JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12148

DETECTING VOLCANIC ERUPTIONS IN TEMPERATURE RECONSTRUCTIONS BY DESIGNED BREAK-INDICATOR SATURATION

Felix Pretis*
University of Oxford

Lea Schneider
Johannes Gutenberg University

Jason E. Smerdon
Lamont-Doherty Earth Observatory, Columbia University

David F. Hendry
University of Oxford

Abstract. We present a methodology for detecting breaks at any point in time-series regression models using an indicator saturation approach, applied here to modelling climate change. Building on recent developments in econometric model selection for more variables than observations, we saturate a regression model with a full set of designed break functions. By selecting over these break functions using an extended general-to-specific algorithm, we obtain unbiased estimates of the break date and magnitude. Monte Carlo simulations confirm the approximate properties of the approach. We assess the methodology by detecting volcanic eruptions in a time series of Northern Hemisphere mean temperature spanning roughly 1200 years, derived from a fully coupled global climate model simulation. Our technique demonstrates that historic volcanic eruptions can be statistically detected without prior knowledge of their occurrence or magnitude- and hence may prove useful for estimating the past impact of volcanic events using proxy reconstructions of hemispheric or global mean temperature, leading to an improved understanding of the effect of stratospheric aerosols on temperatures. The break detection procedure can be applied to evaluate policy impacts as well as act as a robust forecasting device.

Keywords. Break; Climate; Econometrics; Model Selection; Indicator Saturation; Volcano

1. Introduction

Breaks in time series come in many shapes and may occur at any point in time – distorting inference in-sample and leading to forecast failure out-of-sample if not appropriately modelled. Often an approximate

*Corresponding author contact email: felix.pretis@nuffield.ox.ac.uk; Tel.: +44 784 722 1783

Journal of Economic Surveys (2016) Vol. 30, No. 3, pp. 403–429
© 2016 John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

shape of a break can be postulated *a priori*, either from previous observations or theory. For example, smooth transitions are common in economic time series following recessions or policy interventions, while sudden drops followed by smooth reversions to the mean are typical in climate time series such as temperature records after a large volcanic eruption (e.g. Kelly and Sear, 1984). While the approximate form of a break may be known, the timings and magnitudes of breaks are often unknown. Here, we propose an econometric approach for detecting breaks of any specified shape in regression models using an indicator saturation procedure. Our approach is based on recent developments in variable selection within regression models that involve more variables than observations (Castle *et al.*, 2011). By selecting over a complete set of designed break indicators, our approach produces estimates of the break magnitude and timing without imposing limits on the number of breaks that may occur, even at the start or end of a sample.

A structural break is defined as a time-dependent change in a model parameter resulting from a change in the underlying data generating process (DGP). For example, a volcanic eruption leading to a rapid climatic cooling corresponds to a temporary shift in the mean of the surface temperature process. The detection of structural breaks in time series has received significant attention in the recent literature – with a growing interest in econometric models of climate change (e.g. Estrada *et al.*, 2013; Pretis *et al.*, 2015a). The focus has primarily remained on breaks in the mean through the form of step functions (Step-Indicator Saturation – SIS, Castle *et al.*, 2015b; Pretis, 2015b), smooth transition functions (González and Teräsvirta, 2008), breaks in regression coefficients (see, e.g. Bai and Perron, 1998, 2003; Perron and Zhu, 2005; Perron and Yabu, 2009), or individual outliers or groups of outliers that can be indicative of different forms of breaks (Impulse-Indicator Saturation – IIS, see Hendry *et al.*, 2008).

Broadly grouped into 'specific-to-general' and 'general-to-specific', there exist a plethora of approaches for the detection of structural breaks. Perron (2006) provides a broad overview of specific-to-general methods, some of which are subject to an upper limit on the number of breaks, a minimum break length, co-breaking restrictions, as well as ruling out breaks at the beginning or end of the sample, though Strikholm (2006) proposes a specific-to-general algorithm that allows for breaks at the start or end of a sample and relaxes the break length assumption.

Indicator saturation (IIS, SIS) provides an alternative approach using an extended general-to-specific methodology based on model selection. By starting with a full set of step indicators in SIS and removing all but significant ones, structural breaks can be detected without having to specify a minimum break length, maximum break number or imposed co-breaking. Crucially this also allows model selection to be conducted jointly with break detection as non-linearities, dynamics, theory-motivated variables and break functions are selected over simultaneously.

Step functions and impulses are nevertheless only the simplest of many potential break specifications and may not provide the closest approximation to the underlying break. Ericsson (2012) proposes a wide range of extensions to impulse and step shifts. Here, we show that the principle of SIS can be generalized to any form of deterministic break function. An advantage over existing methods is an expected higher frequency of detection when a break function approximates the true break,[1] high flexibility as multiple types of break functions can be selected over and improvements in forecasting where designed functions act as continuous intercept corrections. Moreover, by being a structured search, the retention of irrelevant effects can be controlled.

The method is illustrated using an econometric model of climate variation – detecting volcanic eruptions in a time series of Northern Hemisphere (NH) mean temperature spanning roughly 1200 years, derived from a fully coupled global climate model simulation. Our technique demonstrates that eruptions can be statistically detected without prior knowledge of their occurrence or magnitude– and hence may prove useful for estimating the past impacts of volcanic events using proxy reconstructions of hemispheric or global mean temperatures. Specifically, this can lead to an improved understanding of the effect of stratospheric aerosols on temperatures (with relevance to geo-engineering and pollution control), and more generally, the break detection procedure can be applied to evaluate policy interventions (e.g. the

© 2016 John Wiley & Sons Ltd

Montreal Protocol: see Estrada *et al.*, 2013; and Pretis and Allen, 2013), correct for measurement changes by detecting and subsequently removing shifts, and function as a robust forecasting device.

Section 2 introduces the methodology and investigates the properties of break detection in the presence of breaks and under the null of no breaks. Section 3 applies the method to detect volcanic eruptions in simulated climate data, and considers designed indicator functions as a robust forecasting device. The conclusions of our work are discussed in Section 4.

2. Break Detection Using Designed Indicator Functions

Breaks are intrinsically stochastic without prior knowledge of their timings and magnitudes. Using a full set of break functions allows us to model the responses deterministically. The detection of structural breaks in regression models can be formulated as a model selection problem where we select over a full set of break functions, a subset of which accurately describes the underlying 'true' break. Consider a simple model as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are $(T \times 1)$ vectors, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector and $\mathbf{Z}$ is a $(T \times k)$ matrix $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$ of rank k. We investigate the presence of structural breaks in any of the $\boldsymbol{\beta}$ where $\mathbf{z}$ may be a constant, trend or random variable. For each break type at any point in time for each variable whose coefficient is allowed to break, we augment the above model by a $(T \times T)$ break function matrix $\mathbf{D}$:[2]

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{2}$$

where $\boldsymbol{\gamma}$ is a $(T \times 1)$ vector. The specification of $\mathbf{D}$ is such that the first column $\mathbf{d}_1$ $(T \times 1)$ is set to denote some specified break function $d(t)$ of length L, where $d_{1,t} = d(t)$ for $t \leq L$ and 0 otherwise, $d_{1,t} = 0$ for $t > L$. All further columns $\mathbf{d}_j$ (for $j = 2, \ldots, T$) in $\mathbf{D}$ are set such that $d_{j,t} = d_{j-1,t-1}$ for $t \geq j$ and 0 otherwise. The break matrix $\mathbf{D}$ is then defined as $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_T)$, where $\mathbf{d}_j$ denotes a vector with break at time $t = j$:

$$\begin{aligned}
\mathbf{D} &= (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_T) \\
\mathbf{d}_1 &= (d_1, d_2, \ldots, d_{L-1}, d_L, 0, \ldots, 0)' \\
\mathbf{d}_2 &= (0, d_1, d_2, \ldots, d_{L-1}, d_L, 0, \ldots, 0)' \\
\mathbf{d}_3 &= (0, 0, d_1, d_2, \ldots, d_{L-1}, d_L, 0, \ldots, 0)' \\
&\vdots
\end{aligned} \tag{3}$$

This specification provides a general framework within which multiple break types can be analysed – Table 1 provides a non-exhaustive overview.[3]

The form of the break function $d(t)$ has to be designed *a priori*, but this is implicitly done in most structural break detection methods. For example, outlier detection through finding impulses (IIS, in Hendry *et al.*, 2008) sets the break vector in $\mathbf{d}_1$ such that $d(t) = 1$ and $L = 1$, while a search for step shifts (SIS) sets $d(t) = 1$ and the length to $T - t + 1$, that is, the break function continues until the end of the sample. Breaks in linear trends (see, e.g. Perron and Zhu, 2005; Perron and Yabu, 2009; Estrada *et al.*, 2013) can be constructed by setting $d(t) = t$ and the length to $T - t + 1$. Pretis *et al.* (2015a) apply indicator saturation using broken trends and step shifts to evaluate climate models. Breaks in coefficients on random variables z_t (see, e.g. Bai and Perron, 2003; Ericsson, 2012; Kitov and Tabor, 2015) can be constructed by interacting z_t with a full set of step shifts. Sudden declines followed by a smooth recovery to the mean in hemispheric temperature responses are introduced here as volcanic functions and

© 2016 John Wiley & Sons Ltd

Table 1. Break Function Specifications

	Break Value: $d(t)$	Length:	
Deterministic Breaks			$D = \begin{pmatrix} d_1 & 0 & \dots & \dots & \dots & 0 \\ d_2 & d_1 & 0 & \dots & \dots & \vdots \\ \vdots & d_2 & d_1 & 0 & \dots & \vdots \\ d_L & \vdots & d_2 & d_1 & 0 & \vdots \\ \vdots & d_L & \vdots & d_2 & d_1 & 0 \\ 0 & 0 & d_L & d_3 & d_2 & d_1 \end{pmatrix}$
General Case	$d(t)$	L	
Impulses (IIS)	1	1	
Step Shifts (SIS)	1	$T-t+1$	
Broken Trends	t	$T-t+1$	
Volcanic Functions	see equation (31)	3	
Random Variables			
Coeff. on z_t (MIS)	$z_t \cdot d_{t,SIS}$	$T-t+1$	

considered in Sections 2.1 and 3. Linear combinations of multiple break functions can allow for varying lengths of breaks without pre-specification.

Searching for breaks in k variables implies that the complete break matrix across all k variables $\mathbf{D}$ is of dimension $(T \times kT)$. The inclusion of kT additional variables leads to the total number of variables N exceeding the number of observations, $N > T$, even for $k = 1$. Thus, a methodology allowing for more variables than observations is required.

Selection of models with more variables than observations has primarily relied on either shrinkage-based penalized likelihood methods (Tibshirani, 1996; Zou and Hastie, 2005; Tibshirani, 2011) or general-to-specific methodology in the econometrics literature (see, e.g. Castle *et al.*, 2011). Kock and Teräsvirta (2015) compare the general-to-specific model selection algorithms *Autometrics*, to *QuickNet* – an artificial neural-network method proposed by White (2006), and to a shrinkage-based bridge estimator from Huang *et al.* (2008) in the context of forecasting.

Here we rely on general-to-specific model selection due to methods based on forward stepwise searches not performing as well in break detection contexts (see Section 2.1.2 for a simple comparison, or Epprecht *et al.* (2013), and Hendry and Doornik (2014) for comparisons on general variable selection). Cox and Snell (1974) discuss some of the challenges of the general variable selection problem and Hoover and Perez (1999) show the feasibility of general-to-specific model selection for $N \ll T$. When facing more variables than observations, the general-to-specific approach is closely linked to robust statistics. Saturating a model with a full set of $0/1$ indicator functions from which selections are made is equivalent to a robust one-step M-estimator using Huber's skip function (see Johansen and Nielsen, 2009, 2013 for the iterated case, and Johansen and Nielsen, 2016 for an overview). Here, we generalize this allowing for any form of designed break function in place of impulses, and formulate break detection as a model selection problem.

To estimate model (2) saturated with a full set of break functions $\mathbf{D}$ (so $N > T$), we rely on a block-partitioning estimation procedure (Doornik, 2010; Hendry and Johansen, 2015). For this, we partition $\mathbf{D}$ into b blocks of n_i variables such that $n_i \ll T$ and $\sum_{i=1}^{b} n_i = N$. In the simplest case of testing for a break in a single variable (e.g. the intercept), a split-half approach (see Figure 1 and *Algorithm 1* in the supplementary material) is feasible: initially, we include the first half of $\mathbf{D}_1 = (\mathbf{d}_1, \dots, \mathbf{d}_{T/2})$ and retain only significant break indicators. We repeat the step for the second half of break functions d_j (for $j = T/2 + 1, \dots T$) and finally combine the retained sets and only keep significant indicators. This split-half approach is considered here for analytical tractability in Section 2.1.2.

In practice, however, we rely on a multi-split and multi-path search to lower the variance of the estimators, allow for any number of variables for a given set of observations and to avoid a breakdown of the procedure if the breaks cannot be adequately modelled through split-half indicators.[4] This can be implemented through the general-to-specific model selection algorithm *Autometrics* (*Algorithm 2* in

© 2016 John Wiley & Sons Ltd

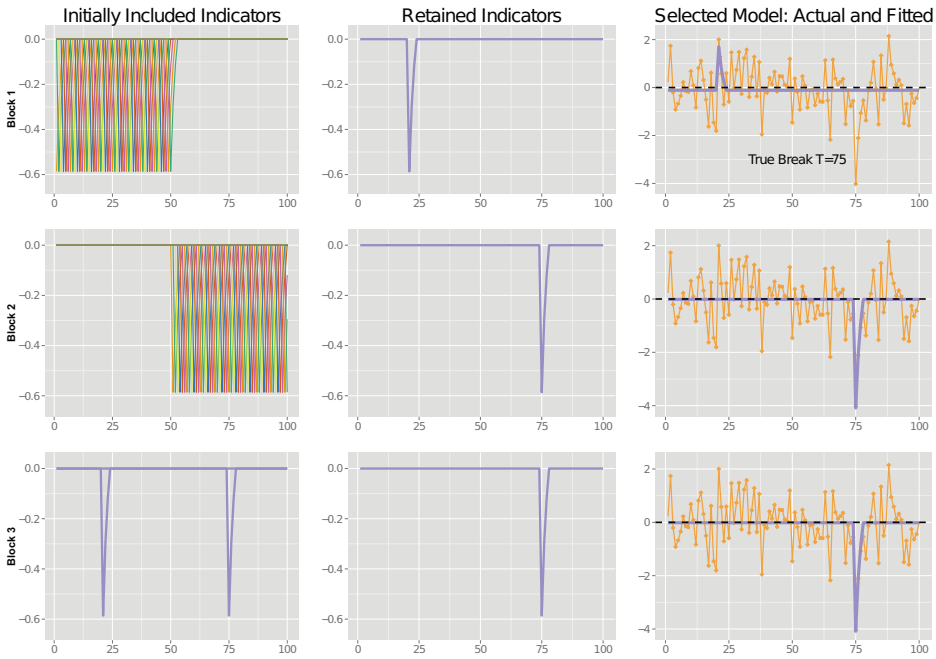

Figure 1. Split-Half Approach for a Single Unknown Break of the Shape of a Volcanic Function at $T = 75$.

Note: Left column shows included indicators in each step, middle column shows the retained indicators and right column graphs the selected model with actual and fitted data. Block 1 (top panel) includes the first half of break functions and retains a single one as the mean is lowered in the second half due to the presence of a break at $t = 75$. Block 2 (middle panel) then includes the second half retaining the correct break function. Block 3 uses the union of retained indicators from blocks 1 and 2 in which now the first indicator is rendered insignificant by the mean being correctly estimated due to the second indicator capturing the break. Using a saturating set of break functions at 1% the break at $T = 75$ is detected without prior knowledge and is the only break function retained.

the supplementary material), described in Doornik, 2009a, or the *gets* package in the statistical software environment *R* (Pretis *et al.*, 2016). The algorithm (referred to as multi-path throughout the paper) avoids path dependence through a tree-structure and uses a parallel stepwise backwards search, while testing for encompassing and congruence (Hendry, 1995). See Hendry and Pretis (2013) for an application of the algorithm to econometric modelling of climate change. A simulation-based comparison to shrinkage methods is provided in Section 2.1.2.

2.1 *Properties of Designed Break Functions in the Presence of Breaks*

To assess the theoretical power of the proposed methodology, we first investigate the properties in the benchmark case of a single break matched by a correctly timed break indicator in Section 2.1.1. Section 2.1.2 then assesses the properties of break-indicator saturation when the break date and magnitude are unknown. Section 2.1.3 investigates uncertainty around the break date and 2.2 describes the properties in

© 2016 John Wiley & Sons Ltd

the presence of no breaks. Theory results are derived for general designed functions, simulation examples are based on a volcanic break as characterized by equation (31).

2.1.1 *Power for Known Break Date*

We investigate the theoretical power of detecting a break in a time series given a known break date. Consider a DGP coinciding with the model for a single known break in an intercept:

$$y_t = \mu + \lambda d_t + \epsilon_t \tag{4}$$

where $\epsilon_t \sim \text{IN}(0, \sigma_\epsilon^2)$. The break shifts μ to $\mu + \lambda d_t$ where d_t is a break function of length L beginning at time $t = T_1$ where $(T_1 + L) \leq T$ such that $d_t \neq 0$ for $T_1 \leq t < (T_1 + L)$ and 0 otherwise. The estimators $\widehat{\mu}$ and $\widehat{\gamma}$ (where $\widehat{\gamma}$ is the estimator for λ) in a correctly specified model for a known break are given by

$$\begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma} - \lambda \end{pmatrix} = \begin{pmatrix} T_d^{-1} \left(\sum_{t=T_1}^{T_1+L-1} d_t^2 \sum_{t=1}^{T} \epsilon_t - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=T_1}^{T_1+L-1} d_t \epsilon_t \right) \\ T_d^{-1} \left(\sum_{t=T1}^{T_1+L-1} d_t \epsilon_t - \sum_{t=T1}^{T_1+L-1} d_t \sum_{t=1}^{T} \epsilon_t \right) \end{pmatrix} \tag{5}$$

where $T_d = T[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \frac{1}{T}(\sum_{t=T_1}^{T_1+L-1} d_t)^2]$. The estimators are unbiased for the break and intercept: $E[\widehat{\mu} - \mu] = 0$ and $E[\widehat{\gamma} - \lambda] = 0$. The variance of the estimators is given by

$$V \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma} - \lambda \end{pmatrix} = \sigma_\epsilon^2 T_d^{-1} \begin{pmatrix} \sum_{t=T_1}^{T_1+L-1} d_t^2 & -\sum_{t=T_1}^{T_1+L-1} d_t \\ -\sum_{t=T_1}^{T_1+L-1} d_t & T \end{pmatrix} \tag{6}$$

The distribution of the break estimator is then:

$$(\widehat{\gamma} - \lambda) \sim N \left(0, \sigma_\epsilon^2 \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \sum_{t=T_1}^{T_1+L-1} d_t \bar{d} \right]^{-1} \right) \tag{7}$$

where $\bar{d} = 1/T \sum_{t=1}^{T} d_t$. For the special case when step-indicators are chosen as the functional form of d_t and the single break lasts from $t = 0$ to $t = T_1 < T$, equation (5) simplifies to $\widehat{\mu} - \mu = \bar{\epsilon}_2$ and $\widehat{\gamma} - \lambda = \bar{\epsilon}_1 - \bar{\epsilon}_2$, where $\bar{\epsilon}_1 = 1/T_1 \sum_{t=1}^{T_1} \epsilon_t$ and $\bar{\epsilon}_2 = 1/(T - T_1) \sum_{t=T_1+1}^{T} \epsilon_t$.[5]

2.1.2 *Potency for an Unknown Break Date*

When the break date is unknown, we propose to saturate the regression model using a full set of specified break indicators and select significant breaks through an extended general-to-specific algorithm (Castle *et al.*, 2011). We assess the methodology in the selection context using two main concepts: the null retention frequency of indicators is called the gauge, comparable to the size of a test denoting its (false) null rejection frequency, but taking into account that indicators that are insignificant on a pre-assigned criterion may nevertheless be retained to offset what would otherwise be a significant misspecification test (see Johansen and Nielsen, 2016, for distributional results on the gauge). The non-null retention frequency when selecting indicators is called its potency, comparable to a similar test's power for rejecting a false null hypothesis.

Here we investigate the feasibility of the proposed method by deriving the analytical properties of the split-half approach for an unknown break. Figure 1 illustrates the split-half method for a single unknown break. In practice, we rely on a multi-path, multi-block search algorithm (such as *Autometrics*, see *Algorithm 2* in the supplementary material) to reduce the variance of the estimators.

© 2016 John Wiley & Sons Ltd

Consider a single break falling into the first half of the sample beginning at time T_1 for L periods such that $0 < T_1 < T_1 + L < T/2$. In matrix form, the DGP is given as

$$\mathbf{y} = \lambda \mathbf{d}_{T_1} + \boldsymbol{\epsilon} \tag{8}$$

where $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ for simplicity and the $(T \times 1)$ vector $\mathbf{d}_{T_1}$ denotes a break at $t = T_1$ for L periods. Using a split-half approach, we assess the properties of detecting the single break when the break date is unknown. The split-half model for the first half of break functions is

$$\mathbf{y} = \mathbf{D}_1 \boldsymbol{\gamma}_{(1)} + \mathbf{v} \tag{9}$$

where $\boldsymbol{\gamma}_{(1)} = (\gamma_1, \gamma_2, \cdots \gamma_{T/2})'$ and $\mathbf{D}_1 = (\mathbf{d}_1, \ldots, \mathbf{d}_{T/2})$. The estimator $\widehat{\boldsymbol{\gamma}}_{(1)}$ equals:[6]

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_{(1)} &= \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} \mathbf{D}_1'\mathbf{y} = \lambda \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} \mathbf{D}_1'\mathbf{d}_{T_1} + \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} \mathbf{D}_1'\boldsymbol{\epsilon} \\ &= \lambda \mathbf{r} + \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} \mathbf{D}_1'\boldsymbol{\epsilon} \end{aligned} \tag{10}$$

where the $(T/2 \times 1)$ vector $\mathbf{r}$ is equal to one at $t = T_1$ and zero otherwise, $r_t = 1_{\{t=T_1\}}$. It follows that $E[\widehat{\boldsymbol{\gamma}}_{(1)}] = \lambda \mathbf{r}$ and $V[\widehat{\boldsymbol{\gamma}}_{(1)}] = \sigma_\epsilon^2 (\mathbf{D}_1'\mathbf{D}_1)^{-1}$. We find for the first half, for normal error terms:

$$\left(\widehat{\boldsymbol{\gamma}}_{(1)} - \lambda \mathbf{r}\right) \sim N\left(\mathbf{0}, \sigma_\epsilon^2 \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\right) \tag{11}$$

Therefore conventional t-tests can be used to assess the significance of individual indicators. The estimator $\widehat{\boldsymbol{\gamma}}_{(2)}$ on the second half of indicators, $\mathbf{D}_2 = (\mathbf{d}_{T/2+1}, \ldots, \mathbf{d}_T)$, will miss the break in the DGP in the first half described by $\mathbf{d}_{T_1}$ and equals

$$\widehat{\boldsymbol{\gamma}}_{(2)} = \lambda \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1} \mathbf{D}_2'\mathbf{d}_{T_1} + \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1} \mathbf{D}_2'\boldsymbol{\epsilon} \tag{12}$$

For step shifts, Castle *et al.* (2015b) show that the indicator in $\mathbf{D}_2$ closest to the sample split will be retained in the second set of indicators. For the general form of break functions, retention in $\mathbf{D}_2$, when there is a break in the first half, will depend on the specific functional form. However, conditional on the break indicator being correctly retained in the first set $\mathbf{D}_1$, retention of irrelevant indicators in $\mathbf{D}_2$ does not affect the correct identification of the break overall: let $\mathbf{D}_{1*}$ and $\mathbf{D}_{2*}$ denote the set of retained break functions in the first and second set, respectively, where retention is based on a retention rule such as $\mathbf{d}_j$ is retained if $|t_{\widehat{\gamma}_j}| \geq c_\alpha$. The final step in the split-half procedure is then to combine the retained indicators using $\mathbf{D}_U = [\mathbf{D}_{1*}\mathbf{D}_{2*}]$ and estimate the model:

$$\mathbf{y} = \mathbf{D}_U \boldsymbol{\gamma}_{(U)} + \mathbf{v} \tag{13}$$

This yields the estimator $\widehat{\boldsymbol{\gamma}}_{(U)}$ unbiased for the true break:[7]

$$\widehat{\boldsymbol{\gamma}}_{(U)} = \lambda \mathbf{r} + \left(\mathbf{D}_U'\mathbf{D}_U\right)^{-1} \mathbf{D}_U'\boldsymbol{\epsilon} \tag{14}$$

The carried-forward break function in $\mathbf{D}_{1*}$ correctly identifies the true break, and coefficients on all other break functions will thus be zero in expectation. The proof is identical to that given for the first half of indicators in the supplementary material. This shows that, conditional on retaining the correct break indicator in $\mathbf{D}_1$, the retention of indicators in $\mathbf{D}_2$ does not affect the correct identification of the break, when the first and second set are combined and reselected over. The distribution of the final split-half estimator is then given by

$$\left(\widehat{\boldsymbol{\gamma}}_{(U)} - \lambda \mathbf{r}\right) \sim N\left(\mathbf{0}, \sigma_\epsilon^2 \left(\mathbf{D}_U'\mathbf{D}_U\right)^{-1}\right) \tag{15}$$

Reselection then results in only the true break indicator being retained in expectation.[8]

This result generalizes the specific case of step indicators presented in Castle *et al.* (2015b). Even though the break date and magnitude are unknown, the use of a fully saturated set of break indicators

© 2016 John Wiley & Sons Ltd

Table 2. Potency of Detecting an Unknown Break When Using Split-Half and Multi-Path Searches.

	Split-Half		Multi-Path	
	Potency	Gauge $\mathbf{D}_1$	Potency	Gauge
$\lambda = 6$, trough $= 3.48$	0.69	0.013	0.88	0.015
$\lambda = 4$, trough $= 2.23$	0.30	0.013	0.50	0.014
$\lambda = 2$, trough $= 1.16$	0.06	0.013	0.11	0.015

Notes: Statistics were generated from 1000 simulations and detection significance was set to $\alpha = 0.01$, with a length of $L = 3$. Break magnitude λ corresponds to the full response in standard deviations of the error term ($\sigma_\epsilon = 1$) over the entire break, the trough is 0.58λ.

allows us to obtain an unbiased estimate of the break magnitude and timing. The estimator then follows a normal distribution subject to correct specification of the break function. Thus the estimated coefficient at the break time, $\widehat{\gamma}_{T_1}$, is in expectation equal to the break magnitude, while all other estimated coefficients are mean-zero in expectation. This result generalizes to multiple breaks falling in a single split. As in the case of the known break timing, the variance of the estimator depends on the specified break function. Let $\delta_{k,j}$ denote the (k, j) element of the matrix $(\mathbf{D}_1'\mathbf{D}_1)^{-1}$. The variance of the coefficient at the breakpoint in the first half is therefore:

$$V[\widehat{\gamma}_{T_1}] = \sigma_\epsilon^2 \delta_{T_1,T_1} \tag{16}$$

For iid error terms $\boldsymbol{\epsilon}$, and $\mathbf{D}$ specified as a full set of step functions, the split-half model (without selection) yields $\delta_{j,j} = 2$, so the break coefficient has twice the error variance. For the proposed volcanic function (derived and assessed in detail in Section 3) modelling a single drop followed by a reversion to the mean, we find that $\delta_{j,j} = 3.7$, thus $V[\widehat{\gamma}_{T_1}] = 3.7\sigma_\epsilon^2$. This can be compared to the known-break/single-indicator case where the variance is given by equation (6) and for the volcanic function equals $2.3\sigma_\epsilon^2$ (for $T = 100$). Due to collinearity of break functions, the variance of the estimator is higher in a fully saturated model. In the more general case, δ_{T_1,T_1} depends on the specification of the break function but can be computed *a priori*. The t-statistic is then given as

$$t_{\widehat{\gamma}_{T_1}} = \frac{\widehat{\gamma}_{T_1}}{\widehat{\sigma}_\epsilon\sqrt{\delta_{T_1,T_1}}} \approx \frac{\left(\widehat{\gamma}_{T_1} - \lambda\right)}{\sigma_\epsilon\sqrt{\delta_{T_1,T_1}}} + \frac{\lambda}{\sigma_\epsilon\sqrt{\delta_{T_1,T_1}}} \sim N\left(\frac{\lambda}{\sigma_\epsilon\sqrt{\delta_{T_1,T_1}}}, 1\right) \tag{17}$$

In practice, we use sequential elimination of the break indicators or a multi-path search to eliminate insignificant indicators reducing the variance of the estimators from a saturated model (16) closer to the single break (6) and increasing the power of detection.

For dynamic time-series models, the above approach can be extended by including time-dependent covariates. Valid conditioning (e.g. through the inclusion of auto-regressive terms in the case of non-iid errors) can be ensured by always including the covariates in each block estimation step and only selecting over the break functions. Johansen and Nielsen (2009) provide the asymptotics under the null of no break for the special case of impulses for stationary and unit-root non-stationary autoregressive processes (see Johansen and Nielsen, 2013, for the iterated version). The case for general break functions is discussed in Section 2.2, and the supplementary material provides simulation results for an AR(1) model and DGP.[9]

Simulation Performance based on Volcanic Break Functions. Table 2 reports simulation results ($T = 100$) for a DGP with a single unknown volcanic break at $t = T_1 = 25$ of magnitude λ followed by a smooth reversion to the mean.[10] Equation (31) provides the exact functional form. Simulations are assessed by

© 2016 John Wiley & Sons Ltd

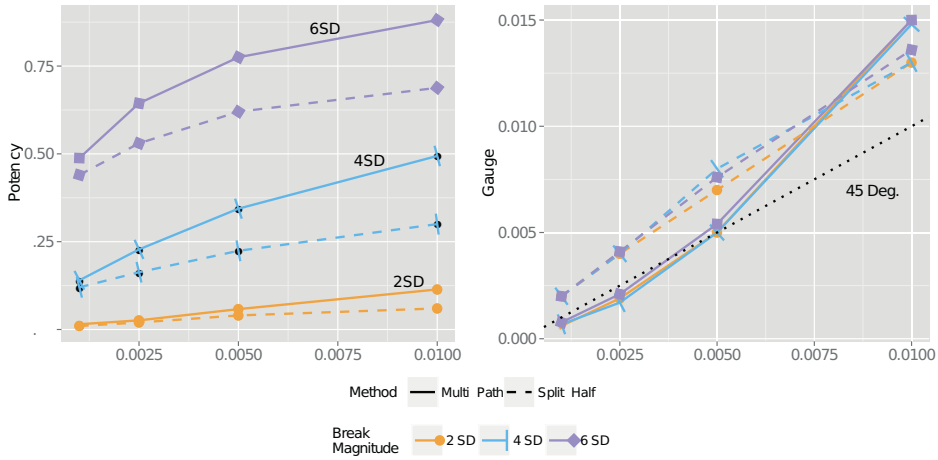

Figure 2. (Left) Potency of Detecting a Volcanic Break of Magnitude λ for Level of Significance α Using Split-Half and Multi-Path Selection and (Right) Proportion of Spuriously Retained Break Indicators (Gauge).

Note: Break magnitude λ corresponds to the full response in standard deviations of the error term ($\sigma_\epsilon = 1$) over the entire break, the trough is 0.58λ, 6 standard deviations (SD) therefore refers to a trough of 3.48SD.

the retention/detection frequency (potency) for a single break and average retention of spurious breaks (gauge).[11]

The trade-off between potency and level of significance of selection α is shown in Figure 2 for a single volcanic break. A multi-path search generally increases the power of detection relative to the split-half approach. Figure 3 shows the results for split-half (dashed) and multi-path (solid) selection when using volcanic functions for a break of $\lambda = 6$. Consistent with derived theory (16), the estimator has 3.7 times the variance of the error term when using split-half estimation for the given function. Using a multi-path search reduces the variance drastically. Any selection bias of the multi-path search estimates can be controlled through bias correction after selection (see Castle *et al.*, 2011 and Pretis, 2015b). The supplementary material provides simulation results for a simple autoregressive DGP and model.

Comparison to Shrinkage-based Methods. Shrinkage-based methods using penalized likelihood estimation (Zou and Hastie, 2005; Tibshirani, 2011) provide an alternative to the general-to-specific algorithm used here in selecting models with more variables than observations. Figure 4 shows the simulation outcomes comparing multi-path indicator saturation (for $\alpha = 0.01$), the Lasso (Tibshirani, 1996, estimated using LARS, see Efron *et al.*, 2004) where cross-validation is used to determine the penalty and the Lasso where the penalty is set such to approximate the false-positive rate of the IS procedure under the null of no breaks ($\approx$0.01). The simulation uses a total break magnitude of six standard deviations (implying a trough of $3.48\sigma_\epsilon$) for an increasing number of evenly spaced breaks from 0 up to 10 in a sample of $T = 100$. The general-to-specific multi-path algorithm exhibits stable power exceeding that of the penalized likelihood methods across any number of breaks. The false-positive rate remains stable and close to the theory level of 0.01. The shrinkage-based procedures, due to their similarity to forward-selection, show decreasing potency as the number of breaks increases, and the false-positive rate is difficult to control.

© 2016 John Wiley & Sons Ltd

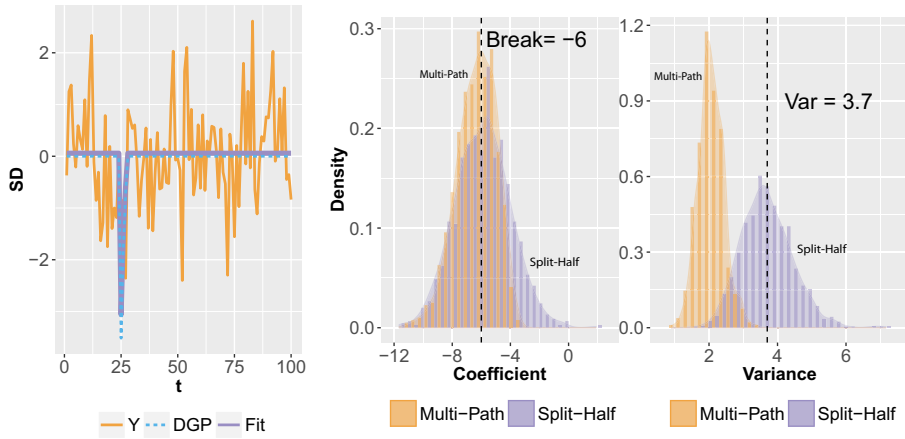

Figure 3. Estimated Break Indicator and Variance for Unknown Break Using Split-Half (purple/shaded dark, on right) and Multi-Path (orange/shaded light, on left) Selection.

Notes: The left panel shows a simulated time series with the true break shown as dotted and the fit as solid. The middle panel shows the distribution of the estimated coefficient, the right panel shows the variance of the coefficient. Vertical dashed lines show the true break magnitude and analytical variance of the split-half coefficient.

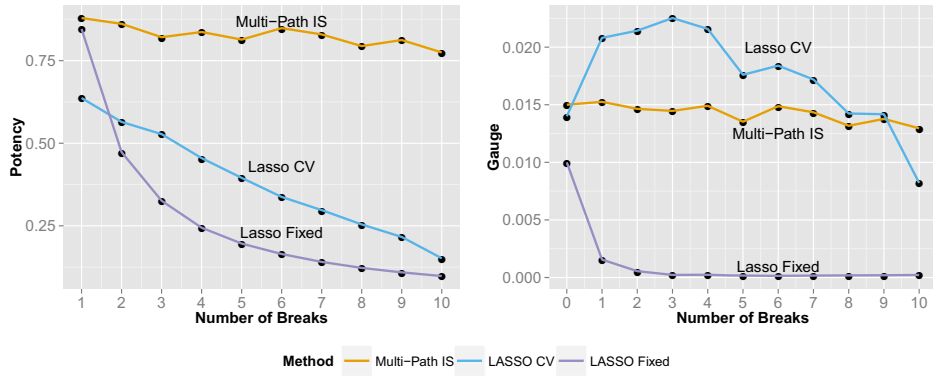

Figure 4. (Left) Average Potency of Detecting Increasing Numbers of Volcanic Breaks Using Multi-Path Indicator Saturation at $\alpha = 0.01$ (IS), Cross-Validated Lasso (CV) and Lasso with Fixed Penalty (Fixed) Where the Penalty is Set Such That the False-Positive Rate Approximates that of the Indicator Saturation Procedure under the Null of No Break; and (Right) Corresponding False-Positive Rate (Gauge).

Note: $M = 1000$ replications.

2.1.3 *Uncertainty on the Break Date*

An estimated uncertainty on the break magnitude and coefficient path (the time-varying intercept in the regression) can be computed given the distribution of the break estimator (see Pretis, 2015b). While of considerable interest, it is non-trivial, however, to quantify the uncertainty around the timing of the break (see Elliott and Müller, 2007). This is particularly true for the literature focusing on break detection using

© 2016 John Wiley & Sons Ltd

general-to-specific methodology. Here we investigate the uncertainty around the timing of estimated break points when using break-indicator saturation by computing the analytical power of a single break indicator when the break function is correctly specified but the break time is not. This is a simplification as it only considers a single mistimed indicator, while the indicator saturation approach includes a saturating set.

Consider a DGP with just a single break in the mean:

$$y_t = \lambda d_{T_1,t} + \epsilon_t \tag{18}$$

The break shifts $E[y_t]$ from 0 to λd_{T_1} at $t = T_1$ where d_{T_1} is a break function of length L beginning at time $t = T_1$ such that $T_1 + L < T$ and $d_{T_1} = (0, \ldots, d_1, d_2, \ldots, d_L, 0, \ldots, 0)$. The corresponding model is then

$$y_t = \gamma d_{j,t} + v_t \tag{19}$$

When the break date is correctly specified, $d_{j,t} = d_{T_1,t}$, so the estimator for λ is given by

$$\widehat{\gamma}_{t=T_1} - \lambda = \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)^{-1} \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}\epsilon_t\right) \tag{20}$$

Similarly for a test of the hypothesis: $\lambda = 0$, the t-statistic has a non-centrality of $E[t_{\widehat{\gamma},t=T_1}] = \psi = \frac{\lambda\sqrt{(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2)}}{\sigma_\epsilon}$ and the normal distribution

$$t_{\widehat{\gamma},t=T_1} \approx \frac{\widehat{\gamma}_{t=T_1}\sqrt{\left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)}}{\sigma_\epsilon} \sim N(\psi, 1) \tag{21}$$

The non-centrality ψ increases in the break magnitude λ, varies with the break length L, and will depend on the underlying break function given by d_t.

Now consider the model being incorrectly specified for the break date, such that $d_{j,t} \neq d_{T_1,t}$ but is shifted by K periods $d_{j,t} = d_{T_1\pm K,t}$. The estimator for λ is then

$$\widehat{\gamma}_{t=T_1\pm K} - \lambda = \lambda\left[\left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2\right)^{-1}\left(\sum_{t=T_1}^{T_1+L} d_{j,t}d_{T_1,t}\right) - 1\right] + \left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2\right)^{-1}\left(\sum_{t=T_1}^{T_1+L} d_{j,t}\epsilon_t\right) \tag{22}$$

For a fixed length L and a forced mistiming, it follows that $\widehat{\gamma}_{t\neq T_1}$ is not an unbiased estimator for λ. Note that if d_j is functionally specified correctly such that the only difference to the true break function is through K lags, $\mathbf{d}_j = \mathbf{d}_{T_1\pm K}$, then it holds that $(\sum_{t=T_1}^{T_1+L} d_{j,t}^2) = (\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2)$. Equally $(\sum_{t=T_1}^{T_1+L} d_{j,t}d_{T_1,t}) = (\sum_{t=T_1}^{T_1+L} d_{T_1\pm K,t}d_{T_1,t})$ for $K \leq L$ and 0 for $K > L$. Using this, we derive an expression for the approximate t-statistic associated with the estimator given a break function time misspecified by K lags:

$$E\left[t_{\widehat{\gamma},t=T_1\pm K}\right] \approx \frac{E\left[\widehat{\gamma}_{t=T_1\pm K}\right]}{\sigma_\epsilon\left(\sum_{t=T_1}^{T_1+L} d_{T_1}^2\right)^{-1/2}} = \frac{\lambda\left(\sum_{t=T_1}^{T_1+L} d_{T_1\pm K}d_{T_1}\right)}{\sigma_\epsilon\left(\sum_{t=T_1}^{T_1+L} d_{T_1}^2\right)^{1/2}} \tag{23}$$

This is equal to the non-centrality of the correct break date ψ scaled by a factor less than one, decreasing with the distance K from the correct date

$$E\left[t_{\widehat{\gamma},t=T_1\pm K}\right] \approx \psi\left(\frac{\sum_{t=T_1}^{T_1+L} d_{T_1\pm K,t}d_{T_1,t}}{\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2}\right) \leq \psi \tag{24}$$

For a given break specification d_t and break length L, the corresponding power function can be computed to provide an approximate measure of power for detection of a break at $t = T_1$ in the neighbourhood of

© 2016 John Wiley & Sons Ltd

T_1. Note that $E[t_{\widehat{\gamma},t=T_1\pm K}]$ is zero outside a neighbourhood of L. The associated t-statistic of a break indicator further away from the true break date T_1 than the break length L is zero in expectation, since $(\sum_{t=T_1}^{T_1+L} d_{j,t}d_{T_1,t}) = 0$ for $K > L$. Intuitively, longer breaks increase the likelihood that a break indicator that is not perfectly coincident with the break date will appear significant, and we can expect the retention to be equal to the nominal significance level outside a $t = T_1 \pm L$ interval.

As before we consider the special case of volcanic functions and also provide results from step shifts for comparison. Figure 5 shows the analytical as well as simulated non-centrality and power around a true break date at $t = 26$ of length $L = 3$ for $\alpha = 0.05$. The Monte Carlo simulations match the theoretical powers and non-centralities closely.

For no break, the analytical power is uniform and equal to the nominal significance level. When there is a break outside of the interval $T_1 \pm L$, the expected retention of the break indicator equals the nominal significance level. For a step shift of a forced length, given (24), the non-centrality decreases linearly as the numerator falls by $1/L$ per shifted period relative to the correct break date. For longer breaks this implies that the power around the true break date is close to uniform. In the case of volcanic functions, due to the particular functional form, the power and retention probability drop more rapidly and peak clearly around the true break date. The special case presented here only considers the properties of a single time-misspecified indicator of a fixed length in the model. However, model selection in the indicator saturation approach alleviates many of these concerns in practice. When selecting from a full set of break functions (see Section 2.1.2) it is less likely that a break function at $T_1 + -K$ appears significant because the correct T_1 indicator is included in the same model, a mistimed indicator in a fully saturated model would likely appear significant only if a chance draw of the error offsets the shift.

2.2 *Properties under the Null of No Break*

Under the null hypothesis when there are no breaks in the DGP, there are two primary concerns regarding the inclusion of a full set of break functions in the statistical model. First, when including a full set of break functions, break indicators may be retained spuriously, and secondly, there may be concerns about the effect on the distributions of coefficients on variables that are known to be relevant – in other words, does saturating a model with irrelevant variables affect relevant ones?

First, we consider the spurious retention of break indicators. Under the null of no breaks, $\lambda = 0$, the DGP from (8) is given by

$$\mathbf{y} = \boldsymbol{\epsilon} \tag{25}$$

Based on the above results, when using a split-half approach with a full set of break indicators, the expectation of the estimated coefficients in the first half is given by

$$E[\widehat{\gamma}_{(1)}] = E\left[\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\boldsymbol{\epsilon}\right] = \mathbf{0} \tag{26}$$

The same result generalizes to the union of retained indicators $\mathbf{D}_U$. Thus, the t-statistics of the included break functions will be centred around zero in expectation when there is no break. Using the selection rule that retains the break function $\mathbf{d}_j$ if $|t_{d_j}| > c_\alpha$, then $\alpha T/2$ indicators will be retained on average in each half. Combining the retained indicators in the final set, αT indicators are retained in expectation. The proportion of spurious indicators can thus be controlled through the nominal significance level of selection. The properties under the null are confirmed below using Monte-Carlo simulations.

Table 3 and Figure 6 report the simulation results when there are no breaks in the DGP but a full set of break functions (of the form of volcanic functions) is included. When using a split-half approach with a one-cut variable selection decision based on the absolute t-statistic, the proportion of irrelevant retained indicators is close to the nominal significance level. In practice, when using a multi-path, multi-split procedure (here implemented through *Autometrics*) the gauge is close to the nominal significance level

© 2016 John Wiley & Sons Ltd

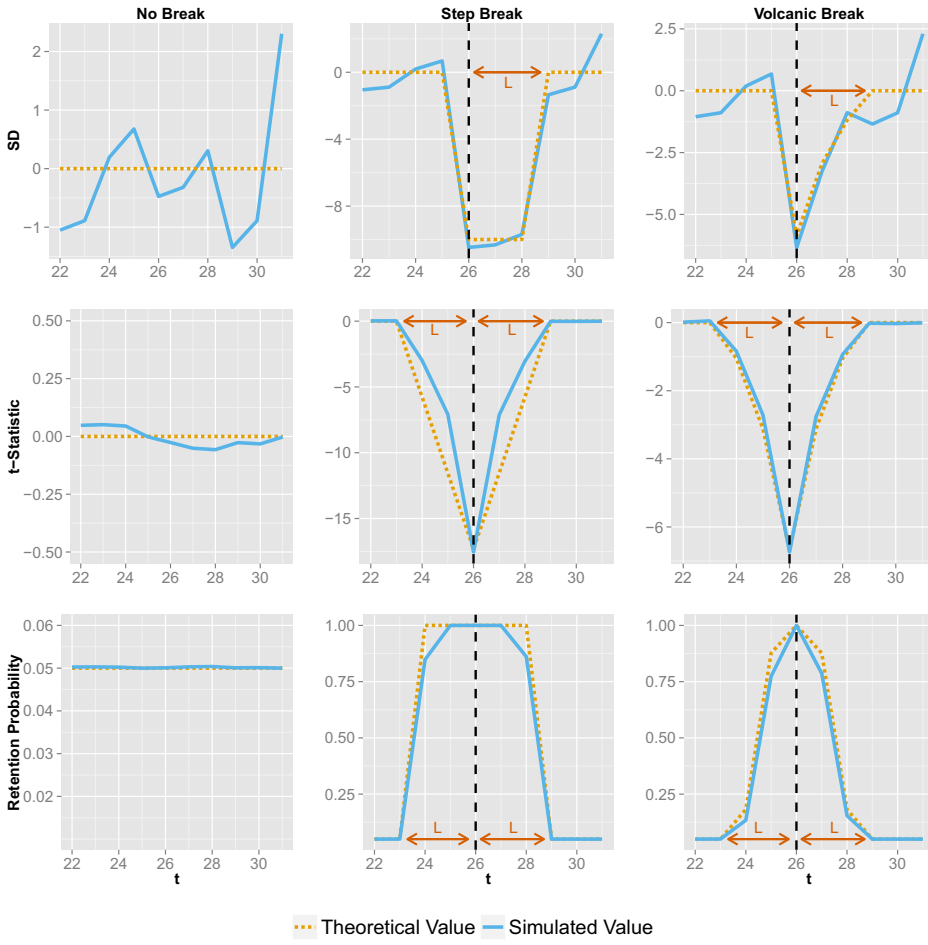

Figure 5. Power and Retention Frequency around the Break Date Where the Timing of the Break Functions is Imposed without Selection.

Notes: Simulated data with and without shifts (top), associated non-centrality and simulated t-statistics (middle), analytical and simulated power (bottom) around break $\lambda = -10$ at $T_1 = 26$ of length $L = 3$ and interval $T_1 \pm K$ for $\alpha = 0.05$. Left shows no break, middle a step-break and right panel a volcanic function break. Analytical non-centralities and powers are shown as dotted, simulated t-statistics and retention are shown as solid. Dashed lines mark the break occurrence. Outside of an interval $T_1 = 26 \pm L$ the retention probability and analytical power are equal to the nominal significance level of $\alpha = 0.05$.

for low levels of α. A conservative approach (low $\alpha \leq 1\%$) is recommended in practice.[12] When compared to results in Castle *et al.* (2015b), there is little notable difference between different specifications of break functions, consistent with the analytical properties of irrelevant indicators.

We now assess the second consideration, which is the effect of including a full set of break indicators when theory variables $\mathbf{X}$ are included in the model but are not selected over ('forced'). These could include contemporaneous covariates or autoregressive dynamic variables. For the specific case when the

© 2016 John Wiley & Sons Ltd

Table 3. Retention of Spurious Volcanic Break Functions When There is No Break.

Significance Level	Split-Half One-Cut		Multi-Path Search
	Gauge $\mathbf{D}_1$	Gauge $\mathbf{D}_2$	Gauge $\mathbf{D}$
$\alpha = 0.05$	0.056	0.054	0.30
$\alpha = 0.01$	0.013	0.012	0.015
$\alpha = 0.005$	0.007	0.007	0.005
$\alpha = 0.0025$	0.004	0.004	0.002
$\alpha = 0.001$	0.002	0.002	0.001

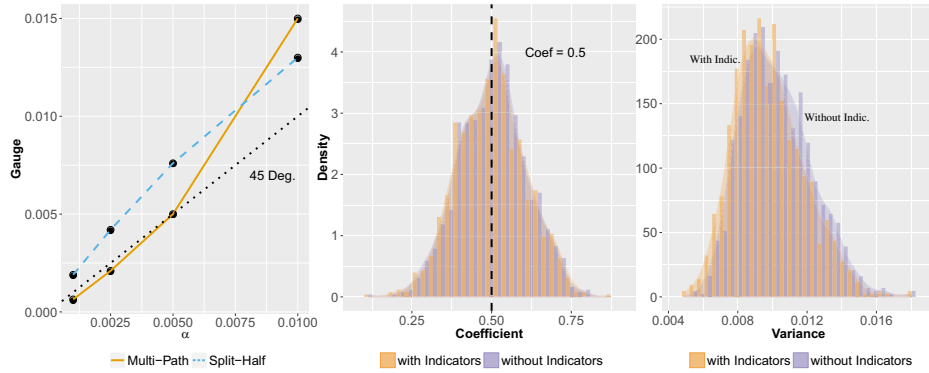

Figure 6. Simulation Results under the Null of No Break.

Notes: (Left) Proportion of irrelevant retained break functions (gauge) using split-half and multi-path selection for varying α when there is no break. (Middle and Right) Simulated distributions and densities of coefficient $\widehat{\beta}$ (true $\beta = 0.5$) on forced parameter x_t: with (orange/shaded light) – and without (purple/shaded dark) – a full set of break functions.

elements of $\mathbf{D}$ are specified to be impulse indicators, Johansen and Nielsen (2009) derive the asymptotic distribution of $\boldsymbol{\beta}$ in the full split-half approach in stationary and unit-root non-stationary regressions using the equivalence of IIS and one step Huber-skip M-estimators. For an iterated procedure (e.g. resembling the multi-block approach in *Autometrics*) the distributional results under the null for IIS are derived in Johansen and Nielsen (2013). For the general form of designed indicator functions, we follow theory for the substeps of split-half estimation where $N \ll T$ in each step, and appeal to simulation results for the overall algorithm. Consider a simple DGP:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{27}$$

where $\boldsymbol{\epsilon} \sim \text{iid}(0, \sigma_\epsilon^2 \mathbf{I})$ and the elements of $\mathbf{X}$ (dynamic or static) are assumed to be relevant and not selected over. The model relying on the split-half approach saturated with the first half of the break functions is then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}_1\boldsymbol{\gamma}_{(1)} + \mathbf{v} \tag{28}$$

where the true $\boldsymbol{\gamma}_{(1)} = \mathbf{0}$. Following Hendry and Johansen (2015), given that there is no break in the DGP, the inclusion of a full set of irrelevant additional variables $\mathbf{D}_1$ need not affect the distribution of the

© 2016 John Wiley & Sons Ltd

included relevant parameters $\boldsymbol{\beta}$. Orthogonalizing $\mathbf{X}$ and $\mathbf{D_1}$ by regressing each column of $\mathbf{D}_1$ on $\mathbf{X}$ yields the estimator $\widehat{\boldsymbol{\beta}}^*$ with asymptotic distribution:[13]

$$\sqrt{T}\begin{pmatrix}\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta} \\ \widehat{\boldsymbol{\gamma}}_{(1)} - \mathbf{0}\end{pmatrix} \xrightarrow{D} N\left[\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}, \sigma_\epsilon^2 \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{D_1 D_1|X}}^{-1}\end{pmatrix}\right] \tag{29}$$

The distribution of the parameters $\widehat{\boldsymbol{\beta}}^*$ on the correct variables $\mathbf{X}$ is unaffected by the inclusion of the orthogonalized break indicators $\mathbf{D}_1$ when there is no break. The equivalent result holds when the second half of break indicators $\mathbf{D}_2$ is included and the resulting union of retained indicators from $\mathbf{D}_1$ and $\mathbf{D}_2$ given that $N < T$. Orthogonalization relative to shifts, however, is not necessary for estimation in practice. Figure 6 shows the simulated distribution of $\widehat{\boldsymbol{\beta}}$ for a single x_t when a full set of break functions is included and selected at $\alpha = 0.005$ (orange/shaded light) and when break functions are not included (purple/shaded dark). The distribution of $\widehat{\boldsymbol{\beta}}$ is unaffected by the saturation of a full set of break functions. In practice, the main risk is the spurious retention of break indicators, but this can be controlled through a conservative selection mechanism (low α).

3. Empirical Illustration for Climate Time Series: Detection of Volcanic Eruptions from Simulated Model Surface Air Temperature Data

Large volcanic eruptions that inject significant amounts of sulphate aerosols into the stratosphere cause short-lived (multi-year) radiative imbalances that induce surface cooling. Over the course of the last several millennia there have been numerous eruptions that have had impacts on global mean temperatures. Identifying their climatic fingerprint is an important scientific endeavour that relies critically on the robust characterization of the timing and magnitude of past volcanism. An accurate understanding of the impact of past eruptions can lead to more accurate estimates of the effect of stratospheric aerosols – to guide policy from geo-engineering to pollution controls. Records of climatically relevant events primarily rely on sulphur deposits in ice cores (see, e.g. Gao *et al.*, 2008; Crowley and Unterman, 2012). However, there remains uncertainty in the precise timing, magnitude and climatic impact of past volcanic activity (Schmidt *et al.*, 2011; Anchukaitis *et al.*, 2012; Brohan *et al.*, 2012; Mann *et al.*, 2012; Baillie and McAneney, 2015). Statistical methods such as the break detection methodology presented herein can therefore augment previous volcanic reconstruction estimates by providing additional characterizations of the timing and magnitude of temperature responses to volcanic eruptions when coupled with large-scale proxy estimates of past temperature variability, for example, from tree-rings. As a synthetic evaluation of the performance of the break-indicator saturation method, we search for volcanic eruptions in surface air temperature output from model simulations. While there is some disagreement on the timing, magnitude and climatic impact of real eruptions over the past several millennia, the present simulation is forced with deterministic (known, imposed) eruptions. It therefore can function as a useful tool for assessing the detection efficacy of the proposed statistical methodology in real-world scenarios when the timing and exact DGP of volcanic eruptions are uncertain.

For our empirical illustration, we use the NH mean surface air temperature from the historical simulation of the National Center for Atmospheric Research (NCAR) Community Climate System Model 4 (CCSM4) and the Last Millennium (LM) simulation (Landrum *et al.*, 2013). These simulations were made available as part of the Coupled and Paleoclimate Model Intercomparison Projects Phases 5 and 3 (CMIP5/PMIP3), respectively (Taylor *et al.*, 2012). Collectively, the two simulations span the period 850–2005 C.E. To imitate potential proxy reconstructions (e.g. tree-ring based), temperatures for extratropical land areas ($30^\circ - 90^\circ$ N) were extracted from the model and only summer months (June–August) were used to build annual averages. This time period is expected to show the strongest cooling in response to an eruption (e.g. Zanchettin *et al.*, 2013 argue for a winter-warming effect) and is associated with the seasonal sampling window of many proxies such as dendroclimatic records. Temperatures are reported as anomalies relative

© 2016 John Wiley & Sons Ltd

to the 1850–1999 mean. The model is forced with the volcanic reconstruction by Gao *et al.* (2008) that reports volcanic activity as stratospheric sulphate loadings in teragrams (Tg). While the model is forced with multiple radiative forcing conditions (e.g. solar irradiance, greenhouse gases, volcanoes, land cover changes and anthropogenic aerosol changes), for the present experiments we treat these as unknown and work with the univariate NH mean temperature series, although multivariate models with more forcing variables could improve the detection algorithm. For a real-world scenario, however, estimates of climate-forcing and -sensitivity are uncertain (IPCC, 2013) and may prove to be of limited use in explaining non-volcanic temperature variation in proxy reconstructions, particularly in the presence of changes in measurement (see, e.g. Pretis and Hendry, 2013).

3.1 *Simulation Setup*

We design a break function to capture the temperature response to a large-scale volcanic eruption using a simple zero-dimensional energy balance model (EBM) that equates incoming to outgoing energy derived from simple physics-based models of climate (see, e.g. section 1 in Rypdal, 2012, section 1 in Schwartz, 2012 or Pretis, 2015a for linking system EBMs to econometric system models)

$$C\frac{dT'}{dt} = F - \theta T' \tag{30}$$

where θ is the climate feedback, C is the heat capacity, T' the temperature deviation from steady state (similar to the measured temperature anomaly as a departure from a long-term average) and F denotes radiative forcing (the variable that in our system describes the volcanic shock). The feedback response time of the model is given by $\tau = \frac{C}{\theta}$. Assuming a volcanic forcing effect of an impulse injection of stratospheric aerosols of F decaying exponentially at rate $-1/\gamma$ yields the following functional form of a volcanic function for the associated temperature response:[14]

$$T'_t = d_t = \begin{cases} \frac{1}{C} e^{\frac{-\theta}{C}t} F \left(\frac{\theta}{C} - \frac{1}{\gamma}\right)^{-1} \left[e^{t\left(\frac{\theta}{C} - \frac{1}{\gamma}\right)} - 1\right] & t \leq L \\ 0 & t < T_1, t > L \end{cases} \tag{31}$$

Intuitively, equation (31) states that a volcanic eruption through F leads to a sudden drop in temperatures, followed by a smooth reversion back to the original equilibrium. Different parameter calibrations are explored in the simulation section below. The main results are reported for a normalized temperature response where the feedback response time is set to 1, and the length of the volcanic impact is set to $L = 3$ to approximate the theory. The decay of stratospheric aerosols is modelled as $\gamma = 0.5$ (function *a*) and $\gamma = 3$ (function *b*) to capture one-period and two-period cooling, respectively. On visual inspection (see Figure 7) these calibrations closely match the average-model response based on a superposed epoch analysis of all large-scale volcanic eruptions in the climate model (Mass and Portman, 1989). The average model response in temperature is a drop by approximately 1–1.5 °C, followed by a smooth reversion to the previous mean over a 3–4 year period. While Gao *et al.* (2008) estimate the retention time for sulphate aerosols to be 2–3 years, a climatic perturbation of 4 years is in line with findings by Landrum *et al.* (2013). It is important to emphasize that the in-sample response to a volcanic eruption is not used to design the break function – the method is not trained and evaluated on the same set of observations.

In a more theoretical approach, which avoids particular shape parameters, a single peak (impulse) could be followed by autoregressive reversion to the mean where we search over a full set of impulses and full set of breaking autoregressive coefficients.

The DGP for the response variable NH temperature (T_t) is

$$T_t = f(X_t, V_t) + \epsilon_t \tag{32}$$

© 2016 John Wiley & Sons Ltd

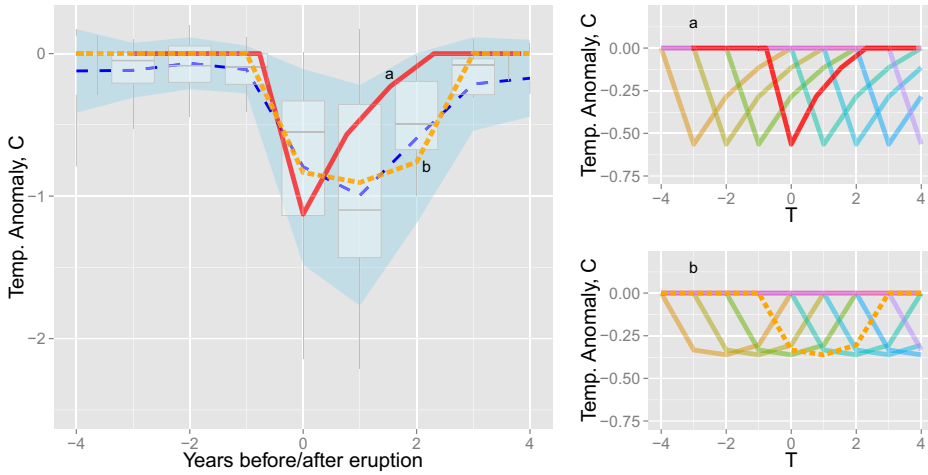

Figure 7. Superposed Epoch Analysis of the Model Temperature Response to Simulated Volcanic Eruptions and Sets of Volcanic Functions.

Notes: (Left) Superposed epoch analysis (Mass and Portman, 1989) of NH mean model temperature response to volcanoes with sulphate emissions >20 Tg (42 events, dashed) with 1 sample standard deviation bands (shaded) and distribution over volcanoes (box-plots). Approximate temperature response using a zero-dimensional energy balance model (EBM) used as volcanic function (a) is given as solid and function (b) in as dotted. (Right) Sets of EBM-based volcanic break functions for the two different specifications (a) (top) and (b) (bottom) to approximate the temperature response in years T relative to an eruption at $t = 0$.

To simulate sampling uncertainty of a proxy-based reconstruction, we generate 100 replications of the outcome by adding $\epsilon_t \sim \text{IN}(0, \sigma_\epsilon^2)$ to the NH mean temperature. The main results here are presented for simulations setting $\sigma_\epsilon = 0.2$ which is half the sample standard deviation of the NH time series of 0.4: the effect of the magnitude of noise is explored in Figure 9. The function $f(X_t, V_t)$ mapping volcanic, V_t, and other forcing, X_t, on to temperature is unknown and the observed forcing variables V_t and X_t are equally treated as unknown. As a proof of concept, we consider two models (intercept-only, and AR(1) with intercept)[15] to detect eruptions:

$$y_t = \mu + \gamma' \mathbf{d}_t + v_t \tag{33}$$

$$y_t = \rho y_{t-1} + \mu + \gamma' \mathbf{d}_t + v_t \tag{34}$$

where $\mathbf{d}_t$ is a full set of volcanic break functions (31) to be selected over.[16] To reduce computational requirements due to the varying simulation setup, the full-sample is split into 10 subsamples of $T = 115$ observations each.[17] There is little difference between full-sample and subsampling performance aside from computational speed (the supplementary material provides the results for a full-sample simulation). Selection is conducted at $\alpha = 0.01$ implying an expected gauge of 1% (approximately one break function spuriously retained per subsample). Higher retention of break functions can be an indicator of model misspecification. Simulations are evaluated based on the retention frequency of known individual volcanic events (potency), the average potency over all volcanoes and the proportion of spurious eruptions detected (gauge).

© 2016 John Wiley & Sons Ltd

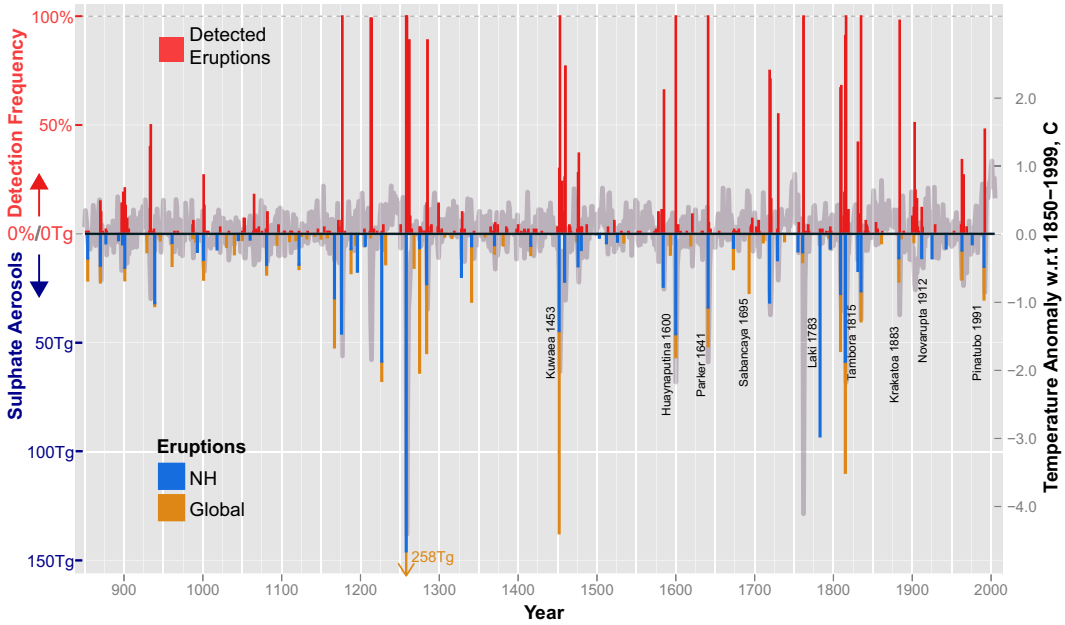

Figure 8. Detected Model Volcanic Eruptions from 850 to 2005.

Notes: Detected (top) volcanic eruptions in the model temperature series from 850 to 2005 using function (a) modelling a single-period drop followed by a reversion to the mean together with an intercept. Bar height indicates detection frequency [0, 100%] across 100 simulations. Stacked sulphur deposition record (bottom) used to force model temperatures are shown for Northern Hemisphere (blue/shaded dark) and global measurements (orange/shaded light) in Tg. Simulated model mean temperature anomalies used to detect the above volcanic eruptions are shown in grey. Mean NH surface temperature data are taken from the Last Millenium and historical simulation of the NCAR CCSM4 model as part of the CMIP5/PMIP3 data archive.

3.2 *Illustration Results*

Figure 8 and Tables 4 and 5 show the results of detected volcanic events in 100 replications of the modelled NH mean temperature[18] using the model (a) volcanic function. The retained volcanic breaks coincide predominantly with the simulated volcanic eruptions. Few spurious volcanoes are detected, and those that are spurious exhibit retention frequencies drastically lower than those of volcanoes used to force the model.

Most large-scale simulated volcanic eruptions are detected consistently: 74% of all larger (>20 Tg) NH eruptions are detected on average within an interval of ± 1 year (57% of all global eruptions, many of which appear to have had little impact on NH temperatures). Consistent with the basic analytical results presented in the previous section, the intervals of selection around the correct break dates are small. While increasing the band from 0 to 1 generally yields an increase in potency, outside of ± 1 year there is little difference (see Table 4). An uncertainty in break dates of ± 1 year can be the result of a monthly dated volcanic forcing record coupled with an annually dated temperature record, for example, a December eruption will mainly affect the following year. The season of sulphur injection – before or after summer – can cause offsets in the timing of the temperature response. Equally there may be regional sampling biases based on the construction of the NH mean surface air temperature.

© 2016 John Wiley & Sons Ltd

Table 4. Potency and Gauge for Volcanic Functions (a).

Function (a)	T	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
Potency NH Tg > 20	0.45	0.74	0.74	0.74
Potency NH Tg > 0	0.17	0.33	0.34	0.35
Potency Global Tg > 20	0.32	0.57	0.59	0.59
Potency Global Tg > 0	0.11	0.22	0.25	0.26
Gauge NH	0.02			
Gauge Global	0.02			
Function (a) + AR(1)				
Potency NH Tg > 20	0.46	0.70	0.70	0.70
Potency NH Tg > 0	0.16	0.30	0.31	0.31
Potency Global Tg > 20	0.31	0.52	0.54	0.54
Potency Global Tg > 0	0.11	0.20	0.22	0.23
Gauge NH	0.02			
Gauge Global	0.02			

Augmenting the designed break functions (a) by an autoregressive model results in nearly similar potency and gauge relative to the baseline model using just a constant (see Table 4 and Figure 9).

The retention frequency of volcanic functions increases with the magnitude of sulphate emissions of the volcanic eruption (Figure 9). While the overall potency for all volcanoes in the NH within a 1-year interval is 33%, this increases to 74% when larger volcanic eruptions over 20 Tg are considered. Given that potency covers all of the volcanic forcing, much of which is small in magnitude, the result is unsurprising. In particular, the lower potency for small eruptions is not driven by an inconsistency in selection of the same volcano over multiple experiments, but rather in the variation in temperature response between volcanoes. Eruptions in 1641 (Parker) and 1600 (Huaynaputina) are detected 100% of the time while the eruption of 1783 (Laki) is not detected in any of the outcomes. In contrast to most of the other volcanoes, Laki is a high-latitude volcano. Because the CCSM4 model uses spatially resolved sulphate estimates, this eruption only affects the northernmost areas and causes only a minor hemispheric cooling of -0.15°, which is much lower in magnitude than that of any of the other major volcanic events (see Figure 7).[19]

Equally, the potency is affected by the chosen standard deviation of the noise process added to the model mean. The main results here are reported for added noise with a standard deviation of half the sample standard deviation. Figure 9 shows the potency for varying levels of noise.

The proportion of spuriously detected volcanoes (gauge) at around 0.02 is close to the nominal significance level ($1/T \approx 0.01$). The fact that it is slightly higher is likely due to the misspecification of the model, which is only run on a constant (including an autoregressive term in the alternate specification) and set of break functions. Any variability in temperature other than volcanic eruptions may be spuriously attributed to the shape of the volcanic functions. This could be controlled by augmenting the model with additional dynamics (e.g. further autoregressive terms, long-term fluctuations through sine-cosine processes) or known forcing series.

Results for volcanic functions (b) are reported in the supplementary material. Volcanic functions (b) that capture the slower initial decline in temperature yield a slightly higher potency when measured at the precise timing (see Figure 9). Potency for $t = T_i$ for all i NH volcanoes using (b) is 0.32 versus 0.17 for (a) (0.23 vs. 0.11 for Global). This result stems from the single drop in function (a) often being most significant in the second period after an eruption if the cooling lasts for two periods. Once we consider the interval of $T_i \pm 1$ years or volcanoes of larger scale the results are nearly identical for functions (a) and (b). Differentiation between one or two-period cooling following an eruption, and

© 2016 John Wiley & Sons Ltd

Table 5. Potency of Detection of Volcanic Eruptions >20 Tg using Volcanic Functions (a) for Intervals $t = T \pm 1, 2, 3$

NH Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
939	31.83	0	0.02	0.02	0.03
1167	29.535	0	0	0	0
1176	45.761	0.06	1	1	1
1227	58.644	0.01	0.02	0.06	0.06
1258	145.8	1	1	1	1
1284	23.053	0.14	0.97	0.97	0.97
1452	44.6	0.3	1	1	1
1459	21.925	0.26	0.98	0.98	0.98
1584	24.228	0.11	0.77	0.8	0.8
1600	46.077	1	1	1	1
1641	33.805	1	1	1	1
1719	31.483	0.75	1	1	1
1783	92.964	0.02	0.02	0.03	0.05
1809	27.558	0.67	0.99	0.99	0.99
1815	58.694	0.91	1	1	1
1835	26.356	1	1	1	1
Global Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
854	21.387	0	0.02	0.03	0.03
870	22.276	0	0.25	0.25	0.25
901	21.283	0	0.34	0.5	0.54
939	33.128	0	0.02	0.02	0.03
1001	21.011	0	0.4	0.4	0.4
1167	52.114	0	0	0	0
1176	45.761	0.06	1	1	1
1227	67.522	0.01	0.02	0.06	0.06
1258	257.91	1	1	1	1
1275	63.723	0	0.06	0.08	0.08
1284	54.698	0.14	0.97	0.97	0.97
1341	31.136	0	0	0	0.01
1452	137.5	0.3	1	1	1
1459	21.925	0.26	0.98	0.98	0.98
1584	24.228	0.11	0.77	0.8	0.8
1600	56.591	1	1	1	1
1641	51.594	1	1	1	1
1693	27.098	0	0	0.03	0.07
1719	31.483	0.75	1	1	1
1783	92.964	0.02	0.02	0.03	0.05
1809	53.74	0.67	0.99	0.99	0.99
1815	109.72	0.91	1	1	1
1835	40.16	1	1	1	1
1883	21.864	0	0.98	0.98	0.98
1963	20.87	0	0.43	0.63	0.63
1991	30.094	0	0.48	0.48	0.48

© 2016 John Wiley & Sons Ltd

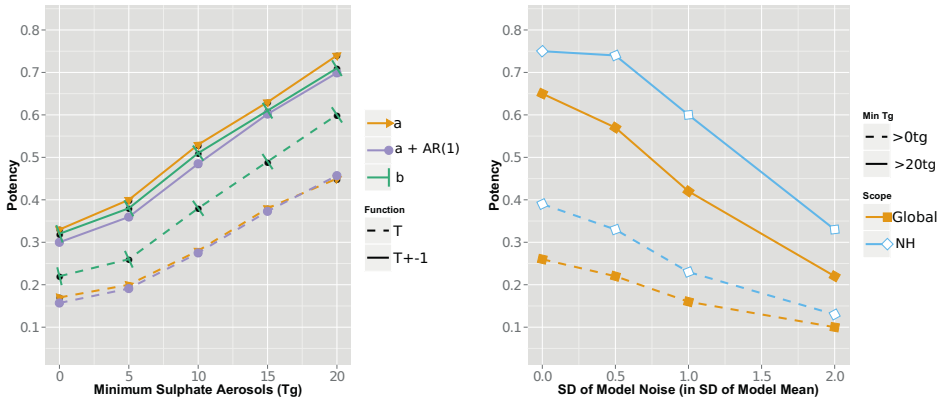

Figure 9. (Left) Detection Potency of NH Eruptions for Given Minimum Sulphate Emissions and Timing for Functions (a), (a) + AR(1) and (b) at the Precise Timing T (Dashed) and in the Interval of $T \pm 1$ (Solid); and (Right) Detection for Varying Levels of Noise Added in the Simulation for Function (a) for All Eruptions (Dashed) and Large Eruptions over 20 Tg (Solid).

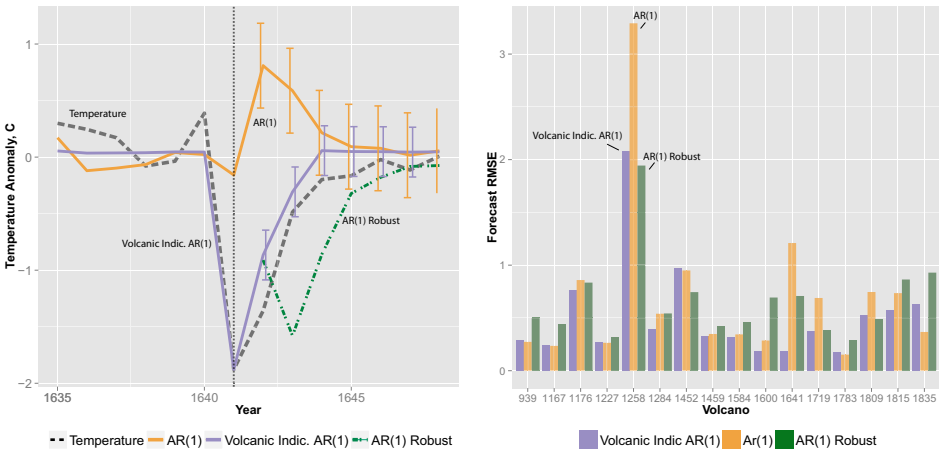

Figure 10. One-Step Forecasts through Volcanic Eruptions using Break Indicators.

Notes: (Left) Forecast performance across different methods: model mean temperature during the simulated 1641 eruption (dashed), one-step forecasts from 1641 onwards are shown for using an AR(1) model with volcanic indicator (purple/shaded dark), an AR(1) model without a volcanic indicator (orange/shaded light) and a robust AR(1) forecast (green/dot-dashed) (Clements and Hendry, 1999). Models are estimated from 1605 until 1641. (Right) one-step forecast root-mean-squared-error (RMSE) over all NH model volcanic eruptions (>20 Tg) for an AR(1) model with volcanic indicator (purple/left), without (orange/middle) and robust AR(1) forecast (green/right). Using volcanic indicators, on average, improves the forecast performance during the break period. However, when no break occurs (little to no temperature response), using a break indicator can result in higher RMSE as seen, for example, for the 1783 model Laki eruption.

© 2016 John Wiley & Sons Ltd

thereby further improvements in detection, could be implemented by searching over functions of type (a) and (b) simultaneously controlling the gauge appropriately.

In summary, large-scale volcanic eruptions can consistently be detected within a ± 1 year interval. Even though the model is likely misspecified when using only a constant, few spurious volcanic eruptions are retained. The signal-to-noise ratio remains, however, crucial in detection. When the method is applied to real-world proxy reconstructions where lower temperature spikes and higher noise levels can be expected, a well-specified baseline model for the temperature process will be required against which volcanic events can be detected to ensure a high power of detection.

3.2.1 *Forecasting during Breaks*

While breaks (such as volcanic eruptions) are by their nature stochastic, using a deterministic approach through a full set of break functions allows us to account for the underlying breaks and model the responses deterministically. This can improve forecasts during breaks if the break function is well specified. Once the break is observed (in this case a volcanic eruption), a forecasting model can be augmented with a break indicator where the magnitude is determined through estimation in the first break period. This indicator then acts as a continuous intercept correction, thereby improving the forecast performance during the break. To illustrate this concept, Figure 10 shows a 1-step forecast for NH model mean temperatures following the simulated 1641 eruption, together with the root-mean-squared (RMSE) forecast errors for all NH ($>$20 Tg) model eruptions based on volcanic function (a). Using volcanic indicators to forecast through the breaks yields on average a lower forecast RMSE (RMSE $= 0.51$) when compared to a simple AR(1) model (RMSE $= 0.71$) or even a robust forecasting device (RMSE $= 0.66$) (Clements and Hendry, 1999).[20] Crucially, this depends on the correct specification of the break function – for volcanic eruptions further improvements could be achieved by switching to volcanic function (b) if the initial cooling lasts for two periods. Detection of breaks based on theory-informed break functions can therefore act as a robust forecasting device through a continuous intercept correction from climate to economic time series.

4. Conclusion

Saturating a regression model with a full set of designed break functions, and removing all but significant ones through a general-to-specific algorithm yields unbiased estimates of the break magnitude and time. By initializing the model with a full set of break functions many of the shortcomings associated with a forward selection or specific-to-general approach in break detection can be avoided. Analytical properties and non-centralities can be derived for any deterministic break function and can be extended to breaks in random variables when interacted with the deterministic break specifications. The break detection procedure exhibits desirable properties both in the presence of breaks (stable potency across multiple breaks) and under the null hypothesis of no breaks where the spurious retention of break functions can be controlled through a chosen significance level of selection. The multi-path algorithm (*Autometrics*) outperforms shrinkage-based estimators, especially when facing multiple breaks. We provide some initial insight into uncertainty on the break date by assessing the retention probability of mistimed break estimators. Break-indicator saturation appears to be effective for detecting large-scale temperature responses to volcanic eruptions. This was shown using surface air temperature output from a combined LM and historical climate simulation. Statistically searching over a set of break functions consistently detects large eruptions from the simulated surface air temperatures without prior knowledge of their occurrence. This holds promise for future volcanic detection efforts using real-world proxy reconstructions of temperature variability over the last several millennia. More broadly, break detection using designed functions and indicator saturation provide a framework to analyse the detection of breaks of any designed shape at any point in time, with applications ranging from the detection of previously unknown events (such as shifts in time series due

© 2016 John Wiley & Sons Ltd

to measurement changes or policy impacts), to acting as a robust forecasting device during breaks – from economic recessions to volcanic eruptions.

Acknowledgments

We thank Vanessa Berenguer-Rico, Guillaume Chevillon, Niels Haldrup, Eric Hillebrand, Søren Johansen, Katarina Juselius, Oleg Kitov, John Muellbauer, Bent Nielsen, Max Roser, Timo Teräsvirta, and anonymous referees for helpful comments and suggestions. Financial support from the Open Society Foundations, the Oxford Martin School, the Robertson Foundation, and the British Academy, is gratefully acknowledged. LDEO contribution number 7983.

Notes

1. For example, SIS exhibits higher power in detecting step shifts than using impulses alone – see Castle *et al.* (2015b).
2. For k breaking variables, this implies augmenting the full-sample model by k $(T \times T)$ matrices.
3. While the framework presented here provides an encompassing specification for many break types, the construction of $\mathbf{D}$ is not limited to this particular case. Additional sets of specifications for step shifts are considered in Castle *et al.* (2015b). The appeal of the specification here is that the definition of $\mathbf{D}$ allows for a general framework under which properties can be analysed where many of the previously proposed cases are a special case of $\mathbf{D}$.
4. In a simple split-half analysis, there may be an identification problem if the sample-split coincides perfectly with a structural break. This is overcome by varying the block partitioning as is done in the software implementations of the algorithm.
5. See the supplementary material for proof.
6. Proof given in the supplementary material.
7. In practice, selection bias can be controlled using bias correction after orthogonalization of the selected regressors – see Hendry and Krolzig (2005) for the orthogonal case, Pretis (2015b) for bias correction of step functions and Castle *et al.* (2015a) for bias correction with correlated variables.
8. The split-half approach is not the only way of analysing the theory of indicator saturation: rather than splitting the functions into a first and second half, alternatively one could consider including every other break function in two sets such that $\mathbf{D}_1$ covers breaks at $t = 1, 3, 5 \ldots$ and $\mathbf{D}_2$ covers breaks at $t = 2, 4, 6 \ldots$. Retention frequencies in this setup can be derived using the results in Section 2.1.3.
9. While our analysis concentrates on small-sample properties, the asymptotic rates of convergence will generally depend on the specification of the break function – varying scaling to obtain non-degenerate limit distributions may therefore be required. In the case of step functions ($d_t = 1$, $L = T$) and the simple no-intercept case, pre-multiplying the estimator by $\sqrt{T}$ yields asymptotic normality for the break estimator when $T^{-1} \sum_{t=T_1}^{T_1+L-1} d_t^2 = T^{-1}L \to \tau$ as $T \to \infty$. In other words, the ratio of break length to the sample size remains constant as the sample size increases – this can be interpreted as obtaining more information on the break period or sampling at higher frequencies as $T \to \infty$. A similar analysis can be applied to the volcanic functions considered here, where either the break length scales with the sample size, or alternatively the magnitude increases similar to the asymptotic analysis for a single impulse in Doornik *et al.* (1998).
10. For a volcanic break, λ denotes the entire temperature response over the specified length L, thus the trough will be less than λ. For the present specification of $L = 3$, the initial trough of the function equals 0.58λ.
11. All simulations and applications using the multi-path search *Autometrics* are coded using the *Ox* programming language (Doornik, 2009b). Simulations using the Lasso are coded using the package `glmnet` (Friedman *et al.*, 2010) in *R*.

© 2016 John Wiley & Sons Ltd

12. Results of high gauge for high significance levels (e.g. $\alpha \geq 0.05$) are consistent with previous results found by Bergamelli and Urga (2013) for step functions. Once a large number of spurious breaks is retained, it becomes more likely to keep additional spurious breaks. The results for the gauge in Table 3 are consistent with the distributional theory for the gauge in Johansen and Nielsen (2016).
13. Where $T^{-1}\mathbf{X}'\mathbf{X} \xrightarrow{P} \Sigma_{XX}$ for stochastic $\mathbf{X}$, and $\mathbf{D}_1$ is scaled such that either the break length scales with the sample size, or alternatively the break magnitude increases such that $T^{-1}\mathbf{D}_1'\mathbf{D}_1 \to \Sigma_{\mathbf{D}_1}$ is constant, and for stochastic $\mathbf{X}$ it holds that $T^{-1}(\hat{\mathbf{u}}'\hat{\mathbf{u}}) \xrightarrow{P} \mathbf{\Sigma}_{\mathbf{D}_1\mathbf{D}_1|\mathbf{X}}$ where $\hat{\mathbf{u}} = \mathbf{D}_1 - \mathbf{X}\hat{\Gamma}$ from orthogonalization regressions. See the supplementary material for a proof based on Hendry and Johansen (2015).
14. See the supplementary material for a derivation. For break detection, the function is normalized to sum to 1 over L.
15. Unless otherwise stated, results refer to the intercept-only case. The intercept term and the autoregressive terms are not selected over.
16. Given the specification of the volcanic break function and if σ_ϵ was the only noise added to the DGP, then the approximate expected non-centrality for a single unknown break using (17) is $\lambda(0.2\sqrt{3.7}))^{-1} \approx 0.4^{-1}\lambda$ where λ is the full temperature response following a volcanic eruption. Since the specified volcanic function has an approximate trough of 0.58 λ, a temperature drop of 1° after a volcanic eruption implies that overall $\lambda \approx 1.7$. Thus in absence of additional noise and for a single volcanic break with an immediate temperature response of 1°, the expected t-statistic is approximately ≈ 4.3. The analytical probability of detecting this eruption is roughly: $P(|t| > c_\alpha) \approx 0.96$ for $\alpha = 0.01$. Large eruptions should be consistently detected if the break function is correctly specified and if σ_ϵ was the only source of noise.
17. The total sample size is $T = 1155$, resulting in nine subsamples of $T = 115$ observations and one subsample of $T = 120$ observations. Significance levels are scaled accordingly. Using a 3 GHz processor, the subsample approach requires ≈5 seconds to cover the entire sample for one replication (across 10 subsamples), compared to ≈5 minutes for one replication using a full-sample approach.
18. Retained volcanic functions with positive coefficients are dropped since these likely constitute positive outliers. The focus here lies on the detection of volcanic events which have a negative temperature response.
19. There is considerable uncertainty on the impact of the Laki eruption, for example, Schmidt *et al.* (2012) find the observed NH peak temperature response to Laki to be around -1°, suggesting that the LM simulation used here may not reflect the entire impact of the eruption, while D'Arrigo *et al.* (2011) argue that Winter impacts were likely independent of the Laki eruption. Notably, the eruption's noxious fumes at the time were discussed in White's (1789) treatment of phenology.
20. The robust forecasting device is based on first differences using the forecasting model for $T + 1|T$ given by: $y_{T+1|T} = y_T + \widehat{\rho}\Delta y_T$ where ρ is estimated using an AR(1) model. No error bars are shown on the robust forecast in Figure 10 (dot-dashed) due to the non-standard distribution of the forecast.

References

Anchukaitis, K.J., Breitenmoser, P., Briffa, K.R., Buchwal, A., Büntgen, U., Cook, E.R., D'Arrigo, R.D., Esper, J., Evans, M.N., Frank, D., Grudd, H., Gunnarson, B.E., Hughes, M.K., Kirdyanov, A.V., Körner, C., Krusic, P.J., Luckman, B., Melvin, T.M., Salzer, M.W., Shashkin, A.V., Timmreck, C., Vaganov, E.A. and Wilson, R.J.S. (2012) Tree rings and volcanic cooling. *Nature Geoscience* 5(12): 836–837.

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

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

Baillie, M. and McAneney, J. (2015) Tree ring effects and ice core acidities clarify the volcanic record of the first millennium. *Climate of the Past* 11(1): 105–114.

© 2016 John Wiley & Sons Ltd

Bergamelli, M. and Urga, G. (2013) Detecting multiple structural breaks: a Monte Carlo study and application to the Fisher equation for US. Discussion Paper, Cass Business School, London.
Brohan, P., Allan, R., Freeman, E., Wheeler, D., Wilkinson, C. and Williamson, F. (2012) Constraining the temperature history of the past millennium using early instrumental observations. *Climate of the Past Discussions* 8(3): 1653–1685.
Castle, J.L. and Shephard, N. (2009) *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.
Castle, J.L., Doornik, J.A. and Hendry, D.F. (2011) Evaluating automatic model selection. *Journal of Time Series Econometrics* 3(1). doi:10.2202/1941-1928.1097.
Castle, J.L., Doornik, J.A. and Hendry, D.F. (2015a) Bias correction after selection with correlated variables. University of Oxford Economics Discussion Paper.
Castle, J.L., Doornik, J.A., Hendry, D.F. and Pretis, F. (2015b) Detecting location shifts by step-indicator saturation during model selection. *Econometrics* 3: 240–264.
Clements, M.P. and Hendry, D.F. (1999) *Forecasting Non-Stationary Economic Time Series*. Cambridge, MA: MIT Press.
Cox, D.R. and Snell, E.J. (1974) The choice of variables in observational studies. *Applied Statistics* 23(1): 51–59.
Crowley, T.J. and Unterman, M.B. (2012) Technical details concerning development of a 1200-yr proxy index for global volcanism. *Earth System Science Data Discussions* 5(1): 1–28.
D'Arrigo, R., Seager, R., Smerdon, J.E., LeGrande, A.N., and Cook, E.R. (2011) The anomalous winter of 1783–1784: Was the Laki eruption or an analog of the 2009–2010 winter to blame? *Geophysical Research Letters*, 38. L05706, doi:10.1029/2011GL046696.
Doornik, J.A. (2009a) *Autometrics*. In J.L. Castle and N. Shephard (eds.), (pp. 88–121). Oxford: Oxford University Press.
Doornik, J.A. (2009b) *An Object-Oriented Matrix Programming Language Ox 6*. London: Timberlake Consultants Press.
Doornik, J.A. (2010) Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.
Doornik, J.A., Hendry, D.F. and Nielsen, B. (1998) Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys* 12: 533–572.
Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. (2004) Least angle regression. *Annals of Statistics* 32(2): 407–499.
Elliott, G. and Müller, U.K. (2007) Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141(2): 1196–1218.
Epprecht, C., Guegan, D. and Veiga, Á. (2013) Comparing variable selection techniques for linear regression: Lasso and autometrics. Documents de travail du Centre d'Economie de la Sorbonne 2013.80.
Ericsson, N.R. (2012) Detecting crises, jumps, and changes in regime. Working Paper, Board of Governors of the Federal Reserve System, Washington, DC.
Estrada, F., Perron, P. and Martínez-López, B. (2013) Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geoscience* 6: 1050–1055.
Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1): 1–22. Retrieved from http://www.jstatsoft.org/v33/i01/
Gao, C., Robock, A. and Ammann, C. (2008) Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. *Journal of Geophysical Research: Atmospheres* 113(D23).
González, A. and Teräsvirta, T. (2008) Modelling autoregressive processes with a shifting mean. *Studies in Nonlinear Dynamics & Econometrics* 12(1): 1–24.
Hendry, D.F. (1995) *Dynamic Econometrics*. Oxford: Oxford University Press.
Hendry, D.F. and Doornik, J.A. (2014) *Empirical Model Discovery and Theory Evaluation*. Cambridge MA: MIT Press.
Hendry, D.F. and Johansen, S. (2015) Model discovery and Trygve Haavelmo's legacy. *Econometric Theory* 31: 93–114.
Hendry, D.F. and Krolzig, H.M. (2005) The properties of automatic Gets modelling. *Economic Journal* 115: C32–C61.

© 2016 John Wiley & Sons Ltd

Hendry, D.F. and Pretis, F. (2013) Anthropogenic influences on atmospheric CO_2. In R. Fouquet (ed.), *Handbook on Energy and Climate Change* (pp. 287–326). Cheltenham: Edward Elgar.

Hendry, D.F., Johansen, S. and Santos, C. (2008) Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 23: 337–339.

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

Huang, J., Horowitz, J.L. and Ma, S. (2008) Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics* 36(2): 587–613.

IPCC (2013) *Fifth Assessment Report: Climate Change 2013: Working Group I Report: The Physical Science Basis*. Geneva: IPCC. Retrieved from https://www.ipcc.ch/report/ar5/wg1/. (accessed on February 2015).

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.) (pp. 1–36). Oxford: Oxford University Press.

Johansen, S. and Nielsen, B. (2013) Outlier detection in regression using an iterated one-step approximation to the Huber-skip estimator. *Econometrics* 1(1): 53–70.

Johansen, S. and Nielsen, B. (2016) Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics* 43(2): 321–348.

Kelly, P.M. and Sear, C.B. (1984) Climatic impact of explosive volcanic eruptions. *Nature* 311(5988): 740–743.

Kitov, O. and Tabor, M.N. (2015) Detecting structural breaks in linear models: a variable selection approach using multiplicative indicator saturation. *University of Oxford Economics Discussion Paper*.

Kock, A.B. and Teräsvirta, T. (2015) Forecasting macroeconomic variables using neural network models and three automated model selection techniques. *Econometric Reviews, Forthcoming.* doi:10.1080/07474938.2015.1035163.

Landrum, L., Otto-Bliesner, B.L., Wahl, E.R., Conley, A., Lawrence, P.J., Rosenbloom, N. and Teng, H. (2013) Last millennium climate and its variability in CCSM4. *Journal of Climate* 26(4): 1085–1111.

Mann, M.E., Fuentes, J.D. and Rutherford, S. (2012) Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. *Nature Geoscience* 5: 202–205.

Mass, C.F. and Portman, D.A. (1989) Major volcanic eruptions and climate: a critical evaluation. *Journal of Climate* 2(6): 566–593.

Perron, P. (2006) Dealing with structural breaks. In *Palgrave Handbook of Econometrics* (Vol. 1, pp. 278–352). London: MacMillan.

Perron, P. and Yabu, T. (2009) Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business & Economic Statistics* 27(3): 369–396.

Perron, P. and Zhu, X. (2005) Structural breaks with deterministic and stochastic trends. *Journal of Econometrics* 129(1): 65–119.

Pretis, F. (2015a) Econometric models of climate systems: the equivalence of two-component energy balance models and cointegrated VARs. *University of Oxford Economics Discussion Paper 750*.

Pretis, F. (2015b) Testing for time-varying predictive accuracy using bias-corrected indicator saturation. *University of Oxford Economics Discussion Paper*.

Pretis, F. and Allen, M. (2013) Climate science: breaks in trends. *Nature Geoscience* 6: 992–993.

Pretis, F. and Hendry, D. (2013) Some hazards in econometric modelling of climate change. *Earth System Dynamics* 4(2): 375–384.

Pretis, F., Mann, M.L. and Kaufmann, R.K. (2015a) Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *Climatic Change* 131(4): 705–718.

Pretis, F., Sucarrat, G. and Reade, J. (2016) General-to-specific modelling and indicator saturation with the R package gets. University of Oxford Economics Discussion Paper 794.

Rypdal, K. (2012) Global temperature response to radiative forcing: solar cycle versus volcanic eruptions. *Journal of Geophysical Research: Atmospheres* 117(D6).

Schmidt, G.A., Jungclaus, J.H., Ammann, C.M., Bard, E., Braconnot, P., Crowley, T., Delaygue, G., Joos, F., Krivova, N.A., Muscheler, R., Otto-Bliesner, B.L., Pongratz, J., Shindell, D.T., Solanki, S.K., Steinhilber, F., and Vieira, L.E.A. (2011) Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geoscientific Model Development* 4(1).

© 2016 John Wiley & Sons Ltd

Schmidt, A., Thordarson, T., Oman, L.D., Robock, A. and Self, S. (2012) Climatic impact of the long-lasting 1783 Lakieruption: inapplicability of mass-independent sulfur isotopic composition measurements. *Journal of Geophysical Research: Atmospheres* 117: D23(16), doi:10.1029/2012JD018414.

Schwartz, S.E. (2012) Determination of Earth's transient and equilibrium climate sensitivities from observations over the twentieth century: strong dependence on assumed forcing. *Surveys in Geophysics* 33(3-4): 745–777.

Strikholm, B. (2006) Determining the number of breaks in a piecewise linear regression model (Tech. Rep.). SSE/EFI Working Paper Series in Economics and Finance.

Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* 93(4): 485–498.

Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)* 58(1): 267–288.

Tibshirani, R. (2011) Regression shrinkage and selection via the Lasso: a retrospective. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 73(3): 273–282.

White, G. (1789) *The Natural History of Selborne*. Oxford: Reprint by Anne Secord, Oxford University Press, 2013.

White, H. (2006) Approximate nonlinear forecasting methods. In G. Elliot, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting* (pp. 459–512). Amsterdam: Elsevier.

Zanchettin, D., Timmreck, C., Bothe, O., Lorenz, S.J., Hegerl, G., Graf, H.-F., Luterbacher, J., Jungclaus, J.H. (2013) Delayed winter warming: a robust decadal response to strong tropical volcanic eruptions? *Geophysical Research Letters* 40(1): 204–209.

Zou, H. and Hastie, T. (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 67(2): 301–320.

Supporting Information

Additional Supporting information may be found in the online version of this article at the publisher's website:

Supplementary Material

© 2016 John Wiley & Sons Ltd