Supplementary Materials for: "Modeling Interval Trendlines Symbolic Singular Spectrum Analysis for Interval Time Series"

Miguel DE CARVALHO[†] and Gabriel MARTOS[‡]

[†]*School of Mathematics, University of Edinburgh, EH9 3FD, UK*

[‡]*Department of Mathematics and Statistics, Universidad Torcuato Di Tella, Argentina*

1 Additional Numerical Evidence

1.1 Supporting Monte Carlo Summaries

The numerical reports below will supplement the Monte Carlo simulation study from the paper (Section 3). In Figure 2 in the paper we depict the empirical distribution of HR in the context of a Monte Carlo simulation experiment. Below, in Tables 1 and 2 we also report the Monte Carlo mean HR and the corresponding standard errors in the experiment corresponding to Scenarios A and B, respectively. As it can be seen from Tables 1 and 2, as the sample size increases, the Monte Carlo mean HR and its standard deviation tend to decrease, regardless of the value of m.

[†]*Address for correspondence*: M. de Carvalho, School of Mathematics, The University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK.

E-mail: miguel.decarvalho@ed.ac.uk

Table 1: Scenario A: Monte Carlo mean HR and standard error (in parenthesis) over different sample sizes (n) and different number of components included in the model (m).

m		Sample size (n)		
		10	250	1000
1	IVSSA on X_t	0.641	0.628	0.622
		(0.015)	(0.008)	(0.004)
	IVSSA on Y_t	1.314	1.305	1.301
		(0.072)	(0.045)	(0.022)
	h–MIVSSA	1.207	1.208	1.21
		(0.032)	(0.020)	(0.010)
	v–MIVSSA	2.299	2.298	2.297
		(0.020)	(0.012)	(0.006)
2	IVSSA on X_t	0.389	0.362	0.347
		(0.053)	(0.031)	(0.015)
	IVSSA on Y_t	0.740	0.675	0.641
		(0.105)	(0.062)	(0.032)
	h–MIVSSA	0.824	0.79	0.785
		(0.111)	(0.092)	(0.071)
	v–MIVSSA	0.943	0.917	0.908
		(0.068)	(0.043)	(0.021)
3	IVSSA on X_t	0.292	0.214	0.152
		(0.064)	(0.039)	(0.017)
	IVSSA on Y_t	0.47	0.282	0.147
		(0.151)	(0.075)	(0.036)
	h–MIVSSA	0.634	0.612	0.616
		(0.078)	(0.049)	(0.026)
	v–MIVSSA	0.651	0.628	0.624
		(0.056)	(0.028)	(0.013)

m		Sample size (n)		
		10	250	1000
4	IVSSA on X_t	0.346	0.245	0.152
		(0.059)	(0.034)	(0.035)
	IVSSA on Y_t	0.589	0.355	0.189
		(0.149)	(0.067)	(0.031)
	h–MIVSSA	0.544	0.387	0.342
		(0.116)	(0.051)	(0.025)
	v–MIVSSA	0.573	0.427	0.350
		(0.107)	(0.059)	(0.022)
5	IVSSA on X_t	0.423	0.307	0.184
		(0.057)	(0.038)	(0.044)
	IVSSA on Y_t	0.728	0.492	0.273
		(0.116)	(0.069)	(0.033)
	h–MIVSSA	0.542	0.318	0.181
		(0.149)	(0.072)	(0.036)
	v–MIVSSA	0.578	0.397	0.289
		(0.130)	(0.069)	(0.034)
6	IVSSA on X_t	0.460	0.328	0.195
		(0.053)	(0.033)	(0.035)
	IVSSA on Y_t	0.806	0.537	0.294
		(0.107)	(0.065)	(0.028)
	h–MIVSSA	0.628	0.378	0.213
		(0.134)	(0.067)	(0.030)
	v–MIVSSA	0.657	0.407	0.157
		(0.124)	(0.092)	(0.036)

Table 2: Scenario B: Monte Carlo mean HR and standard error (in parenthesis) over different sample sizes (n) and different number of components included in the model (m).

m		Sample size (n)		
		10	250	1000
1	IVSSA on X_t	0.639	0.628	0.622
		(0.014)	(0.008)	(0.004)
	IVSSA on Y_t	1.313	1.305	1.301
		(0.071)	(0.044)	(0.022)
	h–MIVSSA	1.206	1.209	1.21
		(0.035)	(0.020)	(0.010)
	v–MIVSSA	2.299	2.298	2.297
		(0.020)	(0.012)	(0.006)
2	IVSSA on X_t	0.388	0.361	0.348
		(0.052)	(0.031)	(0.015)
	IVSSA on Y_t	0.742	0.675	0.640
		(0.111)	(0.062)	(0.032)
	h–MIVSSA	0.826	0.789	0.785
		(0.116)	(0.093)	(0.072)
	v–MIVSSA	0.944	0.918	0.908
		(0.068)	(0.041)	(0.020)
3	IVSSA on X_t	0.295	0.213	0.152
		(0.066)	(0.039)	(0.017)
	IVSSA on Y_t	0.482	0.284	0.150
		(0.162)	(0.079)	(0.040)
	h–MIVSSA	0.641	0.608	0.616
		(0.087)	(0.047)	(0.023)
	v–MIVSSA	0.655	0.625	0.624
		(0.064)	(0.029)	(0.012)

m		Sample size (n)		
		10	250	1000
4	IVSSA on X_t	0.346	0.245	0.152
		(0.058)	(0.034)	(0.035)
	IVSSA on Y_t	0.584	0.360	0.189
		(0.142)	(0.064)	(0.03)
	h–MIVSSA	0.563	0.386	0.343
		(0.127)	(0.050)	(0.025)
	v–MIVSSA	0.585	0.429	0.351
		(0.114)	(0.060)	(0.024)
5	IVSSA on X_t	0.427	0.310	0.186
		(0.058)	(0.037)	(0.042)
	IVSSA on Y_t	0.738	0.493	0.275
		(0.117)	(0.071)	(0.031)
	h–MIVSSA	0.566	0.322	0.181
		(0.153)	(0.074)	(0.034)
	v–MIVSSA	0.595	0.399	0.290
		(0.131)	(0.071)	(0.033)
6	IVSSA on X_t	0.459	0.330	0.195
		(0.054)	(0.033)	(0.036)
	IVSSA on Y_t	0.81	0.535	0.294
		(0.112)	(0.062)	(0.029)
	h–MIVSSA	0.639	0.376	0.211
		(0.143)	(0.066)	(0.030)
	v–MIVSSA	0.662	0.42	0.156
		(0.128)	(0.079)	(0.036)

1.2 Monte Carlo Evidence Under Random Ranges

Here we explore the performance of the proposed methods in a context where the relationship between the lower and upper limits of the set-valued stochastic processes are random. To this aim, consider the following data generation mechanism:

$$x_t = [\mu_t^x + \varepsilon_t^x, \mu_t^x + \theta_t^x + \varepsilon_t^x], \qquad y_t = [\mu_t^y + \varepsilon_t^y, \theta_t^y \mu_t^y + 2 + \varepsilon_t^y], \tag{1}$$

where $t \in T = \{2i\pi/n\}_{i=1}^n$, $\mu_t^x = 8+t+\sin(\pi t)$, $\mu_t^y = \sqrt{t}+\cos(\pi t/2)$, are deterministic functions, and θ_t^x, θ_t^y are random independent and uniformly distributed variables on the interval $[1, 3]$. The error terms ε_t^x and ε_t^y are zero mean normally distributed with covariance function

$$\Sigma(s,t) \equiv \begin{bmatrix} \mathrm{Cov}(\varepsilon_t^x, \varepsilon_s^x) & \mathrm{Cov}(\varepsilon_t^x, \varepsilon_s^y) \\ \mathrm{Cov}(\varepsilon_t^y, \varepsilon_s^x) & \mathrm{Cov}(\varepsilon_t^y, \varepsilon_s^y) \end{bmatrix} = \delta(t-s) \begin{bmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{bmatrix},$$

where $\delta(0) = 1$ and $\delta(t-s) = 0$ for $t \neq s$. Note that the range of the interval-valued process x_t is θ_t^x, and thus $\mathrm{E}(\theta_t^x) = 2$; that is, the scenarios in (1) add a another level of randomness to the scenarios considered in Section 3 in the paper. In Scenario A we set $\rho = 0$ and $\sigma^2 = 1$, thus $\{x_t\}$ and $\{y_t\}$ are independent interval-valued processes, and in Scenario B we set $\rho = 1/2$ and $\sigma^2 = 1$, leading to dependent interval-valued processes.

In Figure 1 we present one instance of an interval trendline estimate yield using our methods corresponding to a one shot experiment for Scenarios A–B. In Figure 2 we depicts side-by-side boxplots of HRs for Scenarios A and B corresponding to a Monte Carlo simulation experiment based on $S = 1\,000$ data simulations. As it can be seen from Figure 2, the simulation results are similar to those presented on Section 3 in the paper: As the sample size increases the HR tends to decrease, regardless of the number of ERC; this thus indicates a better performance, from an Hausdorff residual perspective, of the proposed methods as the number of observations increases. All in all, this supplemental analysis shows that the obtained findings are similar to those in Section 3; as expected, the scale of the y-axis in Figure 2 is not however the same, due to the fact that we have further sources of randomness (θ_t^x and θ_t^y).

1.3 Numerical Experiments on a Block Min–Max Setting

In some applied settings the interest is on modeling intervals defined by the minima and maxima of a process that is sampled on a higher frequency. In this section we illustrate the proposed

method within that framework. We do this by constructing block minima and block maxima versions of Scenarios A and B described in Section 3.1 of the main paper. Specifically, data are generated as follows

$$x_t = [\mu_t^x + w_t^x, \mu_t^x + z_t^x], \quad y_t = [\mu_t^y + w_t^y, \mu_t^y + z_t^y], \tag{2}$$

where $t \in T = \{2i\pi/n\}_{i=1}^n$, $\mu_t^x = 8 + t + \sin(\pi t)$, $\mu_t^y = \sqrt{t} + \cos(\pi t/2)$, $w_t^x = \min\{\varepsilon_{t,1}^x, \ldots, \varepsilon_{t,100}^x\}$, $z_t^x = \max\{\varepsilon_{t,1}^x, \ldots, \varepsilon_{t,100}^x\}$, $w_t^y = \min\{\varepsilon_{t,1}^y, \ldots, \varepsilon_{t,100}^y\}$, and $z_t^y = \max\{\varepsilon_{t,1}^y, \ldots, \varepsilon_{t,100}^y\}$. By keeping in mind the block minima and block maxima nature of the processes above, we will refer to (2) as block min–max processes. In Scenario A, we consider $\{\varepsilon_{t,j}^x\}$ and $\{\varepsilon_{t,j}^y\}$ to be two independent sequences of independent and identically distributed random variables following a zero mean and unit variance normal distribution, thus $\{x_t\}$ and $\{y_t\}$ are independent block min–max processes. In Scenario B, we additionally assume that $\text{Cov}(\varepsilon_{t,j}^x, \varepsilon_{s,j'}^y) = 1/2\delta(t-s)\delta(j-j')$, leading to dependent block min–max processes. In Figure 3 we depict the fits corresponding to a one-shot analysis for $n = 100$, and as can be seen IVSSA and MIVSSA are both able to capture the main dynamics of the block min–max process in both scenarios.

References

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer.

Block minima and block maxima are standard terminology in fields such as extreme value theory so to describe respectively maxima or minima of a block of observations; see, for instance, Coles (2001, Ch. 3).

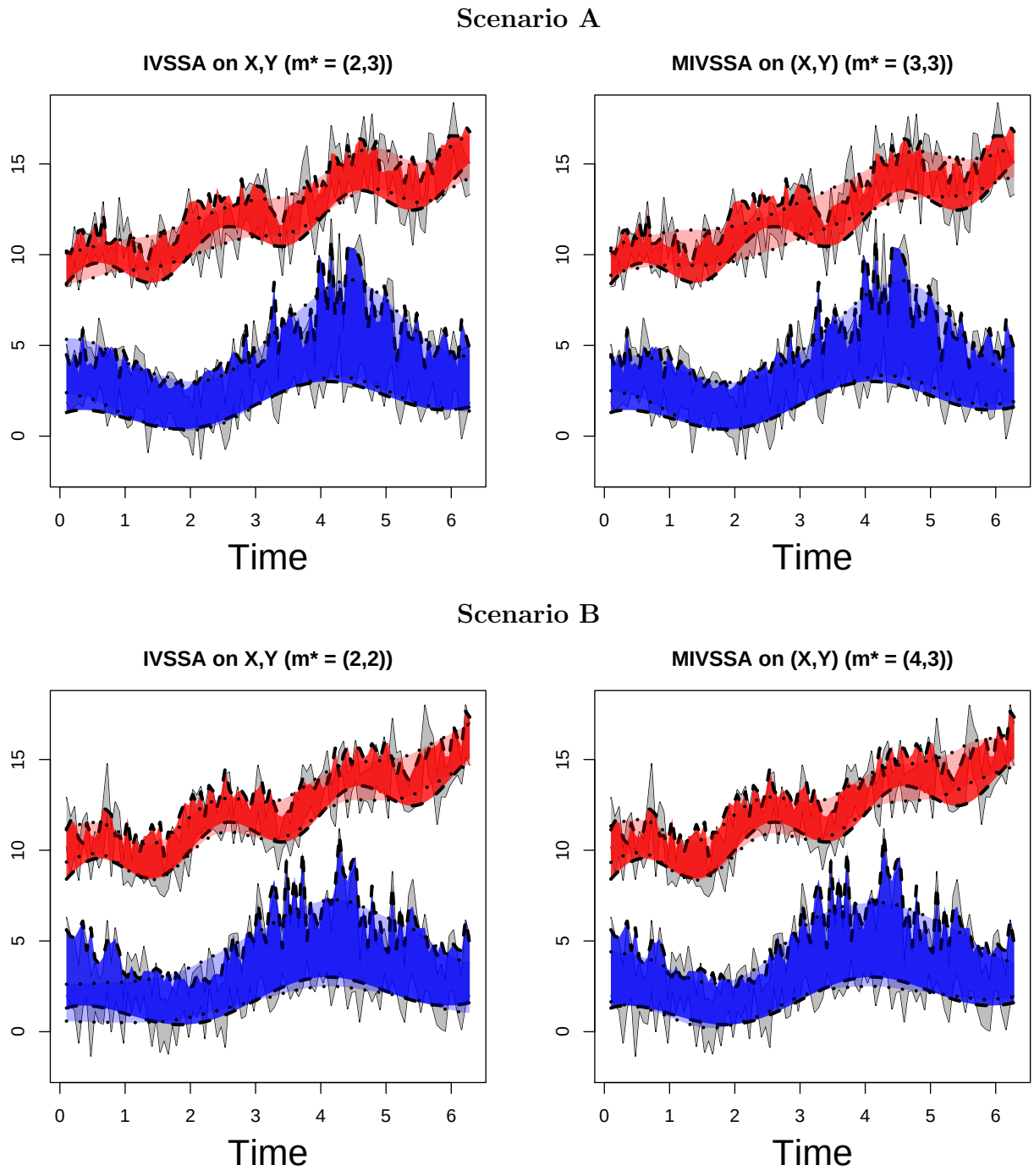

Figure 1: One shot experiments for Scenario A and B from Section 1.2. The solid gray areas corresponds to raw interval data for $\{x_t\}$ and $\{y_t\}$ as in (1). Conditional means $[\mu_t^X, \mu_t^X + 2]$ and $[\mu_t^Y, 2(\mu_t^Y + 1)]$ and trendline estimators are depicted in solid and transparent red and blue respectively.

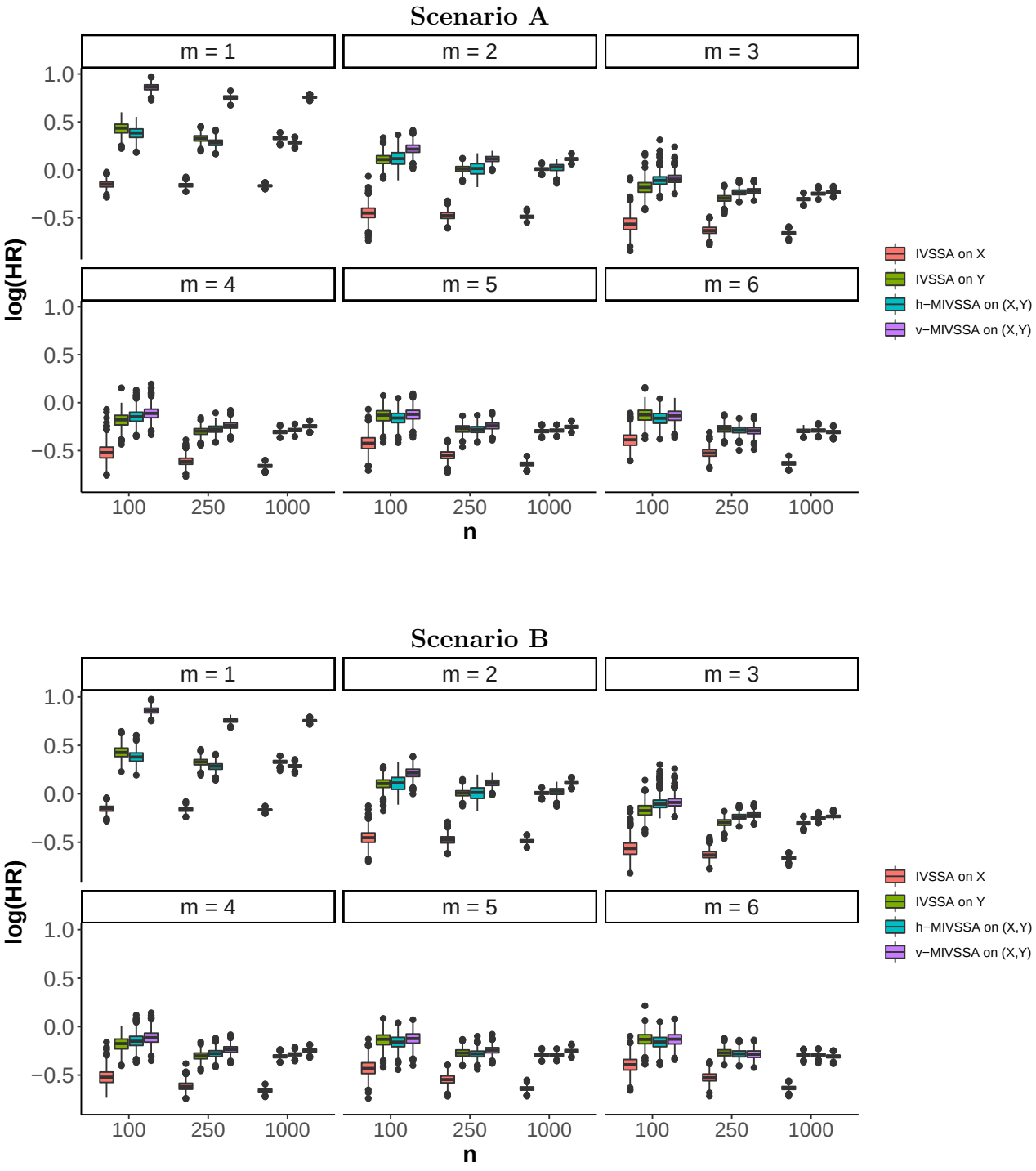

Figure 2: Monte Carlo simulation study Hausdorff residuals (HR). Boxplots of HR for Scenarios A and B from Section 1.2 over different values of m and n.

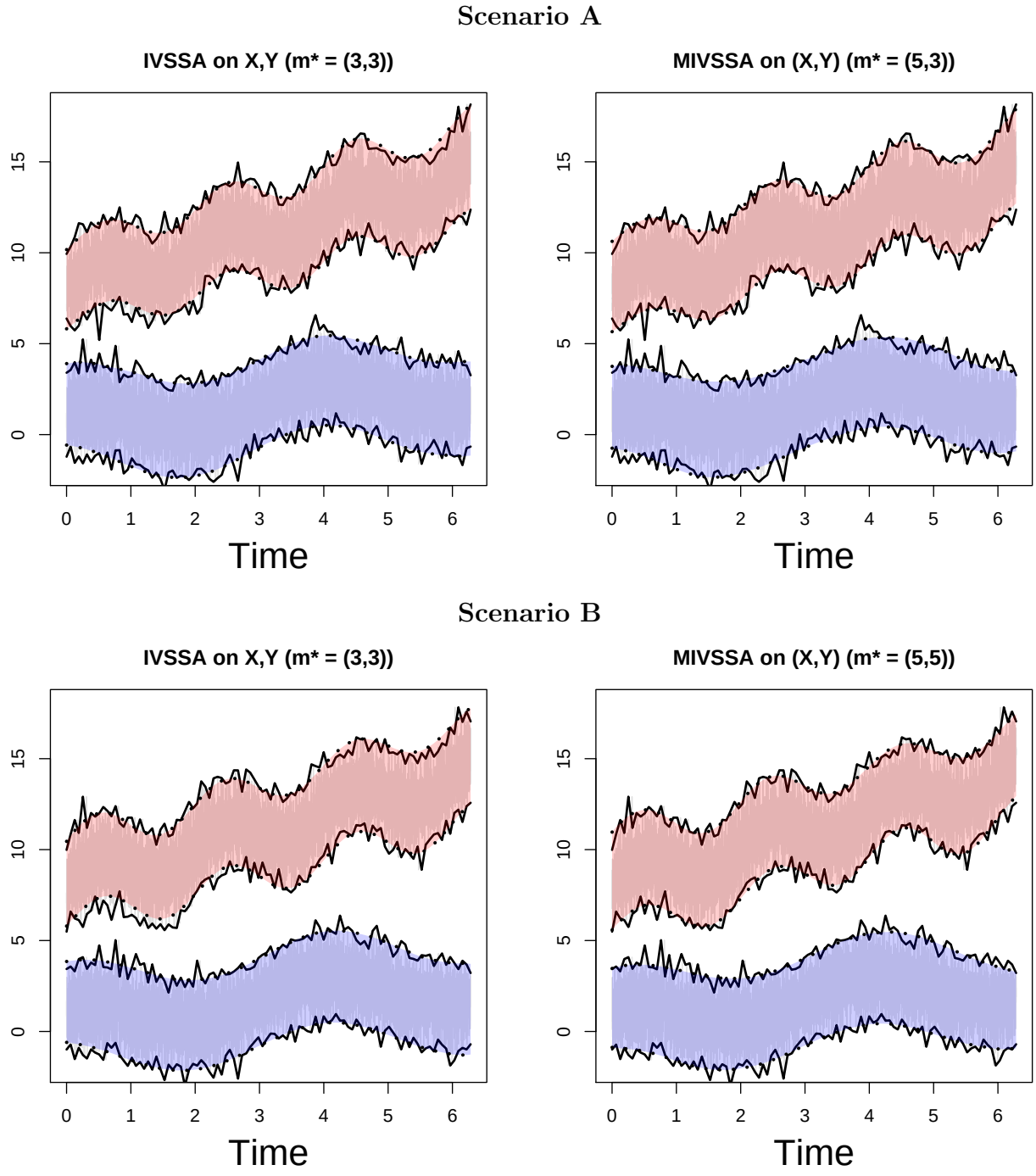

Figure 3: One shot experiments for Scenarios A and B from Section 1.3. The solid gray lines corresponds to raw high frequency data while the solid black lines corresponds to the block min–max processes $\{x_t\}$ and $\{y_t\}$ as in (2). The IVSSA (left) and MIVSSA (right) interval trendlines corresponding to $\{x_t\}$ and $\{y_t\}$ are depicted in transparent red and blue respectively.