Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Finite sample forecast properties and window length under breaks in cointegrated systems

by

Luca Nocciola

Granger Centre Discussion Paper No. 19/07

University of Nottingham
UK | CHINA | MALAYSIA

Finite sample forecast properties and window length under breaks in cointegrated systems

Luca Nocciola[*†]

Abstract

We show that extending the estimation window prior to structural breaks in cointegrated systems can be beneficial for forecasting performance and highlight under which conditions. In doing so, we generalize the Pesaran & Timmermann (2005)'s forecast error decomposition and show that it depends on four terms: 1) a period ahead risk; 2) a bias due to a conditional mean shift; 3) a bias due to a variance mismatch; 4) a gap term valid only conditionally. We also derive new expressions for the estimators of the adjustment matrix and a constant, which are auxiliary to the decomposition. Finally, we introduce new simulation-based estimators for the finite sample forecast properties which are based on the derived decomposition. Our finding points out that, in some cases, we can neglect parameter instability by extending the window backward and be insured against higher forecast risk under this model class as well, generalizing Pesaran & Timmermann (2005)'s result. Our result gives renewed importance to break tests, in order to distinguish cases when break-neglection is (not) appropriate.

JEL: C22, C53
Keywords: Finite sample forecast properties; MSE; Structural breaks; Cointegration; Expanding window estimator

[*]Research Centre, Deutsche Bundesbank, Wilhelm-Epstein-Str. 14, 60431 Frankfurt, Germany, e-mail: luca.nocciola@bundesbank.de.

[†]This paper has been developed during my PhD under the supervision of Prof. Uwe Hassler at the Chair of Statistics and Econometric Methods, Goethe University Frankfurt, Germany. I thank Uwe Hassler, Hashem Pesaran, Kim-Kurz Jeong-Ryeol, Alessandra Luati, Paolo Santucci de Magistris, Xu Han, Mehdi Hosseinkouchak, Ying Lun Cheung and Esteban Prieto for very insightful comments that have significantly improved the quality of the paper. Moreover, I thank the participants of the NBER-NSF time series conference 2019, the 4$^{\text{th}}$ Vienna time series workshop 2019 and the workshop 'Hot topics in econometrics' at Goethe University for useful discussions. The views expressed in this paper are my own and do not represent the views of the Deutsche Bundesbank or the ESCB.

1 Introduction

Which estimation window is optimal for forecasting when the model used is subject to structural breaks?

It is well known that parameter breaks are pervasive to (parametric) models fitted to economic time series (Stock & Watson (1996)) and that instability can be harmful both in terms of forecast bias and error variance (Clements & Hendry (1998, 2001), Pesaran & Timmermann (2004) and Castle *et al.* (2016)). Against this background, a series of papers address the question of whether there exists an estimation window which assures forecast optimality under parameter jumps. For general autoregressive processes, Pesaran & Timmermann (2005) show that, under certain conditions, it is better for estimation to use also pre-break data (vs focus on post-break data only), thanks to declining bias and forecast error variance with expanding estimation window. For the multiple linear regression model with strictly exogenous regressors, Pesaran & Timmermann (2007) reach an identical conclusion, although the bias-efficiency trade-off is binding in that case, i.e. the bias always increase when we include contaminated data. In their article, Pesaran & Timmermann (2007) propose methods based on cross-validation and combination in order to determine the optimal estimation window. A similar contribution is provided by Inoue *et al.* (2017), who offer a new method to determine the optimal rolling window size in a smoothly time-varying parameter model. Other important contributions are by Pesaran & Pick (2011), Pesaran *et al.* (2013) and Giraitis *et al.* (2013).

We tackle the same initial question indirectly as Pesaran & Timmermann (2005), who assess the sensitivity of finite sample properties of forecasts to the starting point of the estimation window. We base our investigation, however, on a multivariate generalization of the autoregressive process. In particular, we contribute by evaluating the forecasting performance under structural breaks of vector error correction (VEC) models (Johansen (1988, 1991)) in relation to the sample size used for estimation. We assess under which conditions and to what extent parameter instability causes adverse effects to forecasting by monitoring the conditional bias and mean square error (MSE) in connection to the initial point of the estimation window and draw conclusions on the window length which ensures best forecasting performance.

We show that extending the estimation window prior to structural breaks in VEC models can be beneficial for forecasting performance and highlight under which conditions. We base our simulation on a generalized Pesaran & Timmermann (2005)'s forecast error decomposition, which we derive, and show that it depends on four components: 1) a period ahead risk; 2) a bias due to a conditional mean shift; 3) a bias due to a variance mismatch; 4) a gap term valid only conditionally. We also derive new expressions for the estimators of the adjustment matrix and a constant, which are auxiliary to the decomposition. Finally, we introduce new simulation-based estimators for the finite sample forecast properties which are based on the derived decomposition.

Our finding points out that, under the derived conditions, we can neglect parameter instability by extending the window backward and be insured against higher forecast risk under this model class as well, generalizing Pesaran & Timmermann (2005)'s result. Our result suggests, in some cases, a reduced need of break-correction strategies and also gives renewed importance to testing for structural breaks, in order to distinguish cases when break-neglection is (not) appropriate. In general, the choice of the estimation window depends on the timing, location and direction of the breaks.

2 Model

2.1 Constant parameters

A general, $N-$dimensional VEC model is given by $\{\boldsymbol{y}_t\}_{t\in\mathbb{N}_0} \sim CI(d,b)$, where $d = \max(d_1, \ldots, d_N)$, $d \in \mathbb{N}$ being the order of integration, $b \in \mathbb{N}$ being the absolute decrease in the order of integration thanks to cointegration, $CI()$, and N is a finite number of series.[1] Generality is given by the presence of k lags, i.e. $\boldsymbol{y}_t \sim \text{VEC}(k)$, and an unrestricted deterministic component.

Assumption 1. *(i) $d = b = 1$, so that $\boldsymbol{y}_t \sim CI(1,1)$; (ii) $k = 1$; (iii) the deterministic term is a constant $\boldsymbol{\pi} = -\boldsymbol{\gamma\rho}$, i.e. restricted to the cointegration space, where $\boldsymbol{\rho}$ is an $(r \times 1)$-dim constant; and (iv) $\boldsymbol{e}_t \overset{w.n.}{\sim} \mathcal{D}(\mathbf{0}_N, \boldsymbol{\Sigma})$, i.e. follows a general distribution with existent and finite first two moments, where $\overset{w.n.}{\sim}$ stands for "white noise" and $\boldsymbol{\Sigma}$ is positive semi-definite.*

Assumption 1 implies formally that

$$\underset{N\times 1}{\Delta \boldsymbol{y}_t} = \boldsymbol{\pi} + \boldsymbol{\gamma\delta}'\boldsymbol{y}_{t-1} + \boldsymbol{e}_t \tag{1}$$

where $\Delta \boldsymbol{y}_t \sim I(0)$, Δ being the difference operator and $I()$ standing for "integrated of some order", $\boldsymbol{y}_t \sim I(1)$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are $(N \times r)$-dim adjustment and cointegration matrices forming a product matrix of reduced rank $rk(\boldsymbol{\gamma\delta}') = r \in (0, N)$, where $rk()$ denotes the rank of a matrix, and $\boldsymbol{e}_t \sim I(0)$.

2.2 Parameter instability

We suppose the parameter space (conditional mean and variance) of (1) to suffer from an instability at T_c.

[1] Hereafter, we simplify the notation for each stochastic process by omitting $\{\}_{t\in\mathbb{N}_0}$. Furthermore, we refer to dimensionality by the symbol $-$dim.

Assumption 2. *For $T_c \leq t < T$, T finite: (i) T_c is known; (ii) the conditional mean*[2] *shifts permanently to $\mathbb{E}_e(\Delta \boldsymbol{y}_t|\boldsymbol{\Omega}_{t-1}) = \boldsymbol{\pi}^* + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}\boldsymbol{y}_{t-1}$ due to $\boldsymbol{\pi}$ shifting to $\boldsymbol{\pi}^*$, $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}^*$ and $\boldsymbol{\delta}$ to $\boldsymbol{\delta}^*$, where $\boldsymbol{\pi}^* = -\boldsymbol{\gamma}^*\boldsymbol{\rho}^*$ and $rk(\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}) = r$; (iii) the conditional variance shifts permanently to $\mathbb{V}_e(\Delta \boldsymbol{y}_t|\boldsymbol{\Omega}_{t-1}) = \boldsymbol{\Sigma}^*$ due to $\boldsymbol{\Sigma}$ shifting to $\boldsymbol{\Sigma}^*$, where the expectation operators ($\mathbb{E}_e()$ and $\mathbb{V}_e()$) are defined over the distribution of $\boldsymbol{e}_t^{(*)}$ and $\boldsymbol{\Omega}_{t-1} = \sigma\{\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots\}$ is a filtration up to $t-1$.*

Equivalently,

$$\underset{N\times 1}{\Delta \boldsymbol{y}_t} = \begin{cases} \boldsymbol{\pi} + \boldsymbol{\gamma}\boldsymbol{\delta}'\boldsymbol{y}_{t-1} + \boldsymbol{e}_t, & \text{if } t < T_c \\ \boldsymbol{\pi}^* + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}\boldsymbol{y}_{t-1} + \boldsymbol{e}_t^*, & \text{if } t \geq T_c \end{cases} \tag{2}$$

where $\boldsymbol{e}_t^*$ is the error process characterized by the new covariance matrix.

3 Conditional forecasting

Following Pesaran & Timmermann (2005, 2007), we define $s - k(> 0)$ to be the starting point of the estimation window, where, following Assumption 1, (ii), k, the number of lags or initial values, is $k = 1$, implying the window to start at $s - 1$. The parameter $s \in \{2, \ldots, T\}$, which allows the break to happen during the estimation period (Pesaran & Timmermann (2005)). We analyze the one-step ahead forecasts from (2) conditioning on $\boldsymbol{\Omega}_T = \boldsymbol{\Omega}_{s,T} = \sigma\{\boldsymbol{y}_T, \boldsymbol{y}_{T-1}, \ldots, \boldsymbol{y}_s\}$, i.e. a filtration up to T which excludes a random initial value $\boldsymbol{y}_{s-1}$.[3] Parameter estimates are therefore random variables whose distribution depends on the distribution of the initial value

[2]It is enough for only one of the parameters among $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ to shift in order to obtain an instability in the conditional mean. Moreover, $\boldsymbol{\pi}$ is not variation free, i.e. a shift, e.g., in $\boldsymbol{\gamma}$ induces a change in $\boldsymbol{\pi}$ as well.

[3]We neglect unconditional results since past information is always available.

(Pesaran & Timmermann (2005), fn. 7).

3.1 Bias and MSE

The estimated parameters (up to T) are functions of s, i.e. $\hat{\boldsymbol{\pi}}(s)$, $\hat{\boldsymbol{\gamma}}(s)$ and $\hat{\boldsymbol{\delta}}(s)$. Conditional forecasts are in turn functions of s too, i.e.

$$\underset{N\times 1}{\Delta\hat{\boldsymbol{y}}_{T+1|T}(s)} = \hat{\boldsymbol{\pi}}(s) + \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)\boldsymbol{y}_T \tag{3}$$

and forecast errors, defined as $\hat{\boldsymbol{\xi}}_{T+1|T}(s) = \Delta\boldsymbol{y}_{T+1} - \Delta\hat{\boldsymbol{y}}_{T+1|T}(s)$, follow

$$\underset{N\times 1}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)} = (\boldsymbol{\pi}^* - \hat{\boldsymbol{\pi}}(s)) + (\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s))\boldsymbol{y}_T + \boldsymbol{e}^*_{T+1} \tag{4}$$

providing a multivariate generalization of Pesaran & Timmermann (2005, 2007), eq.(8), (9) and (3) on p.189-190 and p.137, resp. Under a parameter break, forecasts are conditionally biased and the bias is s-dependent, i.e.

$$\mathbb{E}_e(\underset{N\times 1}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)}|\boldsymbol{\Omega}_T) = (\boldsymbol{\pi}^* - \mathbb{E}_e(\hat{\boldsymbol{\pi}}(s)|\boldsymbol{\Omega}_T)) + (\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} - \mathbb{E}_e(\hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)|\boldsymbol{\Omega}_T))\boldsymbol{y}_T \tag{5}$$

similarly to Pesaran & Timmermann (2005), eq.(33) on p.196.[4] The conditional MSE in matrix form is given by

$$\mathbb{E}_e(\underset{N\times N}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)\hat{\boldsymbol{\xi}}'_{T+1|T}(s)}|\boldsymbol{\Omega}_T) = \mathbb{E}_e\left[((\boldsymbol{\pi}^* - \hat{\boldsymbol{\pi}}(s)) + (\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s))\boldsymbol{y}_T)^2\middle|\boldsymbol{\Omega}_T\right] + \boldsymbol{\Sigma}^* \tag{6}$$

[4] Forecasts are conditionally biased even in the absence of breaks due to the finite sample bias inherent in the estimated parameters (quantified in e.g. van Garderen & Boswijk (2014)). However, in the case of no break, if parameters were utopistically known ex-ante ($\hat{\boldsymbol{\pi}}(s) = \boldsymbol{\pi}$ and $\hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s) = \boldsymbol{\gamma}\boldsymbol{\delta}'$), forecasts would instead be conditionally unbiased, i.e. $\mathbb{E}_e(\hat{\boldsymbol{\xi}}_{T+1|T}(s)|\boldsymbol{\Omega}_T) = \mathbb{E}_e(\boldsymbol{e}_{T+1}|\boldsymbol{\Omega}_T) = \boldsymbol{0}_N$, and the conditional MSE in matrix form would equate the conditional forecast error variance, i.e. $\mathbb{E}_e(\hat{\boldsymbol{\xi}}_{T+1|T}(s)\hat{\boldsymbol{\xi}}'_{T+1|T}(s)|\boldsymbol{\Omega}_T) = \mathbb{V}_e(\hat{\boldsymbol{\xi}}_{T+1|T}(s)|\boldsymbol{\Omega}_T) = \mathbb{E}_e(\boldsymbol{e}_{T+1}\boldsymbol{e}'_{T+1}|\boldsymbol{\Omega}_T) = \boldsymbol{\Sigma}$, providing a lower bound for forecast efficiency (Clements & Hendry (2001), univariate case, eq.(4) on p.3, and Hendry (2006), multivariate case, eq.(19) on p.406, where the horizon is generalized to h).

and is, in turn, dependent on s (a multivariate extension of Pesaran & Timmermann (2005), eq.(10) on p.190).[5] Univariate measures of bias and forecast accuracy are obtained, resp., via the Euclidean distance and by taking the trace of (6), or, equivalently, the expected squared Frobenius norm of (4), i.e. $\mathbb{E}_e(\|\hat{\boldsymbol{\xi}}_{T+1|T}(s)\|_F^2|\boldsymbol{\Omega}_T) = \mathbb{E}_e(tr(\hat{\boldsymbol{\xi}}_{T+1|T}(s)\hat{\boldsymbol{\xi}}'_{T+1|T}(s))|\boldsymbol{\Omega}_T)$, where $tr()$ is the trace operator (linear in its arguments) and $\|\|_F$ is the Frobenius norm. In order to evaluate the bias and MSE analytically as a function of s, explicit expressions in terms of the innovations are needed, where the former depend on the initial value distribution, and can be recovered through the estimated parameters.

3.1.1 Unconditional moments of the initial value distribution

Based on an assumption on the origin of the process, we can compute the first two unconditional moments of the distribution of $\boldsymbol{y}_t$ at the starting point of the estimation window, $s-1$.

Assumption 3. *The value initializing the process is given by $\boldsymbol{y}_1 = \boldsymbol{\pi} + \boldsymbol{e}_1$.*

Lemma 1. *Under Assumptions 1-3, the first two unconditional moments of the distribution of $\boldsymbol{y}_{s-1}$ are given by, for $s-1 < T_c$,*

$$\underset{N\times 1}{\mathbb{E}_e(\boldsymbol{y}_{s-1})} = \sum_{j=0}^{s-2}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^j\boldsymbol{\pi} \tag{7}$$

and

$$\underset{N\times N}{\mathbb{V}_e(\boldsymbol{y}_{s-1})} = \sum_{j=0}^{s-2}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^j\boldsymbol{\Sigma}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{j'} \tag{8}$$

[5]The notation $()^2$ denotes the self outer product.

and by, for $s-1 \geq T_c$,

$$\underset{N\times 1}{\mathbb{E}_e(\boldsymbol{y}_{s-1})} = \sum_{j=0}^{s-T_c-1} (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^j \boldsymbol{\pi}^* + (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{s-T_c} \sum_{j=0}^{T_c-2} (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^j \boldsymbol{\pi} \quad (9)$$

and

$$\begin{aligned} \mathbb{V}_e(\boldsymbol{y}_{s-1}) = & \sum_{j=0}^{s-T_c-1} (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^j \boldsymbol{\Sigma}^* (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{j\prime} + \\ & + (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{s-T_c} \mathbb{V}_e(\boldsymbol{y}_{T_c-1}) (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{s-T_c\prime} \end{aligned} \quad (10)$$

where $\mathbb{V}_e(\boldsymbol{y}_{T_c-1})$ is given by (8), where s is replaced by T_c.

Proof. *Appendix A.1* □

The expressions derived in Lemma 1 are useful in the following subsection (Appendix A.2).

3.1.2 Estimation

Estimation of $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ can be carried out by Reduced Rank Regression (RRR) as in Johansen (1988, 1991).[6] By separately stacking $\boldsymbol{y}_{t-1}$ and $\Delta\boldsymbol{y}_t$ over t (Pesaran & Timmermann (2005)), we obtain the following lemma.

Lemma 2. *Under Assumptions 1-3, the estimators for the adjustment matrix $\hat{\boldsymbol{\gamma}}(s)$ and the constant $\hat{\boldsymbol{\pi}}(s)$ are given by*

$$\begin{aligned} \underset{N\times r}{\hat{\boldsymbol{\gamma}}(s)} = & l^{-1}(s) \cdot (vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s) + \tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))(vec^{-1}(\tilde{\boldsymbol{c}} + \tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s)))' \\ & - l^{-1}(s) \cdot vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s) + \tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))\boldsymbol{1}_{l(s)}\boldsymbol{1}'_{l(s)}(vec^{-1}(\tilde{\boldsymbol{c}} + \tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s)))')\hat{\boldsymbol{\delta}}(s) \end{aligned} \quad (11)$$

[6] Hansen (2018) shows that normality is not necessary to motivate these estimators.

and

$$\underset{N\times 1}{\hat{\boldsymbol{\pi}}(s)} = l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s)+\tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)} - \hat{\gamma}(s)\hat{\boldsymbol{\delta}}'(s)vec^{-1}(\tilde{\mathbf{c}}+\tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)}) \tag{12}$$

where $\boldsymbol{R}$, $\tilde{\boldsymbol{P}}$, $\tilde{\mathbf{c}}$ *and* $\tilde{\boldsymbol{D}}$ *are matrices whose dimensions depend on* s *(Appendix A.2),* $\boldsymbol{\chi}_{0T}(s) = (\boldsymbol{\chi}_s, \dots, \boldsymbol{\chi}_T)' \overset{w.n.}{\sim} \mathcal{D}(\mathbf{0}, \boldsymbol{I}_{l(s)N})$ *and* $\boldsymbol{\chi}_{1T}(s) = (\boldsymbol{\chi}_{s-1}, \dots, \boldsymbol{\chi}_{T-1})' \overset{w.n.}{\sim} \mathcal{D}(\mathbf{0}, \boldsymbol{I}_{l(s)N})$ *are the standardized errors,* $\mathbf{1}_{l(s)}$ *is the unit vector of dimension* $(l(s)\times 1)$*,* $l(s) = l_{pr}(s) + l_{po} = T - s + 1$ *is the length of the estimation window, where, in turn,* $l_{pr}(s) = T_c - s$ *and* $l_{po} = T - T_c + 1$ *are the lengths of the pre- and post-break windows, resp.,* $\cdot$ *is the scalar multiplication,* $vec()$ *is the vectorization operator with respect to the time dimension and* $\hat{\boldsymbol{\delta}}(s)$ *are estimated eigenvectors.*[7]

Proof. *Appendix A.2* □

The eigenvalue problem to find the RRR estimator $\hat{\boldsymbol{\delta}}(s)$ (Appendix A.2) is unchanged, but can be seen to depend on the expressions derived in Appendix A.2, i.e. to be sensitive to the instability in the parameters through the product moment matrices, leading to a change in the estimated eigenvectors. Hence, the eigenvalue problem, (11) and (12) in Lemma 2 depend on all parameters $\boldsymbol{\pi}$, $\boldsymbol{\pi}^*$, $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^*$, $\boldsymbol{\delta}$, $\boldsymbol{\delta}^*$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^*$.

[7] The constant can be estimated jointly with the cointegration space, however to derive and show the usefulfuness of the following decomposition we still estimate it unrestrictedly.

3.2 Forecast error decomposition

Proposition 1. *Using Lemma 2, the one-step ahead forecast errors* $\hat{\boldsymbol{\xi}}_{T+1|T}(s)$ *can be decomposed as*

$$\underset{N\times 1}{\hat{\boldsymbol{\xi}}_{T+1|T}}(s) = \boldsymbol{e}^*_{T+1} - \boldsymbol{A}_{1T}(s) - \boldsymbol{A}_{2T}(s) - \boldsymbol{A}_{3T}(s) \tag{13}$$

where

$$\underset{N\times 1}{\boldsymbol{A}_{1T}}(s) = \hat{\boldsymbol{\gamma}}(s)(\hat{\boldsymbol{\delta}}'(s)\mathbb{E}_e(\boldsymbol{y}_T) - l^{-1}(s)\cdot\hat{\boldsymbol{\delta}}'(s)vec^{-1}(\tilde{\boldsymbol{c}})\mathbf{1}_{l(s)}), \tag{14}$$

$$\underset{N\times 1}{\boldsymbol{A}_{2T}}(s) = l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s)+\tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)}-\hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)vec^{-1}(\tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)}), \tag{15}$$

$$\underset{N\times 1}{\boldsymbol{A}_{3T}}(s) = (\hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s) - \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{y}_T - \mathbb{E}_e(\boldsymbol{y}_T)). \tag{16}$$

Proof. *Appendix A.3* □

The four components of (13) represent: an uncorrelated, zero-mean period ahead risk subject to a break in the covariance structure; a bias due to a conditional mean shift (14); a distortion provoked by the gap between the variation in the differenced process and the variation in the cointegration relations (15); and a bias, valid only conditionally (in general $\boldsymbol{y}_T \neq \mathbb{E}_e(\boldsymbol{y}_T)$), which depends on the distance between estimated and true adjustment and cointegration matrices in the new regime, as well as on the gap between the realized value of the level process and its expected value (16). These expressions generalize Pesaran & Timmermann (2005), eq. (24)-(27), p.193, and can be used to derive simulation-based estimators for the finite sample forecast properties.

3.3 Simulation-based estimators

Proposition 1 can be used to derive simulation-based estimators for the bias and MSE. Let K be the number of repetitions and $i \in \{1, \ldots, K\}$ be the repetition's number. Formally, equation (5) can be simulated by $\hat{\mathbb{E}}_e(\hat{\boldsymbol{\xi}}_{T+1|T}(s)|\boldsymbol{\Omega}_T) \approx \frac{1}{K}\sum_{i=1}^{K} \hat{\boldsymbol{\xi}}^{(i)}_{T+1|T}(s)$, and, similarly, equation (6) can be simulated by $\hat{\mathbb{E}}_e(\hat{\boldsymbol{\xi}}_{T+1|T}(s)\hat{\boldsymbol{\xi}}'_{T+1|T}(s)|\boldsymbol{\Omega}_T) \approx \frac{1}{K}\sum_{i=1}^{K} \hat{\boldsymbol{\xi}}^{(i)}_{T+1|T}(s)\hat{\boldsymbol{\xi}}^{(i)\prime}{}_{T+1|T}(s)$. We propose the following estimators.

Corollary 1. *Using Proposition 1, the simulation-based estimator for the conditonal bias can be obtained as*

$$\hat{\mathbb{E}}_e(\underset{N\times 1}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)}|\boldsymbol{\Omega}_T) = -\frac{1}{K}\sum_{i=1}^{K}(\boldsymbol{A}^{(i)}_{1T}(s) + \boldsymbol{A}^{(i)}_{2T}(s) + \boldsymbol{A}^{(i)}_{3T}(s)) \tag{17}$$

and the simulation-based estimator for the conditional MSE as

$$\hat{\mathbb{E}}_e(\underset{N\times N}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)\hat{\boldsymbol{\xi}}'_{T+1|T}(s)}|\boldsymbol{\Omega}_T) = \boldsymbol{\Sigma}^* + \frac{1}{K}\sum_{i=1}^{K}(\boldsymbol{A}^{(i)}_{1T}(s) + \boldsymbol{A}^{(i)}_{2T}(s) + \boldsymbol{A}^{(i)}_{3T}(s))^2 \tag{18}$$

where (17) and (18) converge in probability as $\hat{\mathbb{E}}_e(\hat{\boldsymbol{\xi}}_{T+1|T}(s)|\boldsymbol{\Omega}_T) \xrightarrow{p} \mathbb{E}_e(\hat{\boldsymbol{\xi}}_{T+1|T}|\boldsymbol{\Omega}_T)$ and $\hat{\mathbb{E}}_e(\hat{\boldsymbol{\xi}}_{T+1|T}(s)\hat{\boldsymbol{\xi}}'_{T+1|T}(s)|\boldsymbol{\Omega}_T) \xrightarrow{p} \mathbb{E}_e(\hat{\boldsymbol{\xi}}_{T+1|T}\hat{\boldsymbol{\xi}}'_{T+1|T}|\boldsymbol{\Omega}_T)$, resp., by the weak law of large numbers thanks to uncorrelatedness of $\hat{\boldsymbol{\xi}}^{(i)}_{T+1|T}$ across repetitions, where $\xrightarrow{p}$ denotes convergence in probability.

4 Window length selection

4.1 Simulation design

We consider $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})' \sim CI(1,1)$ to be 2-dim, i.e. $N = 2$.[8] Assumptions 1-3 hold true. Without further parameter restrictions, the model is given by

$$\underset{2\times 1}{\Delta \boldsymbol{y}_T} = \begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} -\gamma_1 \rho_1 \\ -\gamma_2 \rho_1 \end{bmatrix} + \begin{bmatrix} \gamma_1 & -\gamma_1 \theta_2 \\ \gamma_2 & -\gamma_2 \theta_2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \tag{19}$$

where the constant, restricted to the cointegration space, is given by $\boldsymbol{\pi} = -(\gamma_1\rho_1, \gamma_2\rho_1)'$, the adjustment vector is given by $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ and the (normalized) cointegration vector is given by $\boldsymbol{\delta} = (1, -\theta_2)'$. For simplicity, we impose stricter conditions on the errors, which are given by $\boldsymbol{e}_t = (e_{1,t}, e_{2,t})' \overset{i}{\sim} \mathcal{N}(\mathbf{0}_2, \boldsymbol{\Sigma})$, where $\overset{i}{\sim}$ stands for 'independently distributed' and $\boldsymbol{\Sigma} = (\sigma_1^2, 0 : 0, \sigma_2^2)$. Moreover, we constrain the following parameters: $\gamma_1 = -1$ and $\gamma_2 = 0$, so that $\boldsymbol{\gamma} = (-1, 0)'$. The model becomes

$$\underset{2\times 1}{\Delta \boldsymbol{y}_T} = \begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 & \theta_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \tag{20}$$

where now the constant is given by $\boldsymbol{\pi} = (\rho_1, 0)'$. In single-equation form, $y_{1,t} = \rho_1 + \theta_2 y_{2,t-1} + e_{1,t}$, implying that $y_{1,t}$ is purely adjusting to $y_{2,t}$, or endogenous, as it can be seen from $\boldsymbol{\gamma}$, which is a unit vector, and $y_{2,t} = y_{2,t-1} + e_{2,t}$ is an exogenous random walk.

We report our experiments with associated parameter values in Table

[8]Bi-dimensionality is chosen to insulate from identification issues.

1 (Appendix B). In all experiments we keep $\boldsymbol{\gamma} = (-1, 0)'$. Experiment 1 involves model (20), i.e. a model without breaks (baseline), where $\theta_2 = 0.5$, $\rho_1 = 2$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 1$. Experiments $2 - 9$ vary according to the location (in the parameter space) and direction of the break. Experiments $2-5$ consider a break in the conditional mean. In particular, in experiment 2, $\rho_1^* = 3.5$ so that $\boldsymbol{\pi}^* = (3.5, 0)'$, and $\boldsymbol{\delta}^* = (1, -0.1)'$, since $\theta_2^* = 0.1$ (weaker cointegration). In experiment 3, $\rho_1^* = 3.5$ so that $\boldsymbol{\pi}^* = (3.5, 0)'$ as before, and $\boldsymbol{\delta}^* = (1, -0.9)'$, since $\theta_2^* = 0.9$ (stronger cointegration). Experiment 4 considers $\rho_1^* = 0.5$ so that $\boldsymbol{\pi}^* = (0.5, 0)'$, and $\boldsymbol{\delta}^* = (1, -0.1)'$, since $\theta_2^* = 0.1$ (weaker cointegration). Experiment 5 still considers $\rho_1^* = 0.5$ so that $\boldsymbol{\pi}^* = (0.5, 0)'$, and $\boldsymbol{\delta}^* = (1, -0.9)'$, since $\theta_2^* = 0.9$ (stronger cointegration). Experiments $6-9$ consider a break in the conditional variance. In particular, in experiment 6, $\sigma_1^* = 1.5$ and $\sigma_2^* = 0.5$, so that $\boldsymbol{\Sigma}^* = (2.25, 0 : 0, 0.25)$. Symmetrically, in experiment 7, $\sigma_1^* = 0.5$ and $\sigma_2^* = 1.5$, so that $\boldsymbol{\Sigma}^* = (0.25, 0 : 0, 2.25)$. Experiment 8 considers a joint decrease in the univariate variances $\sigma_1^* = 0.5$ and $\sigma_2^* = 0.5$ so that $\boldsymbol{\Sigma}^* = (0.25, 0 : 0, 0.25)$. Finally, experiment 9 considers a joint increase in the univariate variances $\sigma_1^* = 1.5$ and $\sigma_2^* = 1.5$ so that $\boldsymbol{\Sigma}^* = (2.25, 0 : 0, 2.25)$.

We estimate the parameters by RRR, obtaining $\hat{\boldsymbol{\pi}}_T(s)$, $\hat{\boldsymbol{\gamma}}_T(s)$ and $\hat{\boldsymbol{\delta}}_T(s)$, and compute the one-step ahead conditional forecast errors $\hat{\boldsymbol{\xi}}_{T+1|T}(s)$ as in Proposition 1, equations (13)-(16). The estimators in Lemma 2 are not constrained by the terminal value of the process $\boldsymbol{y}_t$, therefore, to compute the finite sample forecast properties, we condition on a particular realization of $\boldsymbol{y}_T$, which we fix, and we (numerically) integrate with respect to the error distribution. As in (Pesaran & Timmermann, 2005), parameter estimation uncertainty is due to independent (from $\boldsymbol{y}_T$) draws from the error distribution.[9] We run this procedure in a Monte Carlo simulation with $K = 10^3$ replications and get the finite sample properties (conditional bias and MSE) of $\hat{\boldsymbol{\xi}}_{T+1|T}(s)$ as in Corollary 1, eq. (17)-(18). For each experiment, we carry out simulations along two additional dimensions, i.e. s and T_c, or, equivalently, $l_{pr}(s)$ and l_{po}, and let T be fixed ($T = 202$), so that $l_{po} \in \{10, 20, 30, 50, 100\}$ and $l_{pr}(s) \in \{0, 1, 2, 3, 4, 5, 10, 20, 30, 50, 100\}$. Univariate measures for (17) and (18) are then reported for each value of the pre- and post-break window length.

4.2 Results

We report results numerically (Appendix B).

4.2.1 Conditional MSE

Table 2 reports the conditional root trace MSE (RTMSE).

In the case of no break (experiment 1) the RTMSE monotonically decreases with expanding estimation window and depends exclusively on its full length $l(s)$ and not on its single components $l_{pr}(s)$ and l_{po}. In other words, the RTMSE is minimal for maximal $l(s)$ and tends to its theoretical value, i.e. in the absence of parameter estimation uncertainty, $\sqrt{tr(\boldsymbol{\Sigma})} = \sqrt{2}$. This result, in line with Pesaran & Timmermann (2005), is due to the fact that the forecast error variance and the square bias jointly decrease as we add observations to estimate the parameters.[10]

When we introduce a break in θ_2 and ρ_1, and accordingly $\boldsymbol{\delta}$ and $\boldsymbol{\pi}$, (experiment 2-5) the RTMSE needs not monotonically decrease with ex-

[9] We perform simulations remotely via parallel computing in *R 3.5.0*. We set the seeds for the random number generator across experiments and window lengths to ensure reproducibility and comparability. Codes are available upon request.

[10] The (trace of the) forecast error variance can be derived as a reminder in the simulation.

panding window. Although increasing the post-break window (l_{po}) leads to an improvement in forecasting performance, this improvement does not hold when we extend the pre-break window ($l_{pr}(s)$) above the zero length, at least monotonically, mirroring Pesaran & Timmermann (2005)'s finding.

We now discuss the behaviour of the RTMSE for extreme values of the post-break window length as the pre-break window varies, by experiment. In experiment 2 (opposite direction for the breaks: mean up and weaker cointegration), when $l_{po} = 10$ (i.e. when the break is relatively recent) the forecasting performance reaches its lowest quality (RTMSE = 2.807) if we include five pre-break observations. Nevertheless, when we overcome that level the RTMSE decreases monotonically. We have an improvement in performance once we reach $l_{pr}(s) = 20$ against not including any pre-break information at all. The optimal estimation window is an expanding window in this case, where $l_{pr}(s)$ is maximal. For $l_{po} = 100$ (i.e. when the perturbation is relatively old), the RTMSE reaches its highest value (RTMSE = 1.607) if we include thirty pre-break observations, for then experiencing a monotonic decline. However, in this case the optimal estimation window implies $l_{pr}(s) = 3$ (still not null) and no longer an expanding window. In experiment 3 (same upward direction for the breaks), for the shortest l_{po}, the worst performance (RTMSE = 2.877) is reached when we include four pre-break data. After then, the decline is monotonic and the optimal window is expanding, too. For $l_{po} = 100$, the RTMSE reaches its highest value (RTMSE = 1.605) if we include thirty pre-break observations and the optimal window implies $l_{pr}(s) = 2$ (still not null) and no longer an expanding window. In experiment 4 (same downward direction for the breaks), when l_{po} is shortest, the worst performance (RTMSE = 2.753) is reached when we include four pre-break data. After then, the decline is monotonic and the optimal window is expanding, once again. For $l_{po} = 100$, the RTMSE reaches its highest value (RTMSE = 1.624) if we include fifty pre-break observations and the optimal window implies $l_{pr}(s) = 3$ (still not null) and no longer an expanding window. Finally, in experiment 5 (opposite direction for the breaks: mean down and stronger cointegration), when $l_{po} = 10$, the worst performance (RTMSE = 3.030) is reached when we include five pre-break data. After then, the decline is once again monotonic and the optimal window is expanding. For the longest l_{po}, the RTMSE reaches its highest value (RTMSE = 1.819) if we include fifty pre-break observations, but the optimal window implies $l_{pr}(s) = 0$ in this case, i.e. it is best to exclude pre-break data. In all cases, moreover, we are not sure whether the RTMSE still converge to its theoretical value.

In sum, we find that the timing of the break is crucial for determining the optimal window. When the break is recent, an expanding window can be adopted. When the break is old, a window containing few or none pre-break data can be selected. Post-break windows are optimal only in special cases, contrarily to what is assumed in practice. These results are driven by the bias-efficiency and the within bias trade-offs, which depend on the timing, but also the location and direction of the breaks.[11] The joint dynamics of the bias and variance of forecasts under breaks is what eventually matters for forecasting performance.

To shed light on the role of the location of the breaks, we now analyze breaks in the variance. When we introduce a break in σ_1 and σ_2 (experiment 6-9) the RTMSE needs not monotonically decrease with expanding window. Although increasing l_{po} leads to an improvement in accuracy, this improvement does not always hold when we extend $l_{pr}(s)$, once again.

[11] The size of the breaks, not investigated here, is also a key factor in these trade-offs.

In experiment 6 (opposite direction of the break: variance of the cointegration relation up, variance of the random walk down), expanding $l_{pr}(s)$ at its maximum length ($l_{pr}(s) = 100$) turns out to be the best strategy regardless of the value of l_{po}. The same conclusion holds in case of experiment 7 (opposite direction of the break: variance of the cointegration relation down, variance of the random walk up). Hence, in both cases the breaks balance each other and the longest window is optimal eventually. In experiment 8 (same downward direction for the breaks), however, this is no longer true. When l_{po} is shortest, the worst performance (RTMSE $= 1.232$) is reached when we include two pre-break data. After then, the decline is monotonic and the optimal window is expanding. When $l_{po} = 100$, however, the RTMSE reaches its lowest value if we do not include any pre-break observation at all ($l_{pr}(s) = 0$). The expanding window estimator is no longer optimal. This is intuitive since expanding the window backward amounts to introduce more noise. Finally, in experiment 9 (same upward direction for the breaks), including all pre-break data assures the best performance. In this case, instead, extending the window backward reduces the noise. In all cases, we are not sure whether the RTMSE still converge to its theoretical value, which, for the case of a break in variance, it is a varying quantity dependent on the size of the break.

4.2.2 Conditional bias

Table 3 reports the conditional bias. In Table 4 and 5 we report biases in estimating the parameters suffering from a break. This amounts to focus on the estimators for $-\gamma_1\theta_2$ and $-\gamma_1\rho_1$ in (19), whose true values are θ_2 and ρ_1 as displayed in (20). Estimation biases are useful for clarifying the role of the direction of the break.

Experiment 1 (no break) shows that the bias is rather small even for short samples and decreases with increasing sample size. The small size of the bias is due to the well known fact that the cointegration vector is estimated superconsistently.

The introduction of a break in mean induces a substantial increase in the bias. The dynamics of the bias is non-trivial since it depends on the estimation bias as well as the bias introduced by the breaks.

Experiment 2 displays a non-decreasing dynamics for the bias along the pre-break window length. However, experiment 5 shows that the bias tends to increase with expanding pre-break window. When the estimation bias and the bias induced by a break go in the opposite direction, including pre-break information leads to forecasts that are more distorted. This is the reason why in experiment 5 the bias increases with expanding pre-break window. In detail, the estimator for $-\gamma_1\theta_2$ is upward biased (Table 4) and when θ_2 declines (experiment 2) the bias increases even further. Therefore, the inclusion of pre-break data, when the value of θ_2 is higher, leads to a reduction in the overall bias. The estimator for $-\gamma_1\rho_1$ (Table 5) is upward biased, too, however when ρ_1 increases expanding the window backward leads to an increase in bias. The two effects balance each other and the bias does not dissipate eventually. On the other hand, an increasing θ_2 (experiment 5) leads to the opposite phenomenon when we expand the window prior to the break. The new bias which follows adds to the already positive estimation bias and the overall bias is increased. This effect is weakened due to the negative direction of the break in ρ_1, this time. However, the net effect is dominated by the break in θ_2. The final effect is that the bias increases with expanding pre-break window. Interestingly, when the breaks go in the same direction, the dynamics of the bias is non-monotonic. In Experiment 3 and

4, the dynamics of the bias appears to be U-shaped, where the minimum coincides with increasingly long pre-break window lengths as we increase the post-break window.

Instead, breaks in the variance (experiment 6-9) do not materially affect the magnitude of the bias. This result is expected since changes in the variation of the process should not bias the forecasts but influence their variance.

In sum, we shed light on the importance of the direction of the breaks. If breaks happen to be in the mean, then forecasts can suffer from a newly introduced bias by the expansion of the window backward, especially if the estimation bias and the break bias go in the opposite direction. If breaks affect the variance, then forecast efficiency is affected, but its bias is unaffected. When the variance decreases, forecasts are damaged due to the greater risk surrounding them as we expand $l_{pr}(s)$, as shown previously.

4.2.3 Comparison with the univariate case

We assess how important is the multivariate framework with cointegration for the relation between the finite sample properties of forecasts (in particular the MSE) and the window length. We do so by comparing the RTMSE growth rates among window lengths for our experiments vs the ones computed for the autoregressive process (Pesaran & Timmermann (2005), Table 5, p.202). We plot the growth rates against intervals of $l_{pr}(s)$, which can possibly include a break. We do so for the shortest post-break window ($l_{po} = 10$), where differences are starker. Figure 1 reports this comparison by experiment (when comparable). In experiment 1 (no break), for the VEC(1) case there exists larger gains in expanding the window backward for short pre-break windows, while the opposite is true for longer pre-break windows. Indeed, after the interval 10-20, the AR(1) case experiences a faster decrease in the RTMSE. In experiment 3 and 5 (vs 4 in Pesaran & Timmermann (2005)), the break in mean causes a more abrupt behaviour of RTMSE growth (first increase, then decrease, finally increase again). Augmenting the window backward in the VEC(1) case becomes beneficial only after the 4-5 interval. For shorter intervals, expanding the pre-break window is clearly more convenient in the AR(1) case. In experiment 8 (labeled 7 in Pesaran & Timmermann (2005)), the dynamics of the RTMSE is almost identical until a pre-break window interval of length 5-10 is reached. After then expanding the window further backward provides more benefit in the AR(1) case. In experiment 9 (named 6 in Pesaran & Timmermann (2005)), the RTMSE dynamics is also almost identical but with a gap in favour of the VEC(1) case. Expanding the window backward gives more advantage in that case compared to the AR(1) case. Only for very large pre-break windows the RTMSE growth rates seem to converge.

5 Application

We test our theoretical findings on a well known relation in finance among interest rates at different maturities. This is relevant to the term structure of interest rates (also known as the yield curve), a curve suspected to be associated with likelihood of recessions. We focus on the US T-bond interest rates in the period 1970M01-1987M02 as done by Hansen & Johansen (1999), who find evidence of parameter instability, confirming the findings of Hall *et al.* (1992). Figure 2 reports interest rates at four maturities (1M, 2M, 3M and 4M) as well as the periods of identified turbulence by Hall *et al.* (1992) (vertical lines). The four rates are clearly non-stationary series of the unit root type, but are found to cointegrate through their spreads. Indeed, Hansen & Johansen (1999) find three cointegration relations among the rates, which therefore share a common stochastic trend. We use the earliest break point as a starting point for our sample split, since this seems to be the most serious one as found by the joint fluctuation test of the eigenvalues (Hansen & Johansen (1999), bottom-right plot in Fig. 3, on p.319) and also by a visual inspection of Figure 2. We estimate the parameters for each window length by imposing a rank of three for the long run matrix and a restricted constant. We retain one test observation in order to compute the one-step ahead forecast errors, which are calculated using the estimated parameters. Figure 3 reports, by window length, the RTSE conditional on the sample observed. It is clear that extending the window backward is beneficial in this case, even if a break is present in the model parameters. The need of including pre-break data is higher, the shorter the post-break window. Indeed, for a post-break window of twenty observations the longest pre-break window is actually optimal. However, for a post-break window of eighty data points, this is no longer true: the optimal pre-break window length is two. For intermediate post-break window lengths, we observe a reversed ranking in the RTSE. We argue that this is due to the approaching second instability as identified by Hall *et al.* (1992). Indeed, this instabilty is exactly located between these two post-break window lengths, and causes the increase in the RTSE for the longer post-break window ($l_{po} = 60$). For the class of breaks found in this illustration, expanding the estimation window backward is generally beneficial.

6 Conclusion

We assess the sensitivity of the finite sample properties of forecasts (conditional bias and MSE) to the starting point of the estimation window when the VEC model (Johansen (1988, 1991)) experiences breaks. For the type of breaks studied in this paper, we show that extending the estimation window prior to structural breaks in VEC models can be beneficial for forecasting performance and highlight under which conditions.

We base our simulation on a generalized Pesaran & Timmermann (2005)'s forecast error decomposition, which we derive, and show that it depends on four components: 1) a period ahead risk; 2) a bias due to a conditional mean shift; 3) a bias due to a variance mismatch; 4) a gap term valid only conditionally. We also derive new expressions for the estimators of the adjustment matrix and a constant, which are auxiliary to the decomposition. Finally, we introduce new simulation-based estimators for the finite sample forecast properties which are based on the derived decomposition.

Our finding points out that, under the derived conditions, we can neglect parameter instability by extending the window backward and be insured against higher forecast risk under this model class as well, generalizing Pesaran & Timmermann (2005)'s result. Our result suggests, in some cases, a reduced need of break-correction strategies and also gives renewed importance to testing for structural breaks, in order to distinguish cases when break-neglection is (not) appropriate. We draw conclusions on the window length which ensures best forecasting performance. In general, these conclusions are shown to depend on the timing, location and direction of the break. We illustrate the theoretical results in an application based on interest rates at different maturities. Strategies based on excluding contaminated observations are not always justified from a forecasting viewpoint, as also recently shown by Boot & Pick (2019).

Our work can be extended in several directions. For example, the forecast horizon can be generalized to $h > 1$, point forecasting can be extended to density forecasting and prediction can be performed unconditionally. Moreover, several approaches can be adopted to improve on forecasting under breaks in the multivariate case as well. For instance, improved prediction performance under instability can be achieved via forecast combination across estimation windows (Pesaran & Pick (2011)) and via exponential smoothing or robust optimal weights to weigh past observations (Pesaran *et al.* (2013)). We leave these topics for future research.

Appendix

A Proofs

A.1 Proof of Lemma 1

Thanks to Assumption 3, by recursive substitution, $\boldsymbol{y}_{s-1}$ is given by, for $s-1 < T_c$,

$$\underset{N\times 1}{\boldsymbol{y}_{s-1}} = \boldsymbol{\pi} + (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')\boldsymbol{y}_{s-2} + \boldsymbol{e}_{s-1} =$$
$$= \boldsymbol{\pi} + (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')(\boldsymbol{\pi} + (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')\boldsymbol{y}_{s-3} + \boldsymbol{e}_{s-2}) + \boldsymbol{e}_{s-1} =$$
$$= \cdots = \sum_{j=0}^{s-2}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^j(\boldsymbol{\pi} + \boldsymbol{e}_{s-1-j}).$$

Given Assumption 1, (iv), its expectation is given by (7) and its covariance matrix by (8) in Lemma 1. When $s-1 \geq T_c$, thanks to Assumption 3, $\boldsymbol{y}_{s-1}$ is given by

$$\underset{N\times 1}{\boldsymbol{y}_{s-1}} = \boldsymbol{\pi}^* + (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})\boldsymbol{y}_{s-2} + \boldsymbol{e}^*_{s-1} =$$

$$= \cdots = \sum_{j=0}^{s-T_c-1} (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^j(\boldsymbol{\pi}^* + \boldsymbol{e}^*_{s-1-j}) + (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{s-T_c}\boldsymbol{y}_{T_c-1}$$

where

$$\underset{N\times 1}{\boldsymbol{y}_{T_c-1}} = \sum_{j=0}^{T_c-2}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}^{\prime})^j(\boldsymbol{\pi} + \boldsymbol{e}_{T_c-1-j}).$$

Given Assumption 1, (iv), its expectation is given by (9) and its covariance matrix by (10) in Lemma 1.

A.2 Proof of Lemma 2

The Johansen (1988)'s generalized eigenvalue problem is given by $det(\boldsymbol{\Lambda}\boldsymbol{S}_{11}(s) - \boldsymbol{S}_{10}(s)\boldsymbol{S}_{00}^{-1}(s)\boldsymbol{S}_{01}(s)) = 0$, where $det()$ is the determinant of a matrix, $\boldsymbol{S}_{ij}(s) = \boldsymbol{M}_{ij}(s) - \boldsymbol{M}_{i2}(s)\boldsymbol{M}_{22}^{-1}(s)\boldsymbol{M}_{2j}(s)$, $i,j \in \{0,1\}$, where, in turn, $\boldsymbol{M}_{ij}(s) = l^{-1}(s)\cdot\sum_{t=s}^{T}\boldsymbol{z}_{it}\boldsymbol{z}^{\prime}_{jt}$, $i,j \in \{0,1,2\}$, $\cdot$ is the scalar multiplication operator, $\boldsymbol{z}_{0t} = \Delta\boldsymbol{y}_t$, $\boldsymbol{z}_{1t} = \boldsymbol{y}_{t-1}$ and $z_{2t} = 1$. Accordingly, $\boldsymbol{M}_{22}(s) \coloneqq m_{22} = 1$, which implies its inverse to be unity and therefore $\boldsymbol{S}_{ij}(s) = \boldsymbol{M}_{ij}(s) - \boldsymbol{M}_{i2}(s)\boldsymbol{M}_{2j}(s)$. Consistently with our notation, we re-write $\boldsymbol{M}_{i2}(s) \coloneqq \boldsymbol{m}_{i2}(s)$ and $\boldsymbol{M}_{2j}(s) \coloneqq \boldsymbol{m}_{2j}(s)$, $i,j \in \{0,1\}$. The solution to this eigenvalue problem yields the RRR estimator for $\boldsymbol{\delta}$, that is

$$\underset{N\times r}{\hat{\boldsymbol{\delta}}(s)} = (\hat{\boldsymbol{v}}_1(s), \ldots, \hat{\boldsymbol{v}}_r(s))$$

where the $\hat{\boldsymbol{v}}$'s are estimated eigenvectors associated with the r largest eigenvalues. These eigenvalues are collected in the estimated $(r\times r)$-dim eigenvalue matrix given by $\hat{\boldsymbol{\Lambda}}(s) = \hat{\boldsymbol{\delta}}'(s)\boldsymbol{S}_{10}(s)\boldsymbol{S}_{00}^{-1}(s)\boldsymbol{S}_{01}(s)\hat{\boldsymbol{\delta}}(s)$. The RRR estimator for $\boldsymbol{\gamma}$ is derived accordingly as $\hat{\boldsymbol{\gamma}}(s) = \boldsymbol{S}_{01}(s)\hat{\boldsymbol{\delta}}(s)(\hat{\boldsymbol{\delta}}'(s)\boldsymbol{S}_{11}(s)\hat{\boldsymbol{\delta}}(s))^{-1}$, which, after the normalization $\hat{\boldsymbol{\delta}}'(s)\boldsymbol{S}_{11}(s)\hat{\boldsymbol{\delta}}(s) = \boldsymbol{I}_r$ (Johansen (1988)), simplifies to

$$\underset{N\times r}{\hat{\boldsymbol{\gamma}}(s)} = \boldsymbol{S}_{01}(s)\hat{\boldsymbol{\delta}}(s).$$

Finally, $\boldsymbol{\pi}$, given that $m_{22} = 1$, is estimated as

$$\underset{N\times 1}{\hat{\boldsymbol{\pi}}(s)} = \boldsymbol{m}_{02}(s) - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)\boldsymbol{m}_{12}(s).$$

We can re-write the product moment matrices as $\boldsymbol{M}_{ij}(s) = l^{-1}(s)\cdot\sum_{t=s}^{T}\boldsymbol{z}_{it}\boldsymbol{z}'_{jt} = l^{-1}(s)\cdot vec^{-1}(\boldsymbol{Z}_{iT}(s))(vec^{-1}(\boldsymbol{Z}_{jT}(s)))'$, $i,j \in \{0,1,2\}$, where $\boldsymbol{Z}_{0T}(s) = (\Delta\boldsymbol{y}_s,\ldots,\Delta\boldsymbol{y}_T)'$, $\boldsymbol{Z}_{1T}(s) = (\boldsymbol{y}_{s-1},\ldots,\boldsymbol{y}_{T-1})'$ and $\boldsymbol{Z}_{2T}(s) \coloneqq \boldsymbol{z}_{2T}(s) = \mathbf{1}_{l(s)}$, where $vec^{-1}(\boldsymbol{z}_{2T}(s)) = \mathbf{1}'_{l(s)}$. Hence, the estimator for the adjustment matrix $\hat{\boldsymbol{\gamma}}(s)$ becomes

$$\begin{aligned}\underset{N\times r}{\hat{\boldsymbol{\gamma}}(s)} &= \boldsymbol{S}_{01}(s)\hat{\boldsymbol{\delta}}(s) = (\boldsymbol{M}_{01}(s) - \boldsymbol{m}_{02}(s)\boldsymbol{m}_{21}(s))\hat{\boldsymbol{\delta}}(s) = \\ &= l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{Z}_{0T}(s))(vec^{-1}(\boldsymbol{Z}_{1T}(s)))' - \\ &- l^{-1}(s)\cdot vec^{-1}(\boldsymbol{Z}_{0T}(s))\mathbf{1}_{l(s)}\mathbf{1}'_{l(s)}(vec^{-1}(\boldsymbol{Z}_{1T}(s)))')\hat{\boldsymbol{\delta}}(s).\end{aligned}$$

The estimator for the constant $\boldsymbol{\pi}$ becomes

$$\begin{aligned}\underset{N\times 1}{\hat{\boldsymbol{\pi}}(s)} &= \boldsymbol{m}_{02}(s) - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)\boldsymbol{m}_{12}(s) = \\ &= l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{Z}_{0T}(s))\mathbf{1}_{l(s)} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)vec^{-1}(\boldsymbol{Z}_{1T}(s))\mathbf{1}_{l(s)}).\end{aligned}$$

To obtain the estimators as functions of the innovations, an explicit expression for $\boldsymbol{Z}_{1T}(s)$ and $\boldsymbol{Z}_{0T}(s)$ in terms of the errors is needed. Let $\boldsymbol{\chi}_{1T}(s) =$

$(\boldsymbol{\chi}_{s-1}, \ldots, \boldsymbol{\chi}_{T-1})' \overset{w.n.}{\sim} \mathcal{D}(\mathbf{0}, \boldsymbol{I}_{l(s)N})$ be the standardized form of $(\boldsymbol{e}_{s-1}, \ldots, \boldsymbol{e}_{T-1})'$, then

$$\underset{l(s)N\times 1}{\boldsymbol{B}\boldsymbol{Z}_{1T}(s)} = \boldsymbol{c} + \boldsymbol{D}\boldsymbol{\chi}_{1T}(s)$$

where

$$\underset{l(s)N\times l(s)N}{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{I}_N & \mathbf{0}_{N\times l_{pr}(s)N} & \mathbf{0}_{N\times(l_{po}-1)N} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \mathbf{0}_{l_{pr}(s)N\times(l_{po}-1)N} \\ \mathbf{0}_{(l_{po}-1)N\times N} & \boldsymbol{B}_{32} & \boldsymbol{B}_{33} \end{bmatrix},$$

$$\underset{l(s)N\times 1}{\boldsymbol{c}} = \begin{bmatrix} \sum_{j=0}^{s-2}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^j \boldsymbol{\pi} \\ \mathbf{1}_{l_{pr}(s)} \otimes \boldsymbol{\pi} \\ \mathbf{1}_{l_{po}-1} \otimes \boldsymbol{\pi}^* \end{bmatrix},$$

$$\underset{l(s)N\times l(s)N}{\boldsymbol{D}} = \begin{bmatrix} Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} & \mathbf{0}_{N\times l_{pr}(s)N} & \mathbf{0}_{N\times(l_{po}-1)N} \\ \mathbf{0}_{l_{pr}(s)N\times N} & \boldsymbol{I}_{l_{pr}(s)} \otimes \boldsymbol{\Sigma}^{1/2} & \mathbf{0}_{l_{pr}(s)N\times(l_{po}-1)N} \\ \mathbf{0}_{(l_{po}-1)N\times N} & \mathbf{0}_{(l_{po}-1)N\times l_{pr}(s)N} & \boldsymbol{I}_{l_{po}-1} \otimes (\boldsymbol{\Sigma}^*)^{1/2} \end{bmatrix},$$

where, in turn, $Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))$ is a triangular Cholesky factor of $\mathbb{V}_e(\boldsymbol{y}_{s-1})$ derived in (8) and (10) in Lemma 1 ($Ch()$ being the Cholesky factorization operator), $\otimes$ is the Kronecker product and $\boldsymbol{B}_{ij}$, $i \in \{2,3\}$, $j \in \{1,2,3\}$ are defined as

$$\underset{l_{pr}(s)N\times N}{\boldsymbol{B}_{21}} = \begin{bmatrix} -(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}') \\ \mathbf{0}_{N\times N} \\ \vdots \\ \mathbf{0}_{N\times N} \end{bmatrix},$$

$$\underset{l_{pr}(s)N\times l_{pr}(s)N}{\boldsymbol{B}_{22}} = \begin{bmatrix} \boldsymbol{I}_N & \mathbf{0}_{N\times N} & \dots & \dots & \mathbf{0}_{N\times N} \\ -(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}') & \boldsymbol{I}_N & \dots & \dots & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & -(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}') & \boldsymbol{I}_N & \dots & \mathbf{0}_{N\times N} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \mathbf{0}_{N\times N} & \dots & \dots & -(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}') & \boldsymbol{I}_N \end{bmatrix},$$

$$\underset{(l_{po}-1)N\times l_{pr}(s)N}{\boldsymbol{B}_{32}} = \begin{bmatrix} \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} & -(\boldsymbol{I}_N+\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}) \\ \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \end{bmatrix},$$

$$\underset{(l_{po}-1)N\times(l_{po}-1)N}{\boldsymbol{B}_{33}} = \begin{bmatrix} \boldsymbol{I}_N & \mathbf{0}_{N\times N} & \dots & \dots & \mathbf{0}_{N\times N} \\ -(\boldsymbol{I}_N+\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}) & \boldsymbol{I}_N & \dots & \dots & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & -(\boldsymbol{I}_N+\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}) & \boldsymbol{I}_N & \dots & \mathbf{0}_{N\times N} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \mathbf{0}_{N\times N} & \dots & \dots & -(\boldsymbol{I}_N+\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}) & \boldsymbol{I}_N \end{bmatrix}.$$

Matrix $\boldsymbol{B}$, being lower triangular and having non-zero, themselves invertible, diagonal elements, is non-singular and therefore we can re-write the stacked equation as

$$\underset{l(s)N\times 1}{\boldsymbol{Z}_{1T}(s)} = \boldsymbol{B}^{-1}(\boldsymbol{c}+\boldsymbol{D}\boldsymbol{\chi}_{1T}(s)) = \tilde{\boldsymbol{c}} + \tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s)$$

where $\tilde{\boldsymbol{c}} \coloneqq \boldsymbol{B}^{-1}\boldsymbol{c}$ and $\tilde{\boldsymbol{D}} \coloneqq \boldsymbol{B}^{-1}\boldsymbol{D}$ are derived as

$$\underset{l(s)N\times 1}{\tilde{\boldsymbol{c}}} = \begin{bmatrix} \sum_{j=0}^{s-2}(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}')^j\boldsymbol{\pi} \\ \vdots \\ \sum_{j=0}^{T_c-2}(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}')^j\boldsymbol{\pi} \\ (\boldsymbol{I}_N+\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})\sum_{j=0}^{T_c-2}(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}')^j\boldsymbol{\pi}+\boldsymbol{\pi}^* \\ \vdots \\ (\boldsymbol{I}_N+\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{T-T_c}\sum_{j=0}^{T_c-2}(\boldsymbol{I}_N+\boldsymbol{\gamma}\boldsymbol{\delta}')^j\boldsymbol{\pi}+\sum_{j=0}^{T-T_c-1}(\boldsymbol{I}_N+\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^j\boldsymbol{\pi}^* \end{bmatrix},$$

$$\underset{l(s)N\times l(s)N}{\tilde{\boldsymbol{D}}} = \begin{bmatrix} Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} & \mathbf{0}_{N\times l_{pr}(s)N} & \mathbf{0}_{N\times(l_{po}-1)N} \\ \tilde{\boldsymbol{D}}_{21} & \tilde{\boldsymbol{D}}_{22} & \mathbf{0}_{l_{pr}(s)N\times(l_{po}-1)N} \\ \tilde{\boldsymbol{D}}_{31} & \tilde{\boldsymbol{D}}_{32} & \tilde{\boldsymbol{D}}_{33} \end{bmatrix},$$

where, in turn, $\tilde{\boldsymbol{D}}_{ij}$, $i \in \{2,3\}$, $j \in \{1,2,3\}$ are given by

$$\underset{l_{pr}(s)N\times N}{\tilde{\boldsymbol{D}}_{21}} = \begin{bmatrix} (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \\ \vdots \\ (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s}Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \end{bmatrix},$$

$$\underset{l_{pr}(s)N\times l_{pr}(s)N}{\tilde{\boldsymbol{D}}_{22}} = \begin{bmatrix} \boldsymbol{\Sigma}^{1/2} & \boldsymbol{0}_{N\times N} & \dots & \boldsymbol{0}_{N\times N} \\ (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')\boldsymbol{\Sigma}^{1/2} & \boldsymbol{\Sigma}^{1/2} & \dots & \boldsymbol{0}_{N\times N} \\ \vdots & \vdots & \ddots & \vdots \\ (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-1}\boldsymbol{\Sigma}^{1/2} & (\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-2}\boldsymbol{\Sigma}^{1/2} & \dots & \boldsymbol{\Sigma}^{1/2} \end{bmatrix},$$

$$\underset{(l_{po}-1)N\times N}{\tilde{\boldsymbol{D}}_{31}} = \begin{bmatrix} (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{l_{pr}(s)}Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \\ \vdots \\ (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{l_{po}-1}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{l_{pr}(s)}Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \end{bmatrix},$$

$$\underset{(l_{po}-1)N\times l_{pr}(s)N}{\tilde{\boldsymbol{D}}_{32}} = \begin{bmatrix} (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{l_{pr}(s)-1}\boldsymbol{\Sigma}^{1/2} & \dots & (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})\boldsymbol{\Sigma}^{1/2} \\ \vdots & \ddots & \vdots \\ (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{l_{po}-1}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{l_{pr}(s)-1}\boldsymbol{\Sigma}^{1/2} & \dots & (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{l_{po}-1}\boldsymbol{\Sigma}^{1/2} \end{bmatrix},$$

$$\underset{(l_{po}-1)N\times(l_{po}-1)N}{\tilde{\boldsymbol{D}}_{33}} = \begin{bmatrix} (\boldsymbol{\Sigma}^*)^{1/2} & \boldsymbol{0}_{N\times N} & \dots & \boldsymbol{0}_{N\times N} \\ (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{\Sigma}^*)^{1/2} & (\boldsymbol{\Sigma}^*)^{1/2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{l_{po}-2}(\boldsymbol{\Sigma}^*)^{1/2} & (\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{l_{po}-3}(\boldsymbol{\Sigma}^*)^{1/2} & \dots & (\boldsymbol{\Sigma}^*)^{1/2} \end{bmatrix}.$$

Similarly, let $\boldsymbol{\chi}_{0T}(s) = (\boldsymbol{\chi}_s, \dots, \boldsymbol{\chi}_T)' \overset{w.n.}{\sim} \mathcal{D}(\boldsymbol{0}, \boldsymbol{I}_{l(s)N})$ be the standardized version of $(\boldsymbol{e}_s, \dots, \boldsymbol{e}_T)'$, then

$$\underset{l(s)N\times 1}{\boldsymbol{Z}_{0T}(s)} = \boldsymbol{P}\boldsymbol{Z}_{1T}(s) + \boldsymbol{q} + \boldsymbol{R}\boldsymbol{\chi}_{0T}(s)$$

where

$$\underset{l(s)N\times l(s)N}{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{P}_{11} & \boldsymbol{0}_{l_{pr}(s)N\times l_{po}N} \\ \boldsymbol{0}_{l_{po}N\times l_{pr}(s)N} & \boldsymbol{P}_{22} \end{bmatrix},$$

$$\underset{l(s)N\times 1}{\boldsymbol{q}} = \begin{bmatrix} \boldsymbol{1}_{l_{pr}(s)} \otimes \boldsymbol{\pi} \\ \boldsymbol{1}_{l_{po}} \otimes \boldsymbol{\pi}^* \end{bmatrix},$$

$$\underset{l(s)N\times l(s)N}{\boldsymbol{R}} = \begin{bmatrix} \boldsymbol{I}_{l_{pr}(s)} \otimes \boldsymbol{\Sigma}^{1/2} & \boldsymbol{0}_{l_{pr}(s)N\times l_{po}N} \\ \boldsymbol{0}_{l_{po}N\times l_{pr}(s)N} & \boldsymbol{I}_{l_{po}} \otimes (\boldsymbol{\Sigma}^*)^{1/2} \end{bmatrix},$$

where, in turn, $\boldsymbol{P}_{ij}$, $i = j \in \{1, 2\}$ are defined as

$$\underset{l_{pr}(s)N \times l_{pr}(s)N}{\boldsymbol{P}_{11}} = \begin{bmatrix} \boldsymbol{\gamma\delta}' & \mathbf{0}_{N\times N} & \cdots & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \boldsymbol{\gamma\delta}' & \cdots & \mathbf{0}_{N\times N} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{N\times N} & \cdots & \cdots & \boldsymbol{\gamma\delta}' \end{bmatrix},$$

$$\underset{l_{po}N \times l_{po}N}{\boldsymbol{P}_{22}} = \begin{bmatrix} \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} & \mathbf{0}_{N\times N} & \cdots & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} & \cdots & \mathbf{0}_{N\times N} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{N\times N} & \cdots & \cdots & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} \end{bmatrix}.$$

Given Assumption 2, (ii), and substituting $\boldsymbol{Z}_{1T}(s)$ into $\boldsymbol{Z}_{0T}(s)$, we obtain

$$\begin{aligned}\underset{l(s)N\times 1}{\boldsymbol{Z}_{0T}(s)} &= \boldsymbol{P}\boldsymbol{Z}_{1T}(s) + \boldsymbol{q} + \boldsymbol{R}\boldsymbol{\chi}_{0T}(s) = \\ &= \boldsymbol{P}(\tilde{\boldsymbol{c}} + \tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s)) + \boldsymbol{q} + \boldsymbol{R}\boldsymbol{\chi}_{0T}(s) = \\ &= \boldsymbol{P}\tilde{\boldsymbol{c}} + \boldsymbol{q} + \boldsymbol{P}\tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s)) + \boldsymbol{R}\boldsymbol{\chi}_{0T}(s).\end{aligned}$$

Calling $\tilde{\boldsymbol{q}} := \boldsymbol{P}\tilde{\boldsymbol{c}} + \boldsymbol{q}$ and $\tilde{\boldsymbol{P}} = \boldsymbol{P}\tilde{\boldsymbol{D}}$, and noting that $\tilde{\boldsymbol{q}} = \mathbf{0}_{l(s)N}$ by (7) and (9) in Lemma 1 and thanks to, for $s < T_c$, $\boldsymbol{\delta}'\mathbb{E}_e(\boldsymbol{y}_{s-1}) = \boldsymbol{\rho}$ and, for $s \geq T_c$, $\boldsymbol{\delta}^{*\prime}\mathbb{E}_e(\boldsymbol{y}_{s-1}) = \boldsymbol{\rho}^*$ (Hendry (2006), Clements & Hendry (1995)), then

$$\underset{l(s)N\times 1}{\boldsymbol{Z}_{0T}(s)} = \boldsymbol{R}\boldsymbol{\chi}_{0T}(s) + \tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s)$$

where $\tilde{\boldsymbol{P}}$ is given by

$$\underset{l(s)N\times l(s)N}{\tilde{\boldsymbol{P}}} = \begin{bmatrix} \tilde{\boldsymbol{P}}_{11} & \tilde{\boldsymbol{P}}_{12} & \mathbf{0}_{l_{pr}(s)N\times(l_{po}-1)N} \\ \tilde{\boldsymbol{P}}_{21} & \tilde{\boldsymbol{P}}_{22} & \tilde{\boldsymbol{P}}_{23} \end{bmatrix},$$

where, in turn, $\tilde{\boldsymbol{P}}_{ij}$, $i \in \{1,2\}$ and $j \in \{1,2,3\}$, are given by

$$\underset{l_{pr}(s)N\times N}{\tilde{\boldsymbol{P}}_{11}} = \begin{bmatrix} \boldsymbol{\gamma}\boldsymbol{\delta}' Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \\ \boldsymbol{\gamma}\boldsymbol{\delta}'(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \\ \vdots \\ \boldsymbol{\gamma}\boldsymbol{\delta}'(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-1}Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \end{bmatrix},$$

$$\underset{l_{pr}(s)N\times l_{pr}(s)N}{\tilde{\boldsymbol{P}}_{12}} = \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} \\ \boldsymbol{\gamma}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{1/2} & \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} \\ \boldsymbol{\gamma}\boldsymbol{\delta}'(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')\boldsymbol{\Sigma}^{1/2} & \boldsymbol{\gamma}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{1/2} & \dots & \mathbf{0}_{N\times N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\gamma}\boldsymbol{\delta}'(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-2}\boldsymbol{\Sigma}^{1/2} & \boldsymbol{\gamma}\boldsymbol{\delta}'(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-3}\boldsymbol{\Sigma}^{1/2} & \dots & \mathbf{0}_{N\times N} \end{bmatrix},$$

$$\underset{l_{po}N\times N}{\tilde{\boldsymbol{P}}_{21}} = \begin{bmatrix} \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s}Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \\ \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s}Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \\ \vdots \\ \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{T-T_c}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s}Ch(\mathbb{V}_e(\boldsymbol{y}_{s-1}))\boldsymbol{\Sigma}^{1/2} \end{bmatrix},$$

$$\underset{l_{po}N\times l_{pr}(s)N}{\tilde{\boldsymbol{P}}_{22}} = \begin{bmatrix} \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-1}\boldsymbol{\Sigma}^{1/2} & \dots & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}\boldsymbol{\Sigma}^{1/2} \\ \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-1}\boldsymbol{\Sigma}^{1/2} & \dots & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})\boldsymbol{\Sigma}^{1/2} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{T-T_c}(\boldsymbol{I}_N + \boldsymbol{\gamma}\boldsymbol{\delta}')^{T_c-s-1}\boldsymbol{\Sigma}^{1/2} & \dots & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{T-T_c}\boldsymbol{\Sigma}^{1/2} \end{bmatrix},$$

$$\underset{l_{po}N\times(l_{po}-1)N}{\tilde{\boldsymbol{P}}_{23}} = \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} \\ \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{\Sigma}^*)^{1/2} & \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} \\ \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{\Sigma}^*)^{1/2} & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{\Sigma}^*)^{1/2} & \dots & \mathbf{0}_{N\times N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{T-T_c-1}(\boldsymbol{\Sigma}^*)^{1/2} & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{I}_N + \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})^{T-T_c-2}(\boldsymbol{\Sigma}^*)^{1/2} & \dots & \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}(\boldsymbol{\Sigma}^*)^{1/2} \end{bmatrix}.$$

Substituting the expressions for $\boldsymbol{Z}_{1T}(s)$ and $\boldsymbol{Z}_{0T}(s)$ into the expressions for the estimators, we get Lemma 2.

A.3 Proof of Proposition 1

Using (12) in (4), the forecast errors become

$$\begin{aligned}
\underset{N\times 1}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)} &= (\boldsymbol{\pi}^* - \hat{\boldsymbol{\pi}}(s)) + (\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s))\boldsymbol{y}_T + \boldsymbol{e}^*_{T+1} = \\
&= (\boldsymbol{\pi}^* - l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s) + \tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)} - \\
&- \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)vec^{-1}(\tilde{\boldsymbol{c}} + \tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)})) + \\
&+ (\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s))\boldsymbol{y}_T + \boldsymbol{e}^*_{T+1}.
\end{aligned}$$

Since, $\forall t \geq T_c$, $\boldsymbol{\pi}^* = -\boldsymbol{\gamma}^*\boldsymbol{\rho}^* = -\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime}\mathbb{E}_e(\boldsymbol{y}_t)$, by adding and subtracting $\hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)\mathbb{E}_e(\boldsymbol{y}_T)$ and re-arranging, we get

$$\begin{aligned}\underset{N\times 1}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)} &= \boldsymbol{e}^*_{T+1} + \hat{\boldsymbol{\gamma}}(s)(\hat{\boldsymbol{\delta}}'(s)l^{-1}(s)\cdot vec^{-1}(\tilde{\boldsymbol{c}})\mathbf{1}_{l(s)} - \hat{\boldsymbol{\delta}}'(s)\mathbb{E}_e(\boldsymbol{y}_T)) + \\ &+ \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)l^{-1}(s)\cdot vec^{-1}(\tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)} - l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s) + \tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)} + \\ &+ (\boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s))(\boldsymbol{y}_T - \mathbb{E}_e(\boldsymbol{y}_T)).\end{aligned}$$

Re-arranging again, we obtain

$$\begin{aligned}\underset{N\times 1}{\hat{\boldsymbol{\xi}}_{T+1|T}(s)} &= \boldsymbol{e}^*_{T+1} - \hat{\boldsymbol{\gamma}}(s)(\hat{\boldsymbol{\delta}}'(s)\mathbb{E}_e(\boldsymbol{y}_T) - \hat{\boldsymbol{\delta}}'(s)l^{-1}(s)\cdot vec^{-1}(\tilde{\boldsymbol{c}})\mathbf{1}_{l(s)}) - \\ &- l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s) + \tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)vec^{-1}(\tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)}) - \\ &- (\hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s) - \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{y}_T - \mathbb{E}_e(\boldsymbol{y}_T)).\end{aligned}$$

Calling $\boldsymbol{A}_{1T}(s) = \hat{\boldsymbol{\gamma}}(s)(\hat{\boldsymbol{\delta}}'(s)\mathbb{E}_e(\boldsymbol{y}_T) - \hat{\boldsymbol{\delta}}'(s)l^{-1}(s)\cdot vec^{-1}(\tilde{\boldsymbol{c}})\mathbf{1}_{l(s)})$, $\boldsymbol{A}_{2T}(s) = l^{-1}(s)\cdot(vec^{-1}(\boldsymbol{R}\boldsymbol{\chi}_{0T}(s)+\tilde{\boldsymbol{P}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)} - \hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s)vec^{-1}(\tilde{\boldsymbol{D}}\boldsymbol{\chi}_{1T}(s))\mathbf{1}_{l(s)})$ and $\boldsymbol{A}_{3T}(s) = (\hat{\boldsymbol{\gamma}}(s)\hat{\boldsymbol{\delta}}'(s) - \boldsymbol{\gamma}^*\boldsymbol{\delta}^{*\prime})(\boldsymbol{y}_T - \mathbb{E}_e(\boldsymbol{y}_T))$, we get Proposition 1.

B Tables and figures

Table 1: Experimental design ($\boldsymbol{\gamma} = (-1, 0)'$)

Experiment	ρ_1^*	$\boldsymbol{\pi}^*$	θ_2^*	$\boldsymbol{\delta}^*$	σ_1^*	σ_2^*	$\boldsymbol{\Sigma}^*$
1. No break	–	–	–	–	–	–	–
2. ρ_1, θ_2 ($\uparrow$, $\downarrow$)	3.5	$\begin{bmatrix} 3.5 \\ 0 \end{bmatrix}$	0.1	$\begin{bmatrix} 1 \\ -0.1 \end{bmatrix}$	–	–	–
3. ρ_1, θ_2 ($\uparrow$, $\uparrow$)	3.5	$\begin{bmatrix} 3.5 \\ 0 \end{bmatrix}$	0.9	$\begin{bmatrix} 1 \\ -0.9 \end{bmatrix}$	–	–	–
4. ρ_1, θ_2 ($\downarrow$, $\downarrow$)	0.5	$\begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$	0.1	$\begin{bmatrix} 1 \\ -0.1 \end{bmatrix}$	–	–	–
5. ρ_1, θ_2 ($\downarrow$, $\uparrow$)	0.5	$\begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$	0.9	$\begin{bmatrix} 1 \\ -0.9 \end{bmatrix}$	–	–	–
6. σ_1, σ_2 ($\uparrow$, $\downarrow$)	–	–	–	–	1.5	0.5	$\begin{bmatrix} 2.25 & 0 \\ 0 & 0.25 \end{bmatrix}$
7. σ_1, σ_2 ($\downarrow$, $\uparrow$)	–	–	–	–	0.5	1.5	$\begin{bmatrix} 0.25 & 0 \\ 0 & 2.25 \end{bmatrix}$
8. σ_1, σ_2 ($\downarrow$, $\downarrow$)	–	–	–	–	0.5	0.5	$\begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}$
9. σ_1, σ_2 ($\uparrow$, $\uparrow$)	–	–	–	–	1.5	1.5	$\begin{bmatrix} 2.25 & 0 \\ 0 & 2.25 \end{bmatrix}$

Notes. Experiment 1 (no break) involves the following parameter setup: $\rho_1 = 2$, so that $\boldsymbol{\pi} = (2, 0)'$, $\theta_2 = 0.5$, so that $\boldsymbol{\delta} = (1, -0.5)'$, and $\sigma_1 = 1$ and $\sigma_2 = 1$, so that $\boldsymbol{\Sigma} = (1, 0 : 0, 1)$. The $\uparrow$ and $\downarrow$ symbols stand for 'increase' and 'decrease', resp.

Table 2: RTMSE conditional on $\boldsymbol{y}_T = \mathbb{E}_e(\boldsymbol{y}_T) + diag(\boldsymbol{\Sigma}^{(*)1/2})$

Experiment 1 (no break)

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	2.202	1.628	1.516	1.461	1.433
1	2.068	1.609	1.510	1.459	1.433
2	1.984	1.597	1.508	1.457	1.433
3	1.895	1.576	1.503	1.458	1.433
4	1.854	1.568	1.498	1.456	1.432
5	1.794	1.558	1.491	1.455	1.432
10	1.628	1.516	1.477	1.451	1.431
20	1.516	1.477	1.461	1.444	1.430
30	1.477	1.461	1.451	1.440	1.428
50	1.451	1.444	1.440	1.433	1.426
100	1.431	1.430	1.428	1.426	1.422

Experiment 4 (ρ_1, θ_2 (↓, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	2.272	1.699	1.584	1.523	1.493
1	1.993	1.627	1.552	1.508	1.487
2	2.340	1.663	1.535	1.479	1.464
3	2.605	1.792	1.595	1.490	1.459
4	2.753	1.925	1.650	1.509	1.461
5	2.749	1.976	1.694	1.534	1.465
10	2.499	2.080	1.838	1.638	1.501
20	2.160	1.992	1.863	1.713	1.565
30	1.998	1.906	1.843	1.729	1.600
50	1.877	1.830	1.785	1.723	1.624
100	1.801	1.761	1.732	1.694	1.612

Experiment 7 (σ_1, σ_2 (↓, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	2.183	1.702	1.637	1.606	1.592
1	2.059	1.689	1.635	1.606	1.592
2	1.985	1.689	1.634	1.606	1.592
3	1.909	1.680	1.633	1.607	1.592
4	1.874	1.673	1.632	1.606	1.592
5	1.841	1.667	1.630	1.606	1.592
10	1.723	1.645	1.621	1.604	1.592
20	1.651	1.622	1.611	1.601	1.591
30	1.625	1.612	1.605	1.598	1.591
50	1.608	1.602	1.599	1.594	1.589
100	1.594	1.593	1.591	1.589	1.587

Experiment 2 (ρ_1, θ_2 (↑, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	2.365	1.749	1.616	1.542	1.502
1	2.155	1.716	1.611	1.544	1.505
2	2.434	1.735	1.591	1.518	1.491
3	2.688	1.838	1.642	1.528	1.489
4	2.803	1.966	1.698	1.546	1.492
5	2.807	2.017	1.736	1.566	1.498
10	2.542	2.085	1.858	1.659	1.529
20	2.135	1.971	1.860	1.711	1.583
30	1.939	1.868	1.815	1.708	1.607
50	1.785	1.761	1.723	1.674	1.607
100	1.694	1.662	1.645	1.614	1.564

Experiment 5 (ρ_1, θ_2 (↓, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	2.160	1.669	1.571	1.517	1.490
1	2.024	1.668	1.579	1.525	1.494
2	2.519	1.808	1.648	1.555	1.510
3	2.851	1.996	1.739	1.599	1.527
4	3.008	2.135	1.832	1.637	1.546
5	3.030	2.207	1.904	1.675	1.563
10	2.785	2.317	2.064	1.822	1.639
20	2.338	2.195	2.083	1.914	1.746
30	2.104	2.076	2.021	1.916	1.794
50	1.949	1.938	1.920	1.882	1.819
100	1.854	1.827	1.814	1.804	1.789

Experiment 8 (σ_1, σ_2 (↓, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	1.101	0.814	0.758	0.730	0.717
1	1.107	0.830	0.767	0.734	0.718
2	1.232	0.885	0.794	0.741	0.719
3	1.202	0.898	0.807	0.746	0.720
4	1.181	0.909	0.814	0.750	0.722
5	1.145	0.909	0.814	0.754	0.723
10	0.976	0.867	0.810	0.761	0.727
20	0.848	0.806	0.783	0.756	0.731
30	0.798	0.777	0.767	0.750	0.731
50	0.761	0.753	0.748	0.740	0.728
100	0.734	0.733	0.731	0.727	0.722

Experiment 3 (ρ_1, θ_2 (↑, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	2.141	1.633	1.545	1.502	1.482
1	1.951	1.598	1.528	1.495	1.478
2	2.410	1.690	1.566	1.496	1.475
3	2.693	1.852	1.641	1.519	1.477
4	2.877	1.982	1.688	1.543	1.480
5	2.844	2.043	1.746	1.566	1.486
10	2.576	2.129	1.888	1.676	1.523
20	2.157	2.007	1.889	1.736	1.578
30	1.961	1.889	1.831	1.738	1.605
50	1.796	1.768	1.739	1.695	1.604
100	1.693	1.661	1.642	1.613	1.535

Experiment 6 (σ_1, σ_2 (↑, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	4.242	2.496	2.010	1.752	1.633
1	3.955	2.464	2.002	1.753	1.636
2	3.698	2.478	2.053	1.757	1.637
3	3.295	2.343	2.023	1.765	1.636
4	3.010	2.289	1.978	1.756	1.637
5	2.776	2.202	1.927	1.749	1.635
10	2.096	1.903	1.798	1.706	1.631
20	1.767	1.718	1.688	1.656	1.621
30	1.683	1.661	1.651	1.634	1.613
50	1.633	1.626	1.622	1.614	1.604
100	1.603	1.603	1.601	1.598	1.593

Experiment 9 (σ_1, σ_2 (↑, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	3.303	2.441	2.274	2.191	2.150
1	3.085	2.405	2.263	2.188	2.149
2	2.873	2.365	2.250	2.182	2.148
3	2.740	2.325	2.236	2.180	2.147
4	2.679	2.310	2.225	2.177	2.146
5	2.610	2.298	2.216	2.173	2.145
10	2.403	2.241	2.194	2.165	2.142
20	2.244	2.192	2.174	2.154	2.139
30	2.193	2.173	2.161	2.149	2.137
50	2.161	2.153	2.148	2.141	2.133
100	2.137	2.136	2.135	2.133	2.129

Notes. See Table 1.

Table 3: Bias conditional on $\boldsymbol{y}_T = \mathbb{E}_e(\boldsymbol{y}_T) + diag(\boldsymbol{\Sigma}^{(*)1/2})$

Experiment 1 (no break)

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.102	0.039	0.019	0.014	0.009
1	0.094	0.027	0.024	0.013	0.009
2	0.062	0.029	0.024	0.016	0.008
3	0.062	0.030	0.026	0.016	0.008
4	0.045	0.025	0.020	0.014	0.009
5	0.039	0.023	0.016	0.013	0.009
10	0.039	0.019	0.015	0.012	0.007
20	0.019	0.015	0.014	0.010	0.007
30	0.015	0.014	0.012	0.005	0.009
50	0.012	0.010	0.005	0.009	0.006
100	0.007	0.007	0.009	0.006	0.004

Experiment 4 (ρ_1, θ_2 ($\downarrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.351	0.392	0.395	0.401	0.402
1	0.184	0.291	0.336	0.363	0.386
2	0.216	0.021	0.092	0.175	0.269
3	0.379	0.145	0.043	0.062	0.189
4	0.490	0.254	0.138	0.017	0.128
5	0.557	0.327	0.207	0.078	0.080
10	0.685	0.527	0.395	0.260	0.079
20	0.801	0.651	0.559	0.428	0.251
30	0.844	0.718	0.644	0.516	0.345
50	0.913	0.793	0.716	0.619	0.466
100	0.980	0.878	0.808	0.728	0.587

Experiment 7 (σ_1, σ_2 ($\downarrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.082	0.032	0.017	0.008	0.004
1	0.067	0.022	0.021	0.010	0.004
2	0.050	0.025	0.014	0.008	0.004
3	0.050	0.020	0.010	0.009	0.003
4	0.048	0.020	0.015	0.009	0.004
5	0.045	0.022	0.014	0.007	0.003
10	0.034	0.016	0.011	0.008	0.003
20	0.017	0.010	0.008	0.007	0.003
30	0.012	0.008	0.007	0.007	0.002
50	0.008	0.007	0.007	0.004	0.002
100	0.003	0.003	0.001	0.002	0.002

Experiment 2 (ρ_1, θ_2 ($\uparrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.638	0.545	0.498	0.462	0.433
1	0.701	0.575	0.529	0.482	0.446
2	0.477	0.405	0.378	0.368	0.387
3	0.461	0.342	0.315	0.312	0.352
4	0.458	0.312	0.287	0.283	0.330
5	0.454	0.296	0.270	0.262	0.315
10	0.450	0.302	0.261	0.232	0.283
20	0.398	0.303	0.266	0.235	0.278
30	0.394	0.310	0.278	0.243	0.292
50	0.445	0.344	0.295	0.258	0.304
100	0.582	0.431	0.367	0.299	0.317

Experiment 5 (ρ_1, θ_2 ($\downarrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.509	0.472	0.444	0.426	0.412
1	0.634	0.536	0.491	0.454	0.427
2	0.893	0.683	0.594	0.524	0.472
3	1.005	0.774	0.652	0.576	0.506
4	1.058	0.830	0.709	0.614	0.533
5	1.078	0.858	0.749	0.644	0.555
10	1.099	0.950	0.843	0.744	0.644
20	1.060	0.972	0.917	0.825	0.751
30	1.014	0.978	0.928	0.877	0.811
50	1.028	0.974	0.948	0.900	0.871
100	1.070	0.994	0.958	0.923	0.933

Experiment 8 (σ_1, σ_2 ($\downarrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.051	0.019	0.009	0.007	0.004
1	0.042	0.013	0.010	0.006	0.005
2	0.049	0.014	0.012	0.008	0.004
3	0.043	0.014	0.014	0.008	0.005
4	0.040	0.010	0.011	0.008	0.005
5	0.034	0.010	0.006	0.007	0.005
10	0.020	0.010	0.010	0.005	0.004
20	0.010	0.008	0.007	0.002	0.004
30	0.007	0.007	0.005	0.002	0.007
50	0.005	0.003	0.001	0.004	0.003
100	0.003	0.003	0.006	0.004	0.003

Experiment 3 (ρ_1, θ_2 ($\uparrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.229	0.315	0.351	0.369	0.383
1	0.130	0.262	0.300	0.342	0.368
2	0.172	0.248	0.279	0.312	0.346
3	0.171	0.222	0.243	0.295	0.326
4	0.163	0.196	0.229	0.275	0.307
5	0.146	0.171	0.201	0.250	0.291
10	0.178	0.079	0.093	0.169	0.226
20	0.227	0.118	0.038	0.051	0.125
30	0.307	0.166	0.109	0.015	0.057
50	0.398	0.260	0.178	0.109	0.037
100	0.536	0.381	0.300	0.214	0.142

Experiment 6 (σ_1, σ_2 ($\uparrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.133	0.119	0.057	0.045	0.025
1	0.150	0.080	0.075	0.044	0.029
2	0.071	0.062	0.058	0.045	0.024
3	0.058	0.056	0.066	0.043	0.024
4	0.063	0.043	0.040	0.036	0.027
5	0.043	0.042	0.023	0.034	0.026
10	0.071	0.040	0.035	0.030	0.018
20	0.050	0.039	0.036	0.028	0.020
30	0.045	0.038	0.029	0.018	0.021
50	0.034	0.027	0.018	0.022	0.014
100	0.018	0.019	0.019	0.015	0.010

Experiment 9 (σ_1, σ_2 ($\uparrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.152	0.058	0.028	0.021	0.013
1	0.146	0.041	0.037	0.019	0.014
2	0.085	0.044	0.037	0.022	0.012
3	0.094	0.045	0.038	0.022	0.012
4	0.068	0.038	0.029	0.019	0.013
5	0.062	0.035	0.025	0.018	0.012
10	0.061	0.029	0.022	0.018	0.010
20	0.032	0.022	0.022	0.016	0.011
30	0.025	0.024	0.019	0.009	0.012
50	0.019	0.019	0.011	0.013	0.009
100	0.011	0.011	0.013	0.009	0.006

Notes. See Table 1.

Table 4: Bias in estimating $-\gamma_1\theta_2$

Experiment 1 (no break)

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.017	0.029	0.023	0.015	0.008
1	0.021	0.028	0.022	0.015	0.008
2	0.024	0.028	0.021	0.014	0.008
3	0.027	0.028	0.021	0.014	0.008
4	0.028	0.027	0.020	0.014	0.008
5	0.029	0.026	0.019	0.014	0.008
10	0.029	0.023	0.018	0.013	0.007
20	0.023	0.018	0.015	0.012	0.007
30	0.018	0.015	0.013	0.010	0.006
50	0.013	0.012	0.010	0.008	0.005
100	0.007	0.007	0.006	0.005	0.004

Experiment 4 (ρ_1, θ_2 (↓, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.001	0.008	0.006	0.004	0.002
1	-0.000	0.006	0.004	0.003	0.002
2	-0.008	-0.011	-0.013	-0.011	-0.008
3	-0.015	-0.018	-0.020	-0.020	-0.014
4	-0.014	-0.022	-0.024	-0.025	-0.018
5	-0.012	-0.023	-0.026	-0.028	-0.022
10	0.019	-0.015	-0.024	-0.034	-0.031
20	0.059	0.018	-0.005	-0.025	-0.034
30	0.084	0.040	0.013	-0.013	-0.031
50	0.109	0.067	0.041	0.008	-0.021
100	0.147	0.102	0.075	0.040	0.001

Experiment 7 (σ_1, σ_2 (↓, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.013	0.014	0.010	0.006	0.003
1	0.014	0.013	0.009	0.006	0.003
2	0.019	0.014	0.010	0.006	0.003
3	0.020	0.015	0.010	0.006	0.003
4	0.021	0.015	0.010	0.006	0.003
5	0.022	0.015	0.010	0.006	0.003
10	0.022	0.015	0.011	0.007	0.003
20	0.018	0.013	0.010	0.007	0.004
30	0.014	0.011	0.009	0.007	0.003
50	0.010	0.009	0.008	0.006	0.003
100	0.007	0.006	0.005	0.004	0.003

Experiment 2 (ρ_1, θ_2 (↑, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.001	0.008	0.006	0.004	0.002
1	-0.002	0.005	0.004	0.003	0.002
2	-0.009	-0.010	-0.013	-0.012	-0.007
3	-0.015	-0.017	-0.020	-0.020	-0.013
4	-0.016	-0.021	-0.025	-0.025	-0.018
5	-0.014	-0.022	-0.027	-0.028	-0.021
10	0.018	-0.015	-0.026	-0.034	-0.030
20	0.060	0.018	-0.005	-0.026	-0.033
30	0.084	0.039	0.012	-0.014	-0.030
50	0.109	0.066	0.040	0.007	-0.021
100	0.146	0.102	0.075	0.039	0.002

Experiment 5 (ρ_1, θ_2 (↓, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.029	0.040	0.030	0.019	0.010
1	0.020	0.028	0.023	0.015	0.009
2	-0.085	-0.101	-0.102	-0.086	-0.047
3	-0.159	-0.165	-0.169	-0.146	-0.087
4	-0.209	-0.210	-0.211	-0.188	-0.118
5	-0.252	-0.244	-0.242	-0.218	-0.144
10	-0.370	-0.346	-0.333	-0.305	-0.223
20	-0.479	-0.444	-0.423	-0.386	-0.299
30	-0.526	-0.497	-0.473	-0.433	-0.343
50	-0.560	-0.548	-0.530	-0.493	-0.400
100	-0.570	-0.582	-0.579	-0.558	-0.473

Experiment 8 (σ_1, σ_2 (↓, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.017	0.029	0.023	0.015	0.008
1	0.019	0.026	0.021	0.014	0.008
2	0.005	0.019	0.017	0.013	0.007
3	0.003	0.020	0.016	0.013	0.008
4	0.006	0.020	0.017	0.012	0.007
5	0.012	0.020	0.016	0.013	0.007
10	0.025	0.020	0.017	0.013	0.007
20	0.022	0.018	0.016	0.013	0.007
30	0.017	0.015	0.013	0.011	0.007
50	0.012	0.011	0.010	0.009	0.006
100	0.007	0.007	0.006	0.005	0.004

Experiment 3 (ρ_1, θ_2 (↑, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.029	0.040	0.030	0.019	0.010
1	0.022	0.028	0.022	0.015	0.009
2	-0.084	-0.100	-0.104	-0.088	-0.047
3	-0.157	-0.164	-0.171	-0.149	-0.088
4	-0.208	-0.208	-0.214	-0.191	-0.120
5	-0.252	-0.243	-0.245	-0.221	-0.145
10	-0.370	-0.345	-0.337	-0.309	-0.225
20	-0.478	-0.446	-0.426	-0.391	-0.300
30	-0.526	-0.498	-0.475	-0.438	-0.343
50	-0.560	-0.549	-0.532	-0.497	-0.401
100	-0.571	-0.583	-0.580	-0.560	-0.474

Experiment 6 (σ_1, σ_2 (↑, ↓))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.012	0.036	0.027	0.018	0.011
1	0.019	0.036	0.027	0.018	0.011
2	-0.006	0.023	0.022	0.018	0.010
3	-0.009	0.020	0.018	0.016	0.010
4	0.001	0.019	0.018	0.015	0.010
5	0.011	0.021	0.018	0.015	0.010
10	0.027	0.021	0.018	0.015	0.009
20	0.024	0.020	0.016	0.014	0.008
30	0.020	0.017	0.015	0.012	0.007
50	0.014	0.013	0.012	0.009	0.006
100	0.008	0.007	0.007	0.006	0.004

Experiment 9 (σ_1, σ_2 (↑, ↑))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.017	0.029	0.023	0.015	0.008
1	0.022	0.029	0.023	0.015	0.008
2	0.029	0.029	0.022	0.015	0.008
3	0.031	0.029	0.021	0.014	0.008
4	0.033	0.028	0.021	0.014	0.008
5	0.031	0.027	0.020	0.014	0.008
10	0.030	0.024	0.019	0.013	0.007
20	0.024	0.019	0.016	0.012	0.007
30	0.020	0.016	0.013	0.010	0.006
50	0.014	0.013	0.011	0.009	0.005
100	0.008	0.007	0.007	0.006	0.004

Notes. See Table 1.

Table 5: Bias in estimating $-\gamma_1\rho_1$

Experiment 1 (no break)

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.205	0.171	0.112	0.065	0.033
1	0.220	0.158	0.109	0.062	0.034
2	0.227	0.144	0.100	0.061	0.034
3	0.254	0.145	0.094	0.061	0.034
4	0.245	0.134	0.095	0.063	0.032
5	0.226	0.122	0.089	0.061	0.031
10	0.171	0.112	0.081	0.060	0.030
20	0.112	0.081	0.065	0.056	0.028
30	0.081	0.065	0.060	0.048	0.023
50	0.060	0.056	0.048	0.033	0.022
100	0.030	0.028	0.023	0.022	0.019

Experiment 4 (ρ_1, θ_2 ($\downarrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.117	0.080	0.045	0.025	0.012
1	0.045	0.057	0.032	0.023	0.011
2	-0.122	-0.035	-0.050	-0.026	-0.012
3	-0.036	-0.091	-0.085	-0.051	-0.025
4	0.009	-0.087	-0.109	-0.053	-0.036
5	0.119	-0.083	-0.102	-0.055	-0.043
10	0.191	-0.013	-0.038	-0.052	-0.051
20	0.144	0.043	0.051	-0.026	-0.027
30	0.188	0.104	0.084	0.004	0.004
50	0.287	0.166	0.128	0.051	0.050
100	0.474	0.320	0.243	0.159	0.172

Experiment 7 (σ_1, σ_2 ($\downarrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.074	0.070	0.044	0.025	0.012
1	0.095	0.064	0.043	0.024	0.012
2	0.122	0.063	0.041	0.023	0.013
3	0.142	0.069	0.041	0.025	0.013
4	0.146	0.069	0.046	0.027	0.012
5	0.144	0.067	0.044	0.027	0.012
10	0.120	0.072	0.044	0.030	0.013
20	0.093	0.055	0.042	0.032	0.014
30	0.069	0.050	0.040	0.029	0.013
50	0.053	0.044	0.034	0.022	0.014
100	0.026	0.023	0.018	0.017	0.013

Experiment 2 (ρ_1, θ_2 ($\uparrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.367	0.299	0.216	0.139	0.072
1	0.203	0.194	0.148	0.109	0.063
2	-0.897	-0.691	-0.630	-0.499	-0.295
3	-1.297	-1.142	-1.011	-0.838	-0.534
4	-1.565	-1.390	-1.289	-1.046	-0.713
5	-1.618	-1.562	-1.428	-1.197	-0.852
10	-1.912	-1.914	-1.815	-1.631	-1.262
20	-2.193	-2.151	-2.052	-1.946	-1.618
30	-2.272	-2.245	-2.177	-2.103	-1.806
50	-2.307	-2.336	-2.306	-2.254	-2.027
100	-2.276	-2.359	-2.389	-2.405	-2.298

Experiment 5 (ρ_1, θ_2 ($\downarrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.063	0.065	0.031	0.012	0.006
1	0.084	0.046	0.025	0.008	0.006
2	0.312	0.140	0.073	0.008	0.019
3	0.298	0.211	0.107	0.011	0.028
4	0.226	0.239	0.122	0.012	0.035
5	0.142	0.228	0.114	0.013	0.040
10	0.238	0.181	0.095	0.044	0.062
20	0.308	0.163	0.093	0.081	0.094
30	0.301	0.202	0.126	0.121	0.122
50	0.395	0.282	0.198	0.141	0.170
100	0.533	0.376	0.292	0.233	0.287

Experiment 8 (σ_1, σ_2 ($\downarrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.204	0.170	0.114	0.065	0.034
1	0.221	0.140	0.105	0.058	0.033
2	0.127	0.101	0.087	0.057	0.033
3	0.158	0.114	0.072	0.057	0.034
4	0.138	0.094	0.078	0.066	0.031
5	0.130	0.088	0.062	0.063	0.029
10	0.119	0.116	0.073	0.064	0.032
20	0.089	0.078	0.069	0.057	0.030
30	0.073	0.064	0.058	0.049	0.022
50	0.056	0.055	0.047	0.032	0.024
100	0.028	0.026	0.022	0.022	0.019

Experiment 3 (ρ_1, θ_2 ($\uparrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.246	0.210	0.136	0.078	0.039
1	0.149	0.126	0.095	0.056	0.036
2	-1.448	-0.900	-0.697	-0.519	-0.248
3	-2.007	-1.379	-1.113	-0.860	-0.452
4	-2.279	-1.657	-1.394	-1.097	-0.609
5	-2.473	-1.834	-1.591	-1.267	-0.736
10	-2.471	-2.205	-2.019	-1.706	-1.132
20	-2.419	-2.363	-2.271	-2.007	-1.501
30	-2.439	-2.398	-2.340	-2.120	-1.695
50	-2.389	-2.387	-2.375	-2.267	-1.916
100	-2.307	-2.384	-2.411	-2.381	-2.174

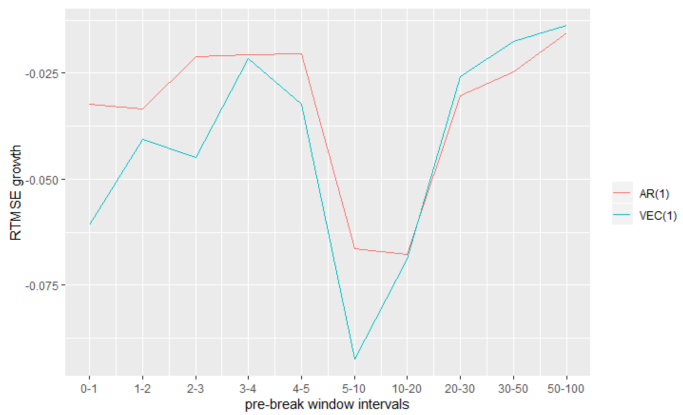
Experiment 6 (σ_1, σ_2 ($\uparrow$, $\downarrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.378	0.276	0.164	0.090	0.046
1	0.432	0.266	0.169	0.088	0.048
2	0.347	0.193	0.154	0.088	0.048
3	0.280	0.183	0.118	0.087	0.048
4	0.237	0.107	0.113	0.096	0.044
5	0.210	0.101	0.093	0.091	0.041
10	0.154	0.106	0.097	0.081	0.040
20	0.095	0.086	0.079	0.067	0.034
30	0.079	0.068	0.072	0.055	0.025
50	0.062	0.064	0.057	0.034	0.025
100	0.033	0.030	0.027	0.023	0.020

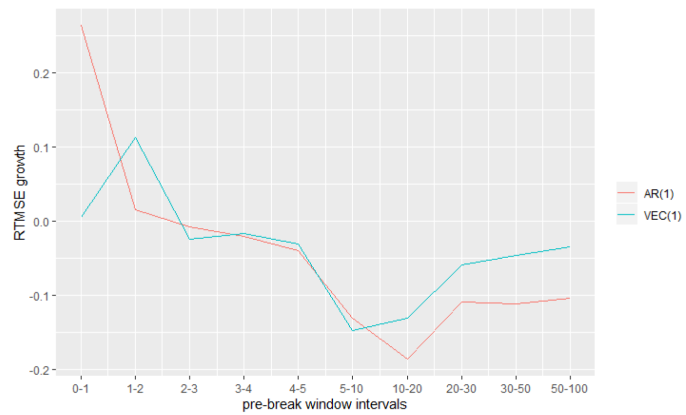
Experiment 9 (σ_1, σ_2 ($\uparrow$, $\uparrow$))

$l_{pr}(s)/l_{po}$	10	20	30	50	100
0	0.205	0.172	0.111	0.064	0.032
1	0.225	0.164	0.109	0.063	0.033
2	0.255	0.154	0.103	0.062	0.034
3	0.276	0.152	0.098	0.062	0.033
4	0.269	0.143	0.098	0.062	0.032
5	0.257	0.133	0.094	0.061	0.031
10	0.200	0.115	0.084	0.059	0.029
20	0.133	0.087	0.068	0.055	0.027
30	0.095	0.070	0.065	0.049	0.025
50	0.068	0.060	0.053	0.036	0.023
100	0.033	0.030	0.026	0.024	0.020

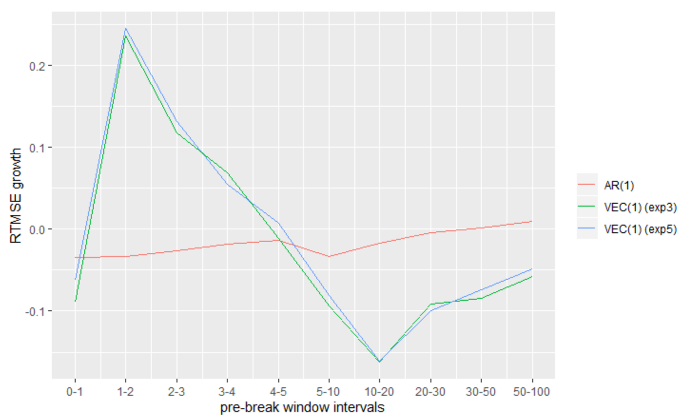
Notes. See Table 1.



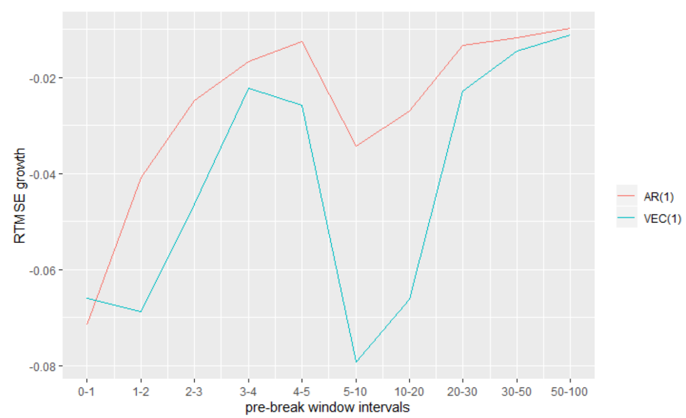
(a) Experiment 1 vs 1 (PT(2005))



(b) Experiment 8 vs 7 (PT(2005))



(c) Experiment 3 & 5 vs 4 (PT(2005))



(d) Experiment 9 vs 6 (PT (2005))

Figure 1: RTMSE growth comparison: Table 2, columns $l_{po} = 10$, vs Pesaran & Timmermann (2005) (PT (2005)).

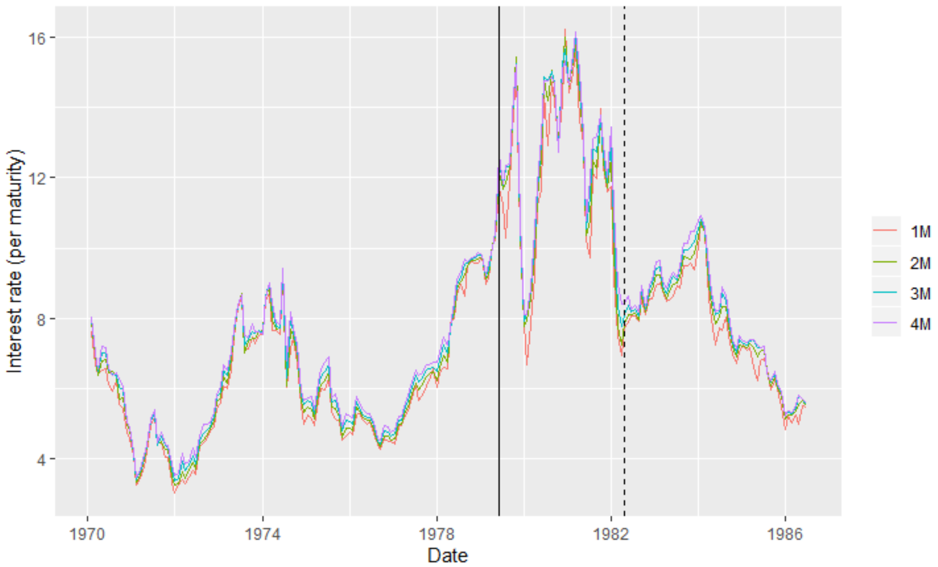


Figure 2: US T-bond interest rates at different maturities: 1-4 months. The vertical lines are the identified period of turbulence (Hall *et al.* (1992)).

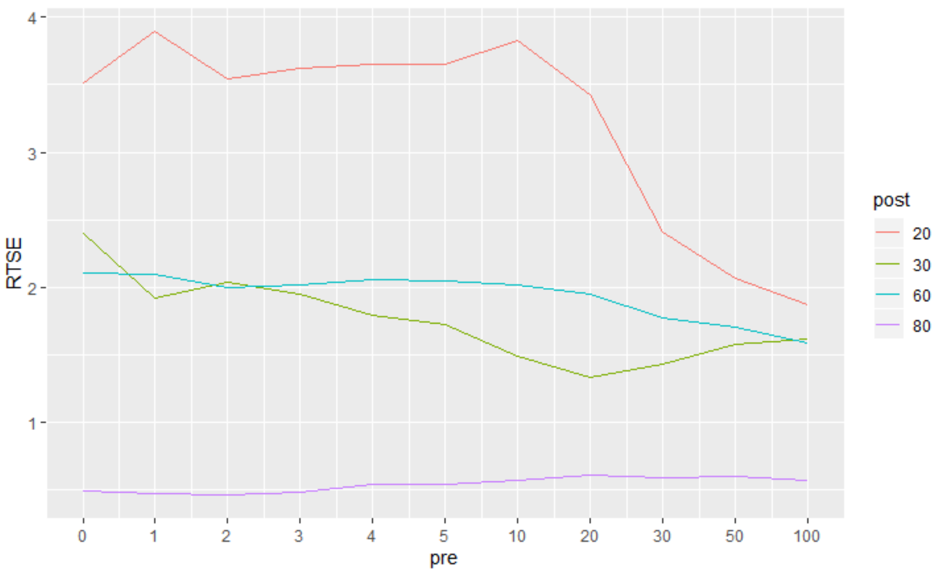


Figure 3: RTSE vs pre-break window length (by post-break window length)

References

Boot, T., & Pick, A. 2019. Does modeling a structural break improve forecast accuracy? *Journal of Econometrics*, (forthcoming).

Castle, J. L., Clements, M. P., & Hendry, D. F. 2016. An overview of forecasting facing breaks. *Journal of Business Cycle Research*, **12**, 3–23.

Clements, M. P., & Hendry, D. F. 1995. Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**(2), 127–146.

Clements, M. P., & Hendry, D. F. 1998. Forecasting economic processes. *International Journal of Forecasting*, **14**, 111–131.

Clements, M. P., & Hendry, D. F. 2001. Forecasting with difference-stationary and trend-stationary models. *Econometrics Journal*, **4**, 1–9.

Giraitis, L., Kapetanios, G., & Price, S. 2013. Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics*, **177**, 153–170.

Hall, A.D., Anderson, H.M., & Granger, C.W.J. 1992. A cointegration analysis of treasury bill yields. *The Review of Economics and Statistics*, **74**, 116–126.

Hansen, B.E. 2018. Johansen's reduced rank estimator is GMM. *Econometrics.*

Hansen, H., & Johansen, S. 1999. Some tests for parameter constancy in cointegrated VAR-models. *The Econometrics Journal*, **2**(2), 306–333.

Hendry, D. F. 2006. Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics*, **135**, 399–426.

Inoue, A., Jin, L., & Rossi, B. 2017. Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, **196**, 55–67.

Johansen, S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**(2), 231–254.

Johansen, S. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 1551–1580.

Pesaran, M. H., & Pick, A. 2011. Forecast combination across estimation windows. *Journal of Business and Economic Statistics*, **29**, 307–318.

Pesaran, M. H., & Timmermann, A. 2004. How costly is to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting*, **20**, 411–425.

Pesaran, M. H., & Timmermann, A. 2005. Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, **129**, 183–217.

Pesaran, M. H., & Timmermann, A. 2007. Selection of estimation window in the presence of breaks. *Journal of Econometrics*, **137**, 134–161.

Pesaran, M. H., Pick, A., & Pranovich, M. 2013. Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, **177**, 134–152.

Stock, J. H., & Watson, M. W. 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Businness and Economic Statistics*, **14**(1), 11–30.

van Garderen, K. J., & Boswijk, H. P. 2014. Bias correcting adjustment coefficients in a cointegrated VAR with known cointegrating vectors. *Economics Letters*, **122**, 224–228.