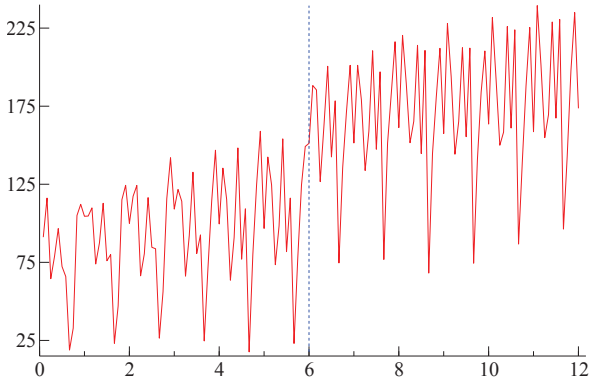
(a) Stable trend and stable seasonal



(b) Unstable trend and stable seasonal

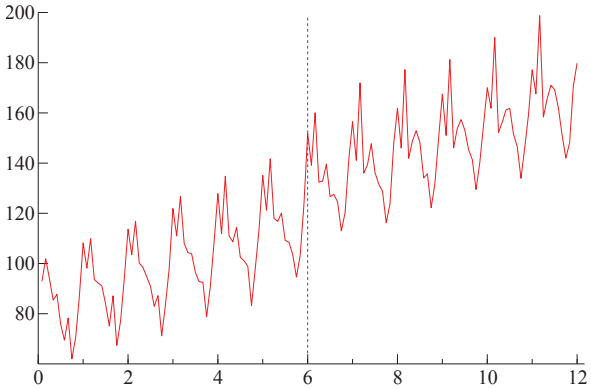


(c) Stable trend and unstable seasonal

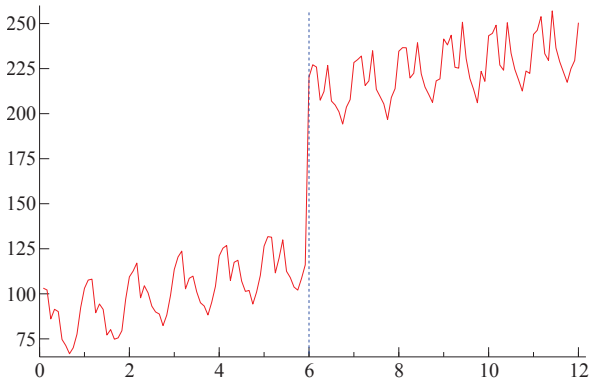


(d) Unstable trend and unstable seasonal

Figure B.4: Examples of series simulated with different variance parameters and a level shift at 0.5 of the sample

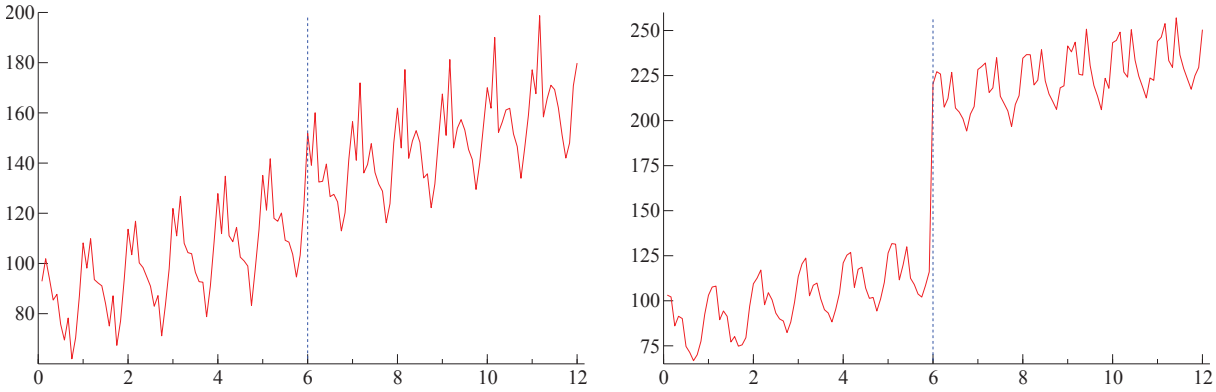


(a) Outlier magnitude given by $2 \cdot PESD$



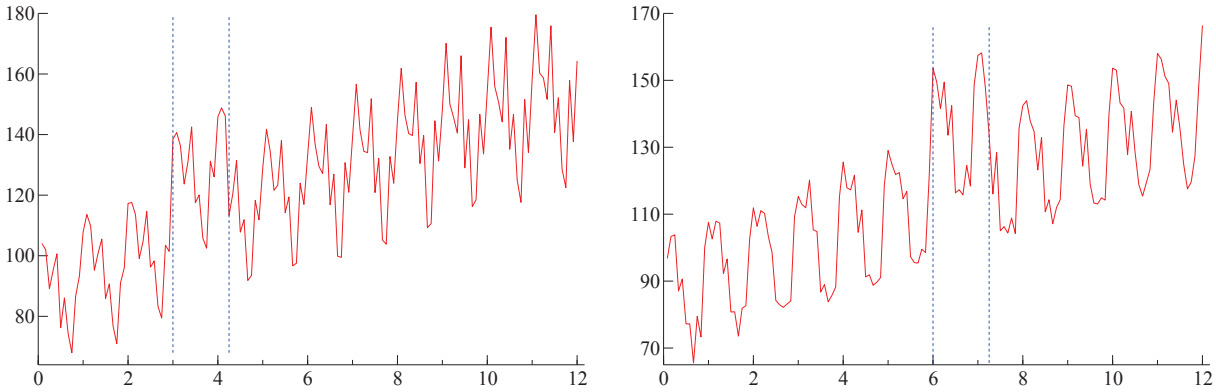
(b) Outlier magnitude given by $14 \cdot PESD$

Figure B.5: Examples of series simulated with a level shift of two different magnitudes at 0.5 of the sample

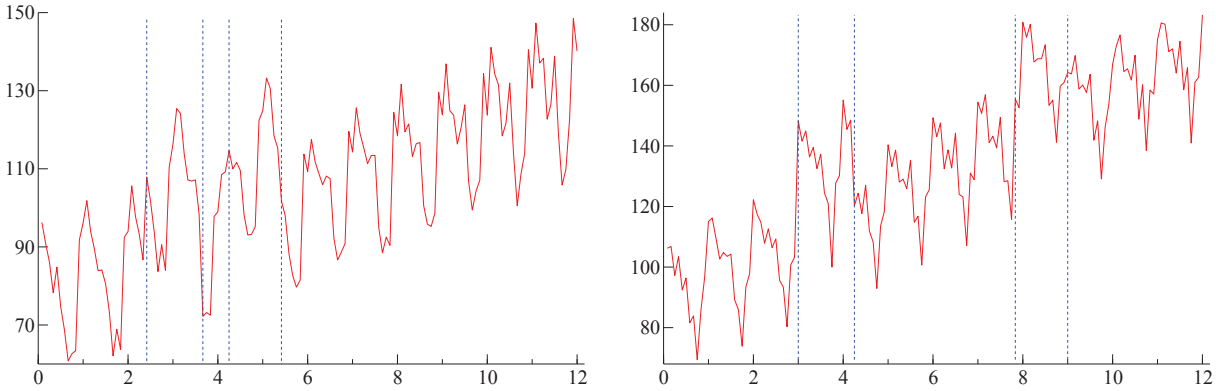


(a) Outlier magnitude given by $2 \cdot PESD$ **(b)** Outlier magnitude given by $14 \cdot PESD$

Figure B.6: Examples of series simulated with a level shift of two different magnitudes at 0.5 of the sample



(a) Level shift between 0.2 and 0.35 of the sample **(b)** Level shift between 0.5 and 0.6 of the sample



(c) Level shifts between 0.2 and 0.3, and between 0.35 and 0.45 of the sample **(d)** Level shifts between 0.25 and 0.35, and between 0.65 and 0.75 of the sample

Figure B.7: Examples of series simulated with the benchmark specification and multiple interval level shifts at different locations

References

Atkinson, A. C., Koopman, S. J., and Shephard, N. (1997). Detecting Shocks: Outliers and Breaks in Time Series. *Journal of Econometrics*, *80*(2), 387–422.

Caporello, G., and Maravall, A. (2004). *Program TSW: Revised Reference Manual* (Manual). Banco de España.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012). Model Selection When There Are Multiple Breaks. *Journal of Econometrics*, *169*(2), 239–246.

Chang, I., Tiao, G. C., and Chen, C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*, *30*(2), 193–204.

Chen, C., and Liu, L.-M. (1993). Joint Estimation of Model Parameters and Outliers Effects in Time Series. *Journal of the American Statistical Association*, *88*(421), 284–297.

Clements, M. P., and Hendry, D. F. (2011). Forecasting from Mis–specified Models in the Presence of Unanticipated Location Shifts. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook of Economic Forecasting* (chap. 10). Oxford: Oxford University Press.

Doornik, J. A. (2008). *Object-oriented Matrix Programming Ox 6.0*. London: Timberlake Consultants.

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

Doornik, J. A. (2009b). *Econometric Model Selection With More Variables Than Observations* (Unpublished paper). Economics Department, University of Oxford.

Doornik, J. A. (2009c). *Object-oriented Matrix Programming using Ox 7*. London: Timberlake Consultants Press.

Doornik, J. A., and Hendry, D. F. (2013). *PcGive 14*. London: Timberlake Consultants. (3 volumes)

Doornik, J. A., Hendry, D. F., and Pretis, F. (2013). *Step–Indicator Saturation* (Discussion Paper No. 658). University of Oxford.

Durbin, J., and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods* (second ed.; O. S. S. Series, Ed.). Oxford: Oxford University Press.

Ericsson, N. R., and Reisman, E. L. (2012). Evaluating a Global Vector Autoregression for Forecasting. *International Advances in Economic Research*, *18*, 247–258.

Gómez, V., and Maravall, A. (1996). *Programs TRAMO and SEATS; Instructions for the User* (Working Paper No. 9628). Servicio de Estudios, Banco de España.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C., and Koopman, J. S. (1992). Diagnostic Checking of Unobserved-Component Time Series Models. *Journal of Business and Economic Statistics*, *10*(4), 377–389.

Harvey, A. C., and Todd, P. H. J. (1983). Forecasting Econometric Time Series with Structural and Box-Jenkins Models (with discussion). *Journal of Business and Economic Statistics*, *1*(4), 299–315.

Hendry, D. F. (1999). An Econometric Analysis of US Food Expenditure, 1931–1989. In J. R. Magnus and M. S. Morgan (Eds.), *Methodology and Tacit Knowledge: Two Experiments in Econometrics* (pp. 341–361). Chichester: John Wiley and Sons.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic Selection of Indicators in a Fully Saturated Regression. *Computational Statistics*, *23*, 317–335.

Hendry, D. F., and Krolzig, H.-M. (2004). *Resolving Three 'Intractable' Problems using a Gets Approach* (Unpublished paper). Economics Department, University of Oxford.

Hendry, D. F., and Mizon, G. E. (2011). Econometric Modelling of Time Series with Outlying Observations. *Journal of Time Series Econometrics*, *3*(1). doi: 10.2202/19411928.1100

Hendry, D. F., and Pretis, F. (2011). *Anthropogenic Influences on Atmospheric CO_2* (Discussion Paper No. 584). University of Oxford.

Hendry, D. F., and Santos, C. (2010). An Automatic Test of Super Exogeneity. In M. W. Watson and T. Bollerslev (Eds.), *Volatility and Time Series Econometrics* (pp. 164–193). Oxford University Press.

Johansen, S., and Nielsen, B. (2009). An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry* (pp. 1–36). Oxford: Oxford University Press.

Koopman, S. J., Harvey, A. C., Doornik, J. A., and Shephard, N. (2009). *STAMP 8.2: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants.

Tsay, R. S. (1986). Time Series Model Specification in the Presence of Outliers. *Journal of the American Statistical Association*, *81*, 132–141.

U.S. Census Bureau. (2013). X–13 ARIMA–SEATS Reference Manual [Computer software manual].

West, M., and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models* (2nd ed.). New York: Springer–Verlag.

RECENT PUBLICATIONS BY *CEIS Tor Vergata*

Publish or Perish? Incentives and Careers in Italian Academia
Richard Damania, Pasquale Lucio Scandizzo and Ann Jeannette Glauber
CEIS Research Paper, *324*,August 2014

Publish or Perish? Incentives and Careers in Italian Academia
Daniele Checchi, Gianni De Fraja and Stefano Verzillo
CEIS Research Paper, *323*,August 2014

The Determinants of Household's Bank Switching
Marianna Brunetti, Rocco Ciciretti and Ljubica Djordjevic
CEIS Research Paper, *322*,August 2014

Direct and Indirect Cost of Diabetes in Italy: a Prevalence Probabilistic Approach
Andrea Marcellusi, Raffaella Viti, Alessandra Mecozzi and Francesco S. Mennini
CEIS Research Paper, *321*,August 2014

150 Years of Italian CO_2 Emissions and Economic Growth
Barbara Annicchiarico, Anna Rita Bennato and Emilio Zanetti Chini
CEIS Research Paper, *320*, July 2014

Exponential Smoothing, Long Memory and Volatility Prediction
Tommaso Proietti
CEIS Research Paper, *319*, July 2014

How To Safeguard World Heritage Sites? A Theoretical Model of "Cultural Responsibility"
Leonardo Becchetti, Nazaria Solferino, M. Elisabetta Tessitore
CEIS Research Paper, *318*, July 2014

Corruption in PPPs, Incentives and Contract Incompleteness
Elisabetta Iossa and David Martimort
CEIS Research Paper, *317*, July 2014

A Proof Without Words and a Maximum without Calculus
Eloisa Campioni and Luca Panaccione
CEIS Research Paper, *316*, May 2014

The causal effect of parents' schooling on children's schooling in Europe. A new IV approach
Enkelejda Havari and Marco Savegnago
CEIS Research Paper, *315*, May 2014

DISTRIBUTION

Our publications are available online at www.ceistorvergata.it

DISCLAIMER

The opinions expressed in these publications are the authors' alone and therefore do not necessarily reflect the opinions of the supporters, staff, or boards of CEIS Tor Vergata.

COPYRIGHT

Copyright © 2014 by authors. All rights reserved. No part of this publication may be reproduced in any manner whatsoever without written permission except in the case of brief passages quoted in critical articles and reviews.

MEDIA INQUIRIES AND INFORMATION

For media inquiries, please contact Barbara Piazzi at +39 06 72595652/01 or by e-mail at ufficiostampa@ceis.uniroma2.it. Our web site, www.ceistorvergata.it, contains more information about Center's events, publications, and staff.

DEVELOPMENT AND SUPPORT

For information about contributing to CEIS Tor Vergata, please contact Carmen Tata at +39 06 72595615/01 or by e-mail at segr.ceis@economia.uniroma2.it