Measuring the degree of non-stationarity of a time series

Sourav Das* and Guy P. Nason

In time series analysis, there is an extensive literature on hypothesis tests that attempt to distinguish a stationary time series from a non-stationary one. However, the binary distinction provided by a hypothesis test can be somewhat blunt when trying to determine the degree of non-stationarity of a time series. This article creates an index that estimates a degree of non-stationarity. This index might be used, for example, to classify or discriminate between series. Our index is based on measuring the roughness of a statistic estimated from the time series, which is calculated from the roughness penalty associated with a spline smoothing/penalized least-squares method. We further use a resampling technique to obtain a likely range of index values obtained from a single realization of a time series. We apply our method to ascertain and compare the non-stationarity index of the well-known earthquake and explosion data. © 2016 The Authors. *Stat* Published by John Wiley & Sons Ltd.

Keywords: bootstrap assessment; non-parametric regression; time series

1 Introduction

Recently, there has been intense interest in tests of second-order stationarity. See, for example, Priestley & Subba Rao (1969), Von Sachs & Neumann (2000), Paparoditis (2009, 2010), Cardinali & Nason (2010), Dette et al. (2011), Dwivedi & Subba Rao (2011), Nason (2013), Taylor et al. (2014), Bašta (2015), Barlett et al. (2015), Jentsch & Subba Rao (2015), Jin et al. (2015), Bandyopadhyay & Subba Rao (2016) and Cardinali & Nason (2016). Most, if not all, of these articles consider the problem of testing the null hypothesis that a time series, $\{X_t\}$, is second-order stationary against the alternative hypothesis that it is not. In practice, such a null/alternative assessment is somewhat blunt. This article takes a different approach: one where we assign an index, $S(X_t) = S(\{X_t\})$, to a time series, $\{X_t\}$, that measures the degree of non-stationarity of $\{X_t\}$. The key advantage of a non-stationarity index is that, with two time series, $\{X_t\}$ and $\{Y_t\}$, one can attempt to ascertain whether series $\{X_t\}$ is more non-stationary than series $\{Y_t\}$.

Section 2 develops an experimental empirical measure of non-stationarity based on the mathematical roughness of the time evolution of fitted parameters of a dynamic linear model. For stationary series, the parameters are constant, which results in zero roughness and a zero value of S. For non-stationary series, greater parameter wiggliness (over time) implies greater roughness and larger positive values of S. Section 3 shows some examples of the measure S for different processes.

We can compute $S(X_t)$ based on one realization of $X_t, t = 1, \ldots, T$. However, we are more interested in S with respect to the stochastic process $\{X_t\}$, and hence, Section 4 proposes a bootstrap technique that obtains a likely distributional assessment of the quantity $S(\{X_t\})$ computed on the process. Such a distribution can be particularly useful when

School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK
*Email: sd15021@bristol.ac.uk

comparing two, or more, series to see whether one is really more non-stationary than the other or whether their distributions are not distinguishable.

Although many of the articles listed earlier focus on hypothesis tests, many of them use metrics that could usefully be employed to measure non-stationarity as we do here, for example, the L_2 metric from Dette et al. (2011). Although we develop a sensible index of stationarity here, it is by no means the only one that could be constructed. Indeed, it would be possible to define measures of non-stationarity that measure different aspects that might, in some cases, give contradictory answers.

2 Non-stationarity as roughness

To assess second-order stationarity, we require some way of modelling at least the second-order structure of our time series. *One* possibility, engaged here, is the use of an underpinning Bayesian dynamic linear model that can flexibly adapt to the underlying process structure, which we review next. The underlying idea is not restricted to dynamic linear models but could also be applied to other moments; local wavelet, $S_j(z)$; or Fourier spectra, $f(z, \omega)$, as functions of time z; see Dahlhaus (1997) or Nason et al. (2000) for example.

Let $y_1, y_2, ..., y_n$ be a realization observed from a Gaussian time-varying autoregressive (TVAR) process $\{Y_t\}$ of order 1, TVAR(1), with an unobservable autoregressive parameter θ_t. Dynamic linear models (e.g. West & Harrison, 1997, Chapter 4) provide a general extendable representation for such series. A TVAR(1) Gaussian dynamic linear model is characterized by two coupled stochastic processes,

$$\begin{aligned} Y_t &= \theta_t Y_{t-1} + v_t; \quad v_t \sim N(0, V) \text{ and} \\ \theta_t &= G_t \theta_{t-1} + w_t; \quad w_t \sim N(0, W), \end{aligned} \quad (1)$$

which are the observation and system equations, respectively, with respective innovation variances of V and W, which we assume to be constant. The quantities G_t, V and W are assumed known (or at least well modelled) and deterministic and can, in practice, be selected using the usual dynamic linear modelling procedures. Let $\mathfrak{D}_t$ denote the usual information set at time t. The system equation in model (1) can represent any of the following: a prior belief, a latent process or an unobservable signal at time t, θ_t, that a domain expert might be interested to estimate. Page 106 of West & Harrison (1997) explains how prediction in this set-up proceeds in a recursive manner and that the posterior distribution $\theta_t \mid \mathfrak{D}_t$ can be shown to follow an $N(m_t, C_t)$ distribution.

For our purposes of measuring stationarity, the key modelling issue is whether θ_t is constant over time. Indeed, $\{Y_t\}$ is second-order stationary if and only if $\{\theta_t\}$ is constant, and this occurs if and only if $W = 0$ and $\theta_t = G_t = c$ for some constant c. For Gaussian dynamic linear models, the necessary information is completely specified by m_t and C_t (as the current prior becomes the next posterior), and naturally, these two are constant for stationary $\{Y_t\}$. To *measure* the degree of non-stationarity, we need some metric for assessing increasing "non-constancy," and one possibility is to measure the "wiggliness" or roughness of the trajectory of m_t. In this article, we study the roughness of the trajectory of $\hat{\theta}_t$, the estimated θ_t.

To assess the roughness of θ_t, we fit a non-parametric function $g(t)$ to the estimated mean response curve $\hat{\theta}_t$ by minimizing the following penalized least-squares criterion:

$$Q = \sum_{i=1}^{n} \{m_{t_i} - g(t_i)\}^2 + \lambda \int_{t_1}^{t_n} \{g''(t)\}^2 \, dt, \quad (2)$$

where $\lambda > 0$ is selected using the generalized cross-validation methodology proposed by Golub et al. (1979). In minimizing Q, we not only obtain a good estimator $g(t)$ of m_t, but, crucially, we compute the roughness measure

$\int_{t_1}^{t_n} \{g''(t)\}^2 \, dt$, which is an estimate of the roughness of θ_t. The integral of the squared second derivative is not the only roughness measure: an extensive literature on global and local roughness measures exists; see for example Good & Gaskins (1971, Section 5), Green & Silverman (1993, Chapter 1) or Ramsay & Silverman (2010, Section 5.2.2).

We have two primary motivations for choosing $\int_{t_1}^{t_n} \{g''(t)\}^2 \, dt$ as the roughness measure. First, the measure can be applied over any time interval and, more importantly, over compactly supported non-overlapping intervals. Second, the theory underlying Q is mature and well understood and is linked to the theory of natural cubic splines. That is, the solution $g(t)$, which minimizes criterion (2), among the class of all continuous functions with continuous first-order derivatives, is a natural cubic spline possessing the following representation (e.g. Silverman, 1985; Green & Silverman, 1993; Wood, 2006):

$$g(t_i) = \sum_{j=1}^{n} \gamma_j \beta_j(t_i); i = 1, 2, \ldots, n; t_1 < t_2 < \ldots < t_n, \tag{3}$$

where the $\beta_j(t)$ are cubic polynomial spline basis functions and the coefficients γ_j are estimated by minimizing Eq. (2). The $t_1 < t_2 < \ldots < t_n$ are the knots where $g(t)$ is continuous and has continuous second-order derivatives at the knot points.

However, our primary interest is not in the fitted g but in the second term of (2), which measures its roughness. Because θ_t is random, it is rational not to interpolate its estimate but find a solution g_t that fits best within the class of all smooth functions. Hence, the second term is a good surrogate estimator for the roughness for m_t. We can give a more parsimonious matrix expression for the roughness measure. Define

$$\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \ldots, \gamma_n\}, B_{ij} = \beta_j(t_i), \Omega_{ij} = \int_{-\infty}^{\infty} \beta_i''(t)\beta_j''(t) \, dt, \tag{4}$$

where $\boldsymbol{\gamma}(Y_t)$ are the γ that arise from fitting the smoothing spline to $\{Y_t\}$. Further, define

$$S(Y_t) = \int_{t_1}^{t_n} \{g''(t)\}^2 \, dt = \boldsymbol{\gamma}^T(Y_t) \, \Omega \, \boldsymbol{\gamma}(Y_t) \tag{5}$$

to be the *non-stationarity index* of the time series $\{Y_t\}$.

3 Some expository examples

This section contains two sets of simple expository examples showing how $S(Y_t)$ changes as $\{Y_t\}$ moves from a stationary model to increasingly non-stationary TVAR(1) models. Smoothing splines were fitted using the routine `smooth.spline`, and the non-stationarity index $S(Y_t)$ was computed using the function `bsplinepen` of the package `fda` (Ramsey & Wickham, 2014), all within the R statistical package.

Figure 1 shows the value of S for four models. In each case, G_t is constant over time. The plots show the result of computing $S(Y_t)$ on different realizations, where one parameter is changed but the others are held constant. So, Figure 1(a) has $W = 0.05$, $V = 1$ and $G_t \in [0.2, 1]$, but all are fixed for each realization. Figure 1(b) has $G_t = 0.8$ and $V = 1$ but changes $W \in [0.05, 1]$. From both of these plots, one can see that increasing G_t or increasing W results in increased roughness of the mean path of θ_t, increasing non-stationarity and increasing $S(Y_t)$. The bottom row shows constant $S(Y_t)$ for two different stationary processes.

The second set of plots shows the spline fitting to a realization from five different models. Model a specifies $\theta_t = 0.8$, model b is $\theta_t = 0.8\theta_{t-1} + w_t$, model c is $\theta_t = \sin(2\pi\omega_0 t) + \cos(2\pi\omega_0 t) + w_t$ with $\omega_0 = 50/1000$, model d is $\theta_t = \exp(-t/T) + w_t$ and model e is $\theta_t = \exp(-t^2/T^2) + w_t$; and, in all cases, $W = 0.03$.
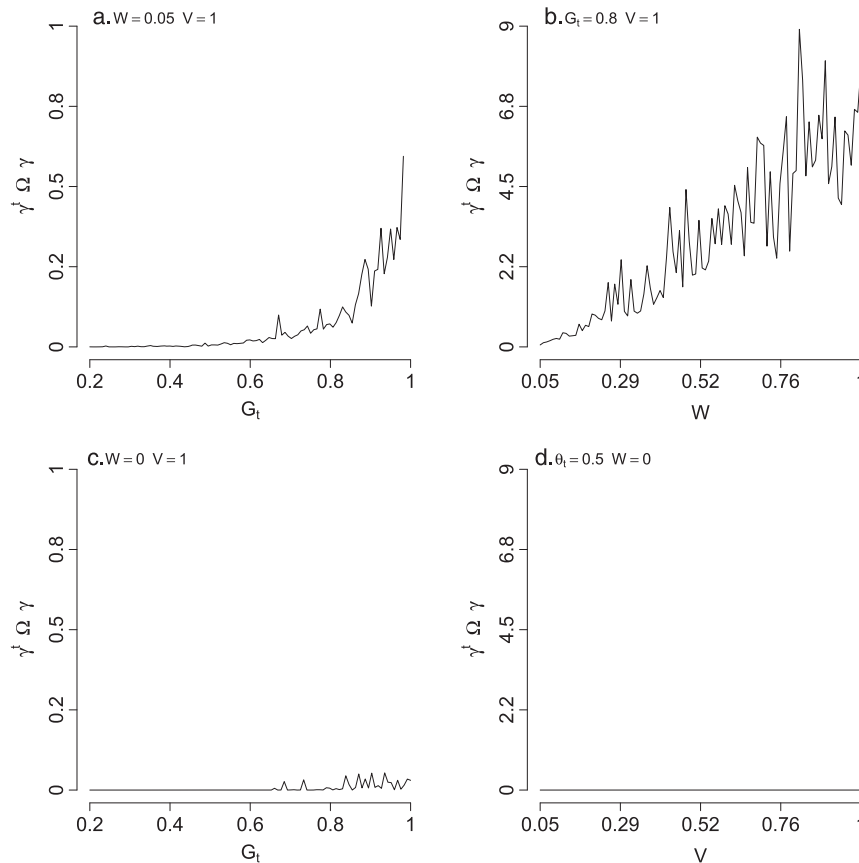
Figure 1. Panels demonstrate the variation in the roughness criterion $S(\cdot)$ as functions of model parameters, G_t and W. In these illustrations, G_t does not depend on time. We depict four scenarios assuming a dynamic linear model (1) generating mechanism. (a) TVAR(1) process with $W = 0.05$, $V = 1$ and $G_t \in [0.2, 1]$. (b) TVAR(1) process with $G_t = 0.8$, $V = 1$ and $W \in [0.05, 1]$. (c) Stationary AR(1) process with $W = 0$, $V = 1$ and $\theta_t \in [0.2, 1]$. (d) Stationary AR(1) process with process variance $V \in [0.05, 1]$.

Figure 2 shows the results with the computed value of $S(Y_t)$ in the top left of each plot. It is easy to see that stationary processes result in a flat line and increasingly non-stationary series give a larger value of S. The S values in Figure 2 are for one realization. We repeated the process for 5000 independent realizations: the summary statistics for each model are shown in Table I. So, for example, both plot and table show that model c exhibits the highest degree of non-stationarity, whereas model a is stationary but, owing to sampling fluctuations, is not precisely zero.

Model c highlights that the non-stationarity index S is an additive measure of the total changes in the gradient of the parameter under study. It cannot distinguish between deterministic and stochastic non-stationarity.

4 Bootstrap distribution of the non-stationarity index

Computation of $S(Y_t)$ on one realization provides limited assessment of the non-stationarity of the underlying process. For a fuller assessment, we construct a conditional bootstrap distribution for the non-stationarity index. For two time series, $\{X_t\}$ and $\{Y_t\}$, a bootstrap estimated distribution of the corresponding non-stationarity indices $S(X_t)$ and $S(Y_t)$ can be plotted to give an idea of whether one *process* is more non-stationary than the other. In recent years, several
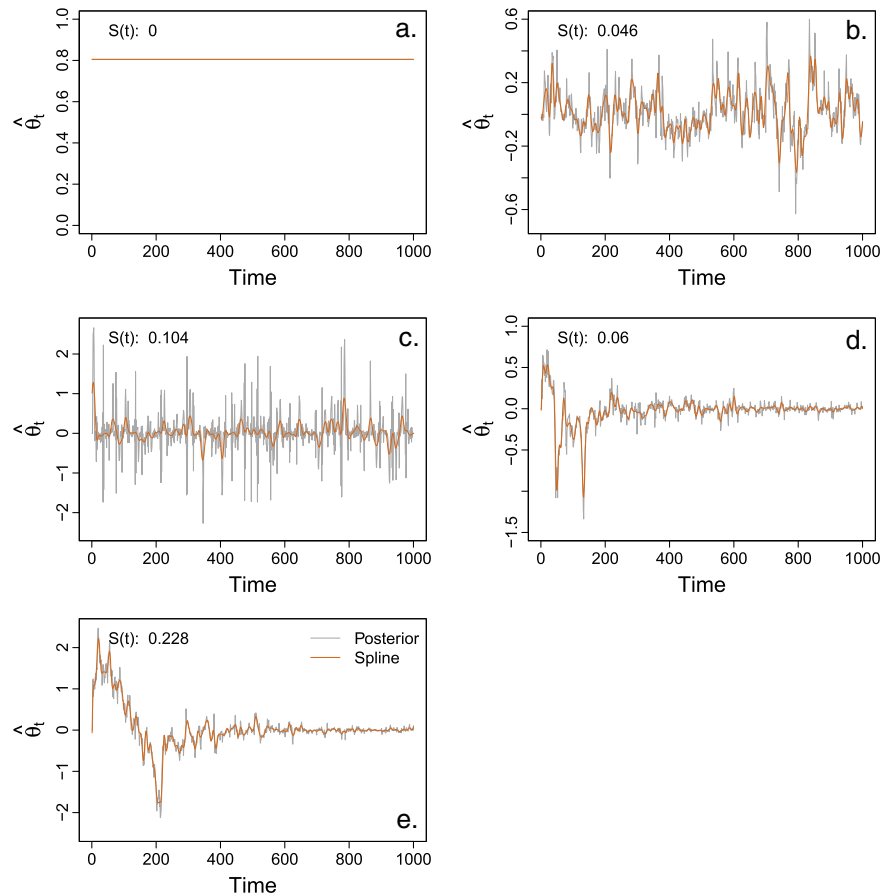
Figure 2. Spline fit (orange) superimposed on a realization (in grey) for estimated posterior mean m_t of the dynamic linear model (1) for different time-varying functions, G_t. Non-stationary index, S, shown in the top left of each panel. Models a to e as defined in the text.

Table I. Summary statistics of S over 5000 realizations.						
Model	Minimum	First quartile	Median	Mean	Third quartile	Maximum	Variance
b	0.01	0.03	0.03	0.04	0.04	0.16	0.00
c	0.00	0.09	0.31	1.10	2.15	5.65	1.50
d	0.01	0.04	0.08	0.12	0.16	0.82	0.01
e	0.02	0.08	0.12	0.16	0.20	0.99	0.01

Note: For model a, $\theta_t = 0.8$, all entries are numerically indistinguishable from zero.

resampling methods for locally stationary processes have been developed. See, for example, Sergides & Paparoditis (2008) and Kreiss & Paparoditis (2012, 2015). Our proposed resampling procedure generates bootstrap innovations corresponding to a fairly general class of TVAR processes. Because our generating model is a Gaussian dynamic linear model, we adapt the state space bootstrap algorithm of Stoffer & Wall (1991), which works by joint estimation and resampling from the system and observation residuals, as described next.

4.1 Assessing distribution of S using a conditional bootstrap algorithm

For the dynamic linear model (1), given $\theta_t \mid \mathfrak{D}_t$, Y_t is independent of its past and future. This implies that to resample from the forecast distribution $Y_t \mid \mathfrak{D}_{t-1}$, at time t, we only need to resample from the conditional distribution $Y_t \mid \theta_t, \mathfrak{D}_{t-1}$. The resulting algorithm is a modified residual bootstrap that uses the forecast distribution $Y_t \mid \mathfrak{D}_{t-1}$ and the parameter distribution $\theta_t \mid \mathfrak{D}_{t-1}$ with information until time $t-1$, to generate bootstrap sample Y_t^* at time t.

(1) Our bootstrap is based on model (1). Then the state at time $t-1$ can be derived from (1) by

$$\begin{pmatrix} \theta_t \\ Y_t \end{pmatrix} \mid \mathfrak{D}_{t-1} \sim N(\boldsymbol{\mu}_t, \Sigma_t), \tag{6}$$

where

$$\boldsymbol{\mu}_t = \begin{pmatrix} a_t \\ f_t \end{pmatrix} \text{ and } \Sigma_t = \begin{bmatrix} R_t & A_t Q_t \\ A_t Q_t & Q_t \end{bmatrix}, \tag{7}$$

where a_t and f_t are the prior mean and forecast mean for θ_t and Y_t, respectively; R_t is the prior variance of θ_t; Q_t is the (prior) forecast variance of Y_t; and A_t is the regression coefficient of θ_t on Y_t (these are standard dynamic linear model quantities used repeatedly by West & Harrison, 1997).

(2) Generate bootstrap residuals at time t using

$$e_t = Q_t^{-1/2} v_t \text{ and } v_t = Y_t - a_t Y_{t-1}. \tag{8}$$

(3) Resample N times from e_t with replacement obtaining $e_t^{(i)}$ for $i = 1, \ldots, N$. Construct N bootstrap samples by

$$\begin{pmatrix} \theta_t^{(i)} \\ Y_t^{(i)} \end{pmatrix} = \hat{\boldsymbol{\mu}}_t + F_t \mathbf{e}_t^{(i)}, \; \boldsymbol{\mu}_t = \begin{pmatrix} G_t \theta_{t-1}^{(i)} \\ a_t Y_{t-1}^{(i)} \end{pmatrix} \text{ and } F_t = \begin{pmatrix} R_t Y_{t-1}^{(i)} Q_t^{1/2} \\ Q_t^{1/2} \end{pmatrix}, i = 1, \ldots, N. \tag{9}$$

(4) Repeat parameter estimation for each bootstrap sample to obtain $\{S^{(i)}\}_{i=1}^N$ using (5).

To assess bootstrap parameter estimation errors, we define two mean-squared loss functions about θ_t and m_t.

$$L_{\text{tru}} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \left\{ m_t^{(i)} - \theta_t \right\}^2 \text{ and } L_{\text{est}} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \left\{ m_t^{(i)} - m_t \right\}^2, \tag{10}$$

where $\{m_t^{(i)}\}_{t=1}^T$ are the bootstrap estimated mean trajectories for bootstrap sample $i = 1, \ldots, N$. We carried out a bootstrap evaluation of S for each of the examples in Table I for $N = 5000$ and $T = 1000$, and the summary statistics are shown in Table II and are broadly comparable with the values in the previous table. Bootstrap histograms of the non-stationarity index S corresponding to the five different TVAR(1) models are shown in Figure 3. The top left

Table II. Summary statistics of bootstrap evaluation of S for system and observation variances of $W = 0.03$ and $V = 1$.

| Model	Minimum	First quartile	Median	Mean	Third quartile	Maximum	Variance	L_{tru}	L_{est}
a	0	0	0	0	0	0	0	3.71×10^{-4}	3.82×10^{-4}
b	0.01	0.02	0.02	0.02	0.03	0.13	0	0.1	0.02
c	0	0.01	0.10	0.54	0.99	8.59	0.84	0.53	0.19
d	0	0.03	0.05	0.10	0.14	0.73	0.01	0.07	0.02
e	0	0.02	0.07	0.09	0.11	1.30	0.01	1.75	1.67
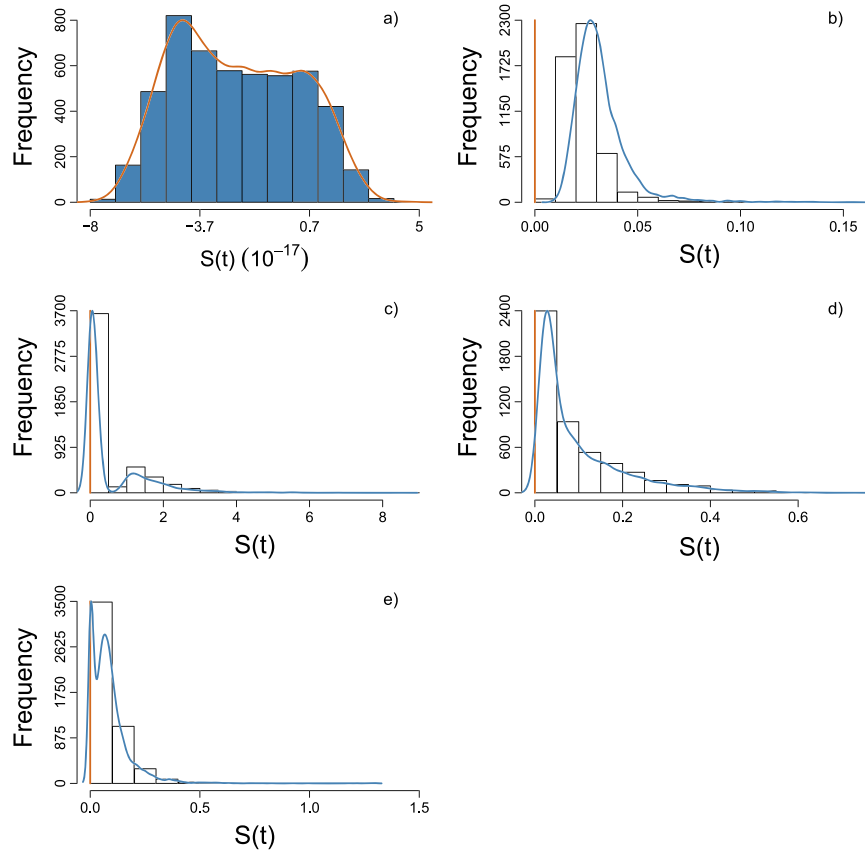
Figure 3. Bootstrap distributions of non-stationarity index S for stationary (steel blue) and non-stationary time series simulated from the Gaussian dynamic linear model (1). (a) Stationary AR(1), $\theta_t = 0.8, W = 0, V = 1$. Every other panel has this stationary density shown in orange. Models b to e as defined in the text.

panel shows the histogram associated with the stationary AR(1) process (in blue) with a kernel density estimate superimposed in orange. Here, the index has a positively skewed distribution with a mode close to zero. For the remaining panels, each histogram shows the bootstrap distribution of S for the other four models with the stationary AR(1) density estimate from the top left panel superimposed for comparison. For all the other, non-stationary models, the bootstrap distributions are shifted significantly to the right compared to the stationary case. The supporting information presents further figures that compare descriptive sample statistics of parameters of original and typical bootstrap samples for the aforementioned models in chronological order, for TVAR(1) in Figure 6, sinusoidal TVAR(1) in Figure 7, exponential TVAR(1) in Figure 8 and Gaussian TVAR(1) in Figure 9.

5 Non-stationarity of the earthquake and explosion data

This section applies the non-stationarity index to compare the earthquake, Y_t, and explosion, X_t, time series from Shumway & Stoffer (2006), which are available in the R statistical package astsa of Stoffer (2016) and shown in Figure 4. Here, $T = 2048$. As before, we use Gaussian dynamic linear model (1) as a model for both time series. Through careful modelling, we selected $W = 0.01, G_t = -0.8$ and $V = 1$. Such models permit flexible, yet simple, incorporation of time-varying mean and variance components for observed and latent dynamic signals.
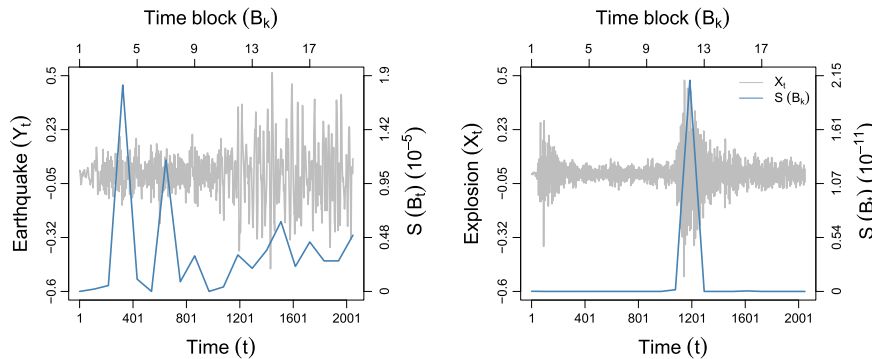
Figure 4. Left: earthquake time series (grey, left axis) with non-stationary index, S, (blue, right axis) superimposed computed on non-overlapping windows, B_k, defined in the text. Right: equivalent plot for explosion series.

We wish to see how S changes over the extent of each series. Hence, we partitioned the series into 20 chronologically ordered non-overlapping time windows, B_k, $k = 1, \ldots, 20$, 19 of length 100 and the final one of length 148, and computed S separately on each window. Following the advice of Shumway & Stoffer (2006, footnote on p. 229), we avoided overlapping windows. Figure 4 shows the two time series with the value of S computed on each window B_k.

The non-stationarity index for the earthquake data shows two bursts of high values of the index around $t = 300$ and $t = 600$. For the former, one can see a slight downward shift in the process itself; for the latter peak, it is difficult to see precisely what might be causing the larger value of S. There are smaller peaks around $t = 1200, 1500$ and 1700, where changes in the stochastic structure appear to occur. For the explosion data, the non-stationarity index is large around the point of the series where the variance rapidly increases ($t = 1200$) but seemingly misses the obvious structural change very early on in the series. The non-stationarity index for the earthquake series is much larger than for the explosion one. Both series are presented on approximately commensurate scales, and so the larger non-stationarity index for the earthquake data indicates a higher degree of non-stationarity. This might be due to it possessing sustained periods of regime change compared to the explosion data, which is akin to singular striking of an object.

6 Correcting for scale effects

The non-stationarity index S is affected by the scale or measured units of an underlying time series. For example, if we multiply a series by a scale factor, $c > 0$, then γ also increases by c and hence S increases by c^2. For situations where series are measured in the same units, for example, the lack of scale invariance might be seen as an advantage. In other situations, one might require a scale-invariant measure of non-stationarity. There is no optimum method for achieving scale invariance, but we have experimented with the following quantities:

$$S_{\boldsymbol{\sigma}}(Y_t) = S\left[Y_t \big/ \mathrm{sd}_{t \in B_k}(\{Y_r\})\right], \qquad (11)$$

$$S_{\boldsymbol{C}}(Y_t) = S\left(C_t^{-1/2} Y_t\right) \text{ and} \qquad (12)$$

$$S_{\boldsymbol{\gamma}}(B_k) = S(B_k) \big/ \left(\boldsymbol{\gamma}_k^T \boldsymbol{\gamma}_k\right), \qquad (13)$$

where $\mathrm{sd}_{t \in B_k}(\{Y_r\})$ is the standard deviation of the series in the block B_k, which overlaps time point y: this normalizes S by the local standard deviation. In $S_{\boldsymbol{C}}(Y_t)$, we normalize by the posterior variance C_t, while in $S_{\boldsymbol{\gamma}}(B_k)$, we divide the original index S by the squared norm of fitted spline coefficients.

Figure 5 plots $S_{\boldsymbol{\sigma}}(Y_t)$ and $S_{\boldsymbol{C}}(Y_t)$ for the earthquake and explosion data, and the supporting information Figure 10 shows the plot for $S_{\boldsymbol{\gamma}}(B_k)$. The normalization has an effect, particularly in the lower left panel of Figure 5 where the
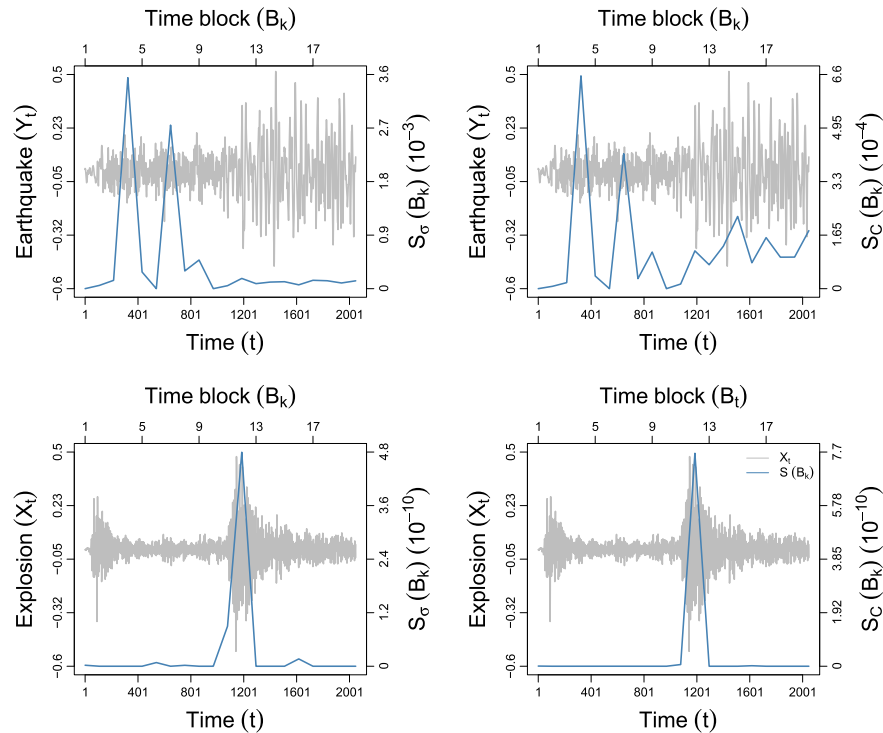
Figure 5. Normalized non-stationarity indices computed for the earthquake and explosion time series. Top row: earthquake. Bottom row: explosion. Left column: $S_{\sigma}(Y_t)$. Right column: $S_C(Y_t)$.

sizes of the non-stationarity indices for the two peaks are closer than previously. However, the normalization may seem to be less influential than expected, possibly because the scales of the earthquake and explosion series are reasonably commensurate to begin with.

7 Discussion

This article proposes a descriptive tool, $S(Y_t)$ from (5), that measures the degree of non-stationarity of a time series. This contrasts with hypothesis tests of stationarity that adopt a coarser binary approach to the question of time series stationarity. The top panels in Figure 1 clearly show that when the modelled θ_t is changing fastest (increasing G_t in different examples), the assessed non-stationarity increases as expected. To prevent relying on the results of S on a single realization, we adopt a bootstrap approach to obtain an estimate of the distribution of S with respect to the underlying stochastic process of the time series.

The non-stationarity index can be flexibly computed on a number of time-varying statistics such as spectra, autocorrelation functions at different lags or, as here, statistics arising from various stochastic models such as dynamic linear models. We can compute the index on the whole series or on a set of windows as shown earlier. The index could also be computed on streaming or online data, taking advantage of fast algorithms for spline computation such as in Section 5 of Weinert (2009).

Although our earlier proposal is one reasonable possible index of non-stationarity, it is certainly not the only one or the optimal (if such a thing exists). For example, we have already remarked that, for a TVAR(1) model with θ_t, following a linear trend as a function of t would result in a zero value of S. Hence, one might consider examination of indices

that depend on the first derivative of the time-varying statistic, for example, $\int \{g'(t)\}^2 \, dt$ and also using some method for estimating g for the particular time series in question. We have mostly focused on Gaussian examples that are well modelled by dynamic linear models, but other models could and should be used in other situations where one is faced with heavy-tailed errors or trend, for example.

Acknowledgements

References

Bandyopadhyay, S & Subba Rao, S (2016), 'A test for stationarity for irregularly spaced spatial data', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **78** (to appear).

Barlett, TE, Sykulski, AM, Olhede, SC, Lilly, JM & Early, JJ (2015), 'A power variance test for stationarity in complex valued signals', In IEEE 14th International Conference on Machine Learning and Applications (ICMLA), 911–916, Miami, Florida.

Bašta, M (2015), 'Wavelet-based test for time series non-stationarity', *Statistika*, **95**, 29–46.

Cardinali, A & Nason, GP (2010), 'Costationarity of locally stationary time series', *Journal of Time Series Econometrics*, **2**, 1–35.

Cardinali, A & Nason, GP (2016), 'Practical powerful wavelet packet tests for second-order stationarity', *Applied and Computational Harmonic Analysis*, **41** (to appear).

Dahlhaus, R (1997), 'Fitting time series models to nonstationary processes', *Annals of Statistics*, **25**, 1–37.

Dette, H, Preuß, P & Vetter, M (2011), 'A measure of stationarity in locally stationary processes with applications to testing', *Journal of the American Statistical Association*, **106**, 1113–1124.

Dwivedi, Y & Subba Rao, S (2011), 'A test for second-order stationarity of a time series based on the discrete Fourier transform', *Journal of Time Series Analysis*, **32**, 68–91.

Golub, GH, Heath, M & Wahba, G (1979), 'Generalized cross-validation as a method for choosing a good ridge parameter', *Technometrics*, **21**, 215–223.

Good, IJ & Gaskins, RA (1971), 'Nonparametric roughness penalties for probability densities', *Biometrika*, **58**, 255–277.

Green, PJ & Silverman, BW (1993), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, CRC Press, Boca Raton, FL.

Jentsch, C & Subba Rao, S (2015), 'A test for second order stationarity of a multivariate time series', *Journal of Econometrics*, **185**, 124–161.

Jin, L, Wang, S & Wang, H (2015), 'A new non-parametric stationarity test of time series in the time domain', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 893–922.

Kreiss, J-P & Paparoditis, E (2012), 'The hybrid wild bootstrap for time series', *Journal of the American Statistical Association*, **107**, 1073–1084.

Kreiss, J-P & Paparoditis, E (2015), 'Bootstrapping locally stationary processes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 267–290.

Nason, GP (2013), 'A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **75**, 879–904.

Nason, GP, von Sachs, R & Kroisandt, G (2000), 'Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**, 271–292.

Paparoditis, E (2009), 'Testing temporal constancy of the spectral structure of a time series', *Bernoulli*, **15**, 1190–1221.

Paparoditis, E (2010), 'Validating stationarity assumptions in time series analysis by rolling local periodograms', *Journal of the American Statistical Association*, **105**, 839–851.

Priestley, MB & Subba Rao, T (1969), 'A test for non-stationarity of time-series', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **31**, 140–149.

Ramsay, JO & Silverman, BW (2010), *Functional Data Analysis*, *Springer Science+ Business Media*, New York, USA.

Ramsey, JO & Wickham, H (2014), *Functional Data Analysis*, *R Foundation for Statistical Computing*, Vienna, Austria. R Package version 2.4.4. https://cran.r-project.org/web/packages/fda/fda.pdf/.

Sergides, M & Paparoditis, E (2008), 'Bootstrapping the local periodogram of locally stationary processes', *Journal of Time Series Analysis*, **29**, 264–299.

Shumway, RH & Stoffer, DS (2006), *Time Series Analysis and Its Applications: With R Examples*, 4th edn., *Springer*, New York.

Silverman, BW (1985), 'Some aspects of the spline smoothing approach to non-parametric regression curve fitting', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **47**, 1–52.

Stoffer, DS & Wall, KD (1991), 'Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter', *Journal of the American Statistical Association*, **86**, 1024–1033.

Stoffer, D (2016), astsa: Applied statistical time series analysis. R package version 1.4. https://CRAN.R-project.org/package=astsa.

Taylor, SL, Eckley, IA & Nunes, MA (2014), 'A test of stationarity for textured images', *Technometrics*, **56**, 291–301.

Von Sachs, R & Neumann, MH (2000), 'A wavelet-based test for stationarity', *Journal of Time Series Analysis*, **21**, 597–613.

Weinert, HL (2009), 'A fast compact algorithm for cubic spline smoothing', *Computational Statistics & Data Analysis*, **53**, 932–940.

West, M & Harrison, J (1997), *Bayesian Forecasting and Dynamic Models*, *Springer*, New York.

Wood, S (2006), *Generalized Additive Models: An Introduction with R*, *CRC Press*, Boca Raton, FL.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.