Graduate School of Decision Sciences

Universität Konstanz

GSDS
Working Paper
No. 2016-05

Volatility forecasting using global stochastic financial trends extracted from non-synchronous data

Lyudmila Grigoryeva
Juan-Pablo Ortega
Anatoly Peresetsky

August 2016

Graduate School of Decision Sciences

All processes within our society are based on decisions – whether they are individual or collective decisions. Understanding how these decisions are made will provide the tools with which we can address the root causes of social science issues.

The GSDS offers an open and communicative academic environment for doctoral researchers who deal with issues of decision making and their application to important social science problems. It combines the perspectives of the various social science disciplines for a comprehensive understanding of human decision behavior and its economic and political consequences.

The GSDS primarily focuses on economics, political science and psychology, but also encompasses the complementary disciplines computer science, sociology and statistics. The GSDS is structured around four interdisciplinary research areas: (A) Behavioural Decision Making, (B) Intertemporal Choice and Markets, (C) Political Decisions and Institutions and (D) Information Processing and Statistical Analysis.

GSDS – Graduate School of Decision Sciences
University of Konstanz
Box 146
78457 Konstanz

Phone: +49 (0)7531 88 3633
Fax: +49 (0)7531 88 5193
E-mail: gsds.office@uni-konstanz.de

-gsds.uni-konstanz.de

ISSN: 2365-4120
September 2016

Volatility forecasting using global stochastic financial trends extracted from non-synchronous data

Lyudmila Grigoryeva[†1] , Juan-Pablo Ortega[2,3], and Anatoly Peresetsky[4]

Abstract

This paper introduces a method based on various linear and nonlinear state space models that are used to extract global stochastic financial trends (GST) out of non-synchronous financial data. More specifically, these models are constructed to take advantage of the intraday arrival of closing information coming from different international markets to improve volatility description and forecasting. A set of three major asynchronous international stock market indices is used in order to empirically show that this forecasting scheme is capable of significant performance gains when compared to standard models like the dynamic conditional correlation (DCC) family.

Keywords: multivariate volatility modeling and forecasting, global stochastic trend, extended Kalman filter, dynamic conditional correlations (DCC), non-synchronous data.

[†]Corresponding author.

[1]Department of Mathematics and Statistics. Graduate School of Decision Sciences. Universität Konstanz. Box 146. D-78457 Konstanz. Germany. Lyudmila.Grigoryeva@uni-konstanz.de

[2]Universität Sankt Gallen. Bodanstrasse 6. CH-9000 Sankt Gallen. Switzerland.

[3]Centre National de la Recherche Scientifique (CNRS). France. Juan-Pablo.Ortega@univ-fcomte.fr

[4]International Laboratory of Quantitative Finance, Higher School of Economics, Myasnitskaya str. 20. Moscow. Russia. aperesetsky@hse.ru

Acknowledgments: LG acknowledges partial financial support from the Faculty for the Future Program of the Schlumberger Foundation. LG and JPO acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493) as well as the hospitality of the International Laboratory of Quantitative Finance of the Higher School of Economics, Moscow, that made possible the collaborative work that lead to the results presented in this paper. The research of AP was funded by the Russian Federation government grant 14.A12.31.0007.

Contents

1 Introduction

Many frameworks for the description of financial returns have as their first building block a factor model of the form

$$r_t = \alpha + \beta y_t + u_t \quad \text{with} \quad \{u_t\} \sim \text{WN}(0, \sigma^2),$$

where the instantaneous returns r_t at time t of individual assets are presented as a mean-zero stochastic stationary additive perturbation of an affine function of a common factor y_t. This factor usually accounts for a common market feature to which all the assets under study are exposed. Consequently, this functional dependence allows to determine, for a given asset return r_t, how much of it has to do with the market situation (through the coefficient β, which is a function of the correlation between r_t and y_t) and how much comes from an idiosyncratic perturbation u_t specifically related to the individual asset. In particular applications of this model, the factor values y_t are sometimes computed by using an index constructed out of a set of assets that represent the class to which r_t is naturally associated. An alternative to this approach consists of treating the common factor values as a non-observable variable and of extracting them using observed individual returns via a Kalman-type state-space model. This direction has already been profusely explored in the literature. In the early available works (see for instance [Jeon 91, Kasa 92, Chun 94, Sikl 01, Rang 01, Rang 02, Phen 04]) the authors consider low sampling frequencies in order to be able to neglect asynchronicity issues. In several more recent references [Dung 01, Chou 07, Chan 09, Luce 11, Cart 11, Bae 11, Feli 12, Bent 15] daily quoted data are used but always using a synchronized approach.

In this work we use a different point of view introduced in [Korh 13] and later on extended in [Durd 14, Pere 15], in which the market returns are thought of as a non-observable global stochastic trend (GST) whose value is ruled by the arrival of information coming from different local markets. In this framework, the returns of the GST are estimated several times per day at time points that are synchronized with the closing times of the markets that are assumed to drive it. This approach is implemented by setting a state-space model in which the observation equation writes the different observable individual market returns that we are interested in as a stochastic perturbation of an affine function of the estimated GST return accumulated during the 24 hours that precede this quote. It is assumed that the observed returns are those that drive the GST and hence its returns are estimated as many times per day as different closing times are included in the list of markets considered.

This point of view has been studied in [Korh 13] using three different setups, namely: three world indices (NIKKEI, MICEX, S&P500) with three different closing times, five world indices (NIKKEI, MICEX, DAX, PX, S&P500) with four different closing times, and ten world indices (NIKKEI, HSI, SENSEX, MICEX, DAX, PX, FTSE, IBOV, DJI, S&P500) with seven different closing times. The estimates of the GST obtained in these different situations are remarkably similar. The robustness that these results indicate allowed the authors to identify, for each market, the relative importance of local with respect to global news in stock prices formation.

The main goal of our work is modifying this approach in order to make it amenable to volatility forecasting and to prove the pertinence of the resulting method when compared to more standard families of models designed to specifically carry out this task. The rationale behind this attempt is that the error inherent to the filtering and forecasting of an unobserved variable like the GST is compensated by the more frequent information updates that the use of asynchronous information carries in its wake.

Since the models introduced in [Korh 13, Pere 15] are intrinsically homoscedastic, they are not appropriate to handle financial volatility modeling and forecasting. The heteroscedastic generalization needed for this purpose can be naturally implemented by using two different approaches. The simplest one consists of using the linear state-space approach in [Korh 13] in a first step to estimate the GST and to subsequently model the volatility and conditional correlation of the resulting global trend and idiosyncratic term using an adapted multivariate correlation model; for this purpose, we consider in this work adapted scalar and non-scalar versions of the dynamic conditional correlation (DCC) model introduced in [Engl 02, Tse 02]. The non-scalar models are estimated using the techniques introduced in [Chre 14, Bauw 15]. The adjustments of these standard models for the handling of the GST are implemented at the level of the so called “deGARCHing” or “first estimation step” in which a model for the conditional variances of the assets of interest is chosen; in our context, we put to work in this step two different GARCH-type models that take into account in their specification the chronology with which the different intraday trend returns are quoted.

A more sophisticated approach that we also study is the inclusion of the heteroscedasticity assumption on the GST returns directly in the formulation of the state-space model by using a GARCH-type and GST-adapted prescription of the type that we just described. The main complication that arises in this setup is the nonlinearity of the resulting modeling scheme that we handle using the extended Kalman filter (EKF) (see [Durb 12] and references therein).

The paper is organized as follows. In Section 2.1 we explain in detail the linear and nonlinear state-space models that we propose. We give details on how they handle the asynchronous character of the observable data and prove rigorous sufficient conditions that ensure their proper identification. Section 2.2 contains details on the Kalman filter based model estimation techniques that we use in the paper, as well as on the model specifications for the conditional variances incorporated in the nonlinear state-space model, together with the positivity and stationarity constraints that need to be imposed at the time of estimation. The GST-based volatility forecasting scheme is described in Section 3. Section 4 contains an empirical study using the adjusted closing values of three major indices (NIKKEI, FTSE, and S&P500) that are quoted at different times due to the time zones in which they are geographically based. In this experiment, we use the model confidence set (MCS) approach of [Hans 03, Hans 11] and we implement it with loss functions constructed with the conditional covariance matrices implied by the different models under consideration. *The results show that the proposed forecasting scheme exhibits excellent and statistically significant performance improvements when compared to the use of standard multivariate parametric correlation models.* Section 5 concludes the paper. The proofs of various technical results in the paper are contained in the appendices in Section A. Additionally, the last appendix in this section contains a replicate of the empirical study in Section 4 in which the estimation period does not include the Fall 2008 volatility events and the out-of-sample forecasting period comprises the Great Recession; this experiments aims at illustrating the robustness of our empirical results with respect to the estimation period used and the pertinence in some situations of the nonlinear state-space model.

Notation and conventions: column vectors are denoted by a bold lower case symbol like $\mathbf{v}$ and $\mathbf{v}^\top$ indicates its transpose. Given a vector $\mathbf{v} \in \mathbb{R}^n$, we denote its entries by v_i, with $i \in \{1, \ldots, n\}$; we also write $\mathbf{v} = (v_i)_{i\in\{1,\ldots,n\}}$. The symbols $\mathbf{i}_n, \mathbf{0}_n \in \mathbb{R}^n$ stand for the vectors of length n consisting of ones and zeros, respectively. We denote by $\mathbb{M}_{n,m}$ the space of real $n \times m$ matrices with $m, n \in \mathbb{N}$. When $n = m$, we use the symbols $\mathbb{M}_n$ and $\mathbb{D}_n$ to refer to the space of square and diagonal matrices of order n, respectively. Given a matrix $A \in \mathbb{M}_{n,m}$, we denote its components by A_{ij} and we write $A = (A_{ij})$, with $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots m\}$. We use $\mathbb{S}_n$ to denote the subspace $\mathbb{S}_n \subset \mathbb{M}_n$ of symmetric matrices:

$$\mathbb{S}_n = \left\{ A \in \mathbb{M}_n \mid A^\top = A \right\},$$

and we use $\mathbb{S}_n^+$ (respectively $\mathbb{S}_n^-$) to refer to the cone $\mathbb{S}_n^+ \subset \mathbb{S}_n$ (respectively $\mathbb{S}_n^- \subset \mathbb{S}_n$) of positive (respectively negative) semidefinite matrices. When $A \in \mathbb{S}_n^+$ (respectively, $A \in \mathbb{S}_n^-$) we write $A \succeq 0$ (respectively, $A \preceq 0$). The symbol $\mathbb{I}_n \in \mathbb{D}_n$ denotes the identity matrix. Given two matrices $A, B \in \mathbb{M}_{n,m}$, we denote by $A \odot B \in \mathbb{M}_{n,m}$ their elementwise multiplication matrix or Hadamard product, that is:

$$(A \odot B)_{ij} := A_{ij} B_{ij} \quad \text{for all} \quad i \in \{1, \ldots, n\}, \quad j \in \{1, \ldots, m\}. \tag{1.1}$$

We denote as Diag the operator $\mathrm{Diag} : \mathbb{M}_n \longrightarrow \mathbb{D}_n$ that sets equal to zero all the components of a square matrix except for those that are on the main diagonal. The operator $\mathrm{diag} : \mathbb{R}^n \longrightarrow \mathbb{D}_n$ takes a given vector and constructs a diagonal matrix with its entries in the main diagonal.

2 State-space model for the global stochastic trend

We start by recalling the description of the global stochastic trend as the state variable in a state-space model, as it was introduced in [Korh 13]. In order to keep the presentation simple, we will carry this out for only three different non-synchronous assets, which is the framework in which our subsequent empirical analysis takes place. The generalization to more assets and quoting times is straightforward.

Let $\mathbf{r}_t \in \mathbb{R}^3$ be a vector containing three non-synchronous stock market returns (typically based on adjusted closing prices) quoted at different times of the same calendar date $t \in \mathbb{N}$. The different intraday quoting times have typically to do with lags in the closing times of the different markets. The intraday moments of time t_i, $i \in \{1,2,3\}$ of the given day t at which the components of $(r_{1,t}, r_{2,t}, r_{3,t})^\top$ of $\mathbf{r}_t$ become available are labeled as $t_i := 3(t-1)+i$, $t \in \mathbb{N}$.

We now assume the existence of an underlying and non-observable global stochastic trend and we denote by s_{t_i}, $i \in \{1,2,3\}$, its intraday log-values for the given calendar date $t \in \mathbb{N}$. We now define $\boldsymbol{\varepsilon}_t \in \mathbb{R}^3$ as the vector that contains the intra-day stochastic trend log-return components of a given calendar day t, that is,

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix} := \begin{pmatrix} s_{t_1} - s_{(t-1)_3} \\ s_{t_2} - s_{t_1} \\ s_{t_3} - s_{t_2} \end{pmatrix}. \tag{2.1}$$

Following the factor model scheme discussed in the introduction, for every $t \in \mathbb{Z}$ we write the log-returns of the components of $\mathbf{r}_t$ as excess returns with respect to an affine function of the entries of the vector $\mathbf{y}_t \in \mathbb{R}^3$, which is constructed with the daily global stochastic trend returns computed at the moments in time in which the components of $\mathbf{r}_t$ are quoted. More specifically $\mathbf{y}_t := (s_{t_1} - s_{(t-1)_1}, s_{t_2} - s_{(t-1)_2}, s_{t_3} - s_{(t-1)_3})^\top$ and

$$r_{i,t} = \alpha_i + \beta_i y_{i,t} + u_{i,t}, \quad i = \{1,2,3\}, \quad t \in \mathbb{Z}, \tag{2.2}$$

with the regression intercepts $\alpha_i \in \mathbb{R}$, $i = \{1,2,3\}$, and the parameters $\boldsymbol{\beta} := (\beta_1, \beta_2, \beta_3)^\top \in \mathbb{R}^3$. For the time being, in this relation we only assume that the residuals $\mathbf{u}_t$ are serially uncorrelated (they are a white noise) with mean zero and unconditional diagonal covariance matrix $\Sigma_u \in \mathbb{S}_3^+$, that is $\{\mathbf{u}_t\} \sim \mathrm{WN}(\mathbf{0}_3, \Sigma_u)$, $\Sigma_u := \mathrm{diag}\left(\sigma_{u,1}^2, \sigma_{u,2}^2, \sigma_{u,3}^2\right)$ with entries $\sigma_{u,1}, \sigma_{u,2}, \sigma_{u,3} \in \mathbb{R}^+$.

Using the definition (2.1), we write the returns $\mathbf{r}_t$ in (2.2) in terms of the global stochastic trend returns in the preceding twenty-four hours. It is easy to verify that using (2.1) in the regression expression (2.2) yields the following identity

$$\mathbf{r}_t = \boldsymbol{\alpha} + B\boldsymbol{e}_t + \mathbf{u}_t, \tag{2.3}$$

with $\boldsymbol{e}_t := (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}, \varepsilon_{2,t-1}, \varepsilon_{3,t-1})^\top$, where $\boldsymbol{\alpha} \in \mathbb{R}^3$, $\{\mathbf{u}_t\} \sim \mathrm{WN}(\mathbf{0}_3, \Sigma_u)$ as in (2.2), and the matrix $B \in \mathbb{M}_{5,3}$ is of the form

$$B := \begin{pmatrix} \beta_1 & 0 & 0 & \beta_1 & \beta_1 \\ \beta_2 & \beta_2 & 0 & 0 & \beta_2 \\ \beta_3 & \beta_3 & \beta_3 & 0 & 0 \end{pmatrix}. \tag{2.4}$$

This equation describes the returns dynamics in terms of the non-observable global stochastic trend returns in the preceding twenty-four hours. In order to estimate this model and to filter out of it the values of the GST, we will proceed by considering (2.3) as the observation equation of several linear and nonlinear state-space models that we design in order make possible their subsequent use for volatility forecasting.

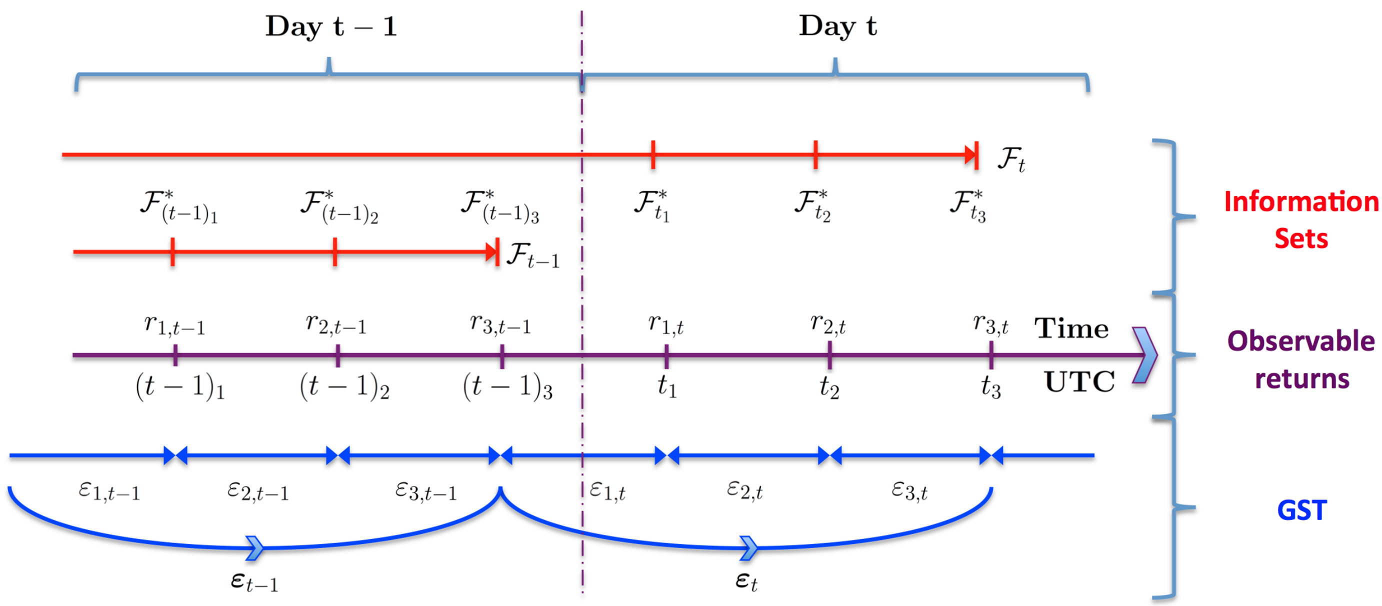

Figure 1: Diagram representing the different variables, time labels, and information sets used in the study as well as their chronology.

2.1 The linear and nonlinear state-space models

The linear state-space model. The first model that we present in this subsection is identical to the one originally considered in [Korh 13]:

$$\begin{cases} \mathbf{r}_t = \boldsymbol{\alpha} + B\boldsymbol{e}_t + \mathbf{u}_t, & \{\mathbf{u}_t\} \sim \mathrm{WN}\left(\mathbf{0}_3, \Sigma_u\right), \text{with } \Sigma_u := \mathrm{diag}\left(\sigma^2_{u,1}, \sigma^2_{u,2}, \sigma^2_{u,3}\right), \quad (2.5a) \\ \boldsymbol{e}_t = T\boldsymbol{e}_{t-1} + R\mathbf{v}_{t-1}, & \{\mathbf{v}_t\} \sim \mathrm{WN}(\mathbf{0}_3, \mathbb{I}_3), \quad (2.5b) \end{cases}$$

where $\boldsymbol{e}_t \in \mathbb{R}^5$, $\boldsymbol{\alpha} \in \mathbb{R}^3$, the matrix $B \in \mathbb{M}_{5,3}$ is provided in (2.4), the matrices $R \in \mathbb{M}_{3,5}$ and $T \in \mathbb{M}_5$ are given by

$$R := \begin{pmatrix} \sigma_{v,1} & 0 & 0 \\ 0 & \sigma_{v,2} & 0 \\ 0 & 0 & \sigma_{v,3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \qquad (2.6)$$

and $\sigma_{u,1}, \sigma_{u,2}, \sigma_{u,3}, \sigma_{v,1}, \sigma_{v,2}, \sigma_{v,3} \in \mathbb{R}^+$. We emphasize that (2.5a)-(2.5b) constitutes a linear state-space model in which the dynamical behavior of the GST is prescribed by the corresponding state equation and where the observation equation establishes a relation between the time evolution of the GST and the observed returns.

In the following proposition, whose proof is provided in an appendix, we study the identification of this model and state sufficient conditions that ensure it.

Proposition 2.1 *The linear state-space model (2.5a)-(2.5b) is well identified with respect to the natural invariance properties spelled out in the Appendix A.1 if one of the elements of the vector* $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ *that define the matrix* B *in (2.4) is set equal to a constant or, alternatively, when one of the unconditional variances* $\sigma^2_{v,1}, \sigma^2_{v,2}, \sigma^2_{v,3}$ *that define* R *in (2.5b) is set equal to a positive constant.*

The nonlinear state-space model. The dynamic specification (2.5b) does not introduce any dependence between the components $(\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$ of the GST $\boldsymbol{\varepsilon}_t$, even though this feature is empirically observed. This has motivated the introduction in [Durd 14, Pere 15] of a VAR-type prescription for the dynamics of $\boldsymbol{\varepsilon}_t$ that allows for correlation between its components while preserving the linearity of the corresponding state-space model. Since in this work we focus on volatility forecasting, we will instead introduce dependence between the components of $\boldsymbol{\varepsilon}_t$ via a specific functional choice for the time evolution of their conditional variances. This approach obliges us to formulate a nonlinear state-space model which makes us use an extended Kalman filter (EKF) instead of a standard one.

There are many different approaches that can be taken in order to implement this strategy. The most comprehensive one consists of building into the state-space model the entire conditional covariance dynamics of both the GST and the residuals. In unreported numerical experiments, we observed that the complexity of the resulting specification makes its estimation difficult to implement. This limitation makes advisable the adoption of an intermediate two-steps solution in which the nonlinear state-space model incorporates the modeling of the conditional variances and then the conditional covariances are handled separately in a second step.

Consider the following modified nonlinear state-space model:

$$\begin{cases} \mathbf{r}_t = \boldsymbol{\alpha} + B\boldsymbol{e}_t + \mathbf{u}_t, & \{\mathbf{u}_t\} \sim \mathrm{WN}\left(\mathbf{0}_3, \Sigma_u\right), \quad \Sigma_u := \mathrm{diag}\left(\sigma_{u,1}^2, \sigma_{u,2}^2, \sigma_{u,3}^2\right), \quad (2.7a) \\ \boldsymbol{e}_t = T\boldsymbol{e}_{t-1} + R_{t-1}(\boldsymbol{e}_{t-1})\mathbf{v}_{t-1}, & \{\mathbf{v}_t\} \sim \mathrm{WN}(\mathbf{0}_3, \mathbb{I}_3), \quad (2.7b) \end{cases}$$

where $\boldsymbol{e}_t \in \mathbb{R}^5$, $\boldsymbol{\alpha} \in \mathbb{R}^3$, the matrix $B \in \mathbb{M}_{5,3}$ is provided in (2.4), $\sigma_{u,1}, \sigma_{u,2}, \sigma_{u,3} \in \mathbb{R}^+$, and the matrices $R \in \mathbb{M}_{3,5}$ and $T \in \mathbb{M}_5$ are defined as

$$R_{t-1}(\boldsymbol{e}_{t-1}) := \begin{pmatrix} \sigma_{1,t}(\boldsymbol{e}_{t-1}) & 0 & 0 \\ 0 & \sigma_{2,t}(\boldsymbol{e}_{t-1}) & 0 \\ 0 & 0 & \sigma_{3,t}(\boldsymbol{e}_{t-1}) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \quad (2.8)$$

Additionally, we set

$$Q_t := R_{t-1}(\boldsymbol{e}_{t-1})R_{t-1}(\boldsymbol{e}_{t-1})^\top. \quad (2.9)$$

The choice of matrix forms of the type (2.8)-(2.9) has two important consequences: first, the resulting model falls in the framework of the EKF, which can be hence put to work to estimate the GST. Second, as it can be seen using the EKF iterations that we spell out later on in the text, the model provides a dynamic description of the conditional variances that we are interested in forecasting, but it is static as far as the covariances is concerned. These will be modeled in a second step.

Later on in the paper we describe two specific functional dependences for the specifications $\sigma_{i,t}(\boldsymbol{e}_{t-1})$, $i \in \{1, 2, 3\}$ in (2.8) that we work with and that are consistent with the chronology with which the components $\varepsilon_{i,t}$ of the GST $\boldsymbol{\varepsilon}_t$ are quoted.

2.2 The linear and extended Kalman filters for state and parameter estimation

We now recall the linear (LKF) and extended (EKF) Kalman filters corresponding to the models (2.5a)-(2.5b) and (2.7a)-(2.7b), respectively. An in-depth treatment of this topic can be found in [Durb 12].

Let $\mathbf{r} := \{\mathbf{r}_1, \ldots, \mathbf{r}_\mathrm{T}\}$ be a sample containing T three-dimensional observed log-returns and for any $t \leq T$ denote by $\mathcal{F}_t$ the information set generated by the observed returns up to time t, that is, $\mathcal{F}_t = \sigma\left(\mathbf{r}_1, \ldots, \mathbf{r}_t\right)$. The Kalman recursions yield minimum variance linear unbiased estimates of the forecasted and updated (or filtered) state vectors and of their covariance matrices. We denote by

$\boldsymbol{\epsilon}_{t|t} := \mathrm{E}\left[\boldsymbol{e}_t|\mathcal{F}_t\right]$ (respectively, $\boldsymbol{\epsilon}_{t+1} := \mathrm{E}\left[\boldsymbol{e}_{t+1}|\mathcal{F}_t\right]$) the **updated or filtered** (respectively, **forecasted**) state vector and by $P_{t|t} := \mathrm{E}\left[\boldsymbol{e}_t\boldsymbol{e}_t^\top|\mathcal{F}_t\right]$ (respectively, $P_{t+1} := \mathrm{E}\left[\boldsymbol{e}_{t+1}\boldsymbol{e}_{t+1}^\top|\mathcal{F}_t\right]$) the corresponding covariance matrices. Additionally, let $H_t := \mathrm{E}[(\mathbf{r}_t-\boldsymbol{\alpha})(\mathbf{r}_t-\boldsymbol{\alpha})^\top \mid \mathcal{F}_{t-1}]$ be the forecasted conditional covariance matrices of the returns. The elements that we just introduced can be recursively obtained out of the Kalman recursions (see [Durb 12]) once $\boldsymbol{\epsilon}_1$ and P_1 have been provided. More specifically:

$$\mathbf{u}_t = \mathbf{r}_t - (\boldsymbol{\alpha} + B\boldsymbol{\epsilon}_t), \tag{2.10}$$

$$H_t = BP_tB^\top + \Sigma_u, \tag{2.11}$$

$$\boldsymbol{\epsilon}_{t|t} = \boldsymbol{\epsilon}_t + K_t\mathbf{u}_t, \ \text{ with } \ K_t := P_tB^\top H_t^{-1}, \tag{2.12}$$

$$P_{t|t} = P_t - K_tBP_t, \tag{2.13}$$

$$\boldsymbol{\epsilon}_{t+1} = T\boldsymbol{\epsilon}_{t|t}, \tag{2.14}$$

$$P_{t+1} = TP_{t|t}T^\top + Q_t, \ \text{ with } \ Q_t := R_t(\boldsymbol{\epsilon}_{t|t})R_t(\boldsymbol{\epsilon}_{t|t})^\top. \tag{2.15}$$

The matrix $K_t \in \mathbb{M}_{5,3}$ is referred to as the **Kalman gain**, the relations (2.12)-(2.13) are called the **updating step**, and (2.14)-(2.15) are the **prediction step** of the Kalman filter, respectively.

If the model parameters are known, the Kalman recursions make possible the filtering of the state vectors for a given observed sample. Otherwise, the model parameters can be estimated via quasi-maximum likelihood estimation using a log-likelihood constructed out of the innovations $\{\mathbf{u}_t\}_{t\in\{1,\ldots,T\}}$, namely,

$$\log L(\mathbf{r};\boldsymbol{\theta}) = -\frac{n\mathrm{T}}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{\mathrm{T}}\left[\log(\det(H_t)) + \mathbf{u}_t^\top H_t^{-1}\mathbf{u}_t\right], \tag{2.16}$$

where $\boldsymbol{\theta} \in \mathbb{R}^s$ is a vector that contains the parameters of the state-space model, and the innovations $\mathbf{u}_t$ and the covariance matrices H_t are obtained out of the observed sample $\mathbf{r} := \{\mathbf{r}_1,\ldots,\mathbf{r}_\mathrm{T}\}$ using the Kalman recursions (2.10)-(2.15). The vector $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^s$ of estimated parameters can be hence obtained by maximizing the loglikelihood function $\log L(\mathbf{r};\boldsymbol{\theta})$ in (2.16).

This optimization is subjected to various constraints that depend on the particular case that we are handling. In the linear case (2.5a)-(2.5b) the only constraint is associated to the proper identification of the model. As it is explained in Proposition 2.1, two possibilities are available: one of the elements of the vector $\boldsymbol{\beta} = (\beta_1,\beta_2,\beta_3)^\top$ that define the matrix B in (2.4) can be set equal to a constant or, alternatively, one of the unconditional variances $\sigma_{v,1}^2, \sigma_{v,2}^2, \sigma_{v,3}^2$ that define R in (2.5b) can be set equal to a positive constant. In the nonlinear case, apart from the identification constraints that we specify later on in Proposition 2.2, one has to make sure that the volatility specifications $\sigma_{i,t}(\boldsymbol{e}_{t-1})$, $i \in \{1,2,3\}$ in (2.9) yield positive values and that the resulting process has stationary solutions. This obviously depends on the specific parametric dependence chosen to define the functions $\sigma_{i,t}(\boldsymbol{e}_{t-1})$. We will consider two different models in the empirical work that we spell out in detail in the following paragraphs.

Model 1 for the conditional variances in the nonlinear state-space model. In this first model we define recursively the values $\sigma_{i,t}(\boldsymbol{e}_{t-1})$, $i \in \{1,2,3\}$, using a GARCH-type functional dependence adapted to the chronology of the GST components. We set:

$$\sigma_{1,t}^2 = a_1 + \delta_1\sigma_{3,t-1}^2 + \gamma_1\varepsilon_{3,t-1}^2, \tag{2.17}$$

$$\sigma_{2,t}^2 = a_2 + \delta_2\sigma_{1,t}^2, \tag{2.18}$$

$$\sigma_{3,t}^2 = a_3 + \delta_3\sigma_{2,t}^2. \tag{2.19}$$

In order to insure the positivity of the elements $\sigma_{i,t}^2$, the model parameters are required to satisfy the constraints $\gamma_1 \geq 0$ and $a_i > 0$, $\delta_i \geq 0$, for all $i \in \{1,2,3\}$. In order to provide sufficient conditions for the stationarity of the process, we rewrite (2.17)-(2.19) as

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \varepsilon_{3,t-1}^2, \tag{2.20}$$
$$\sigma_{2,t}^2 = (a_2 + a_1\delta_2) + \delta_1\delta_2 \sigma_{3,t-1}^2 + \gamma_1\delta_2 \varepsilon_{3,t-1}^2, \tag{2.21}$$
$$\sigma_{3,t}^2 = (a_3 + a_2\delta_3 + \alpha_1\delta_2\delta_3) + \delta_1\delta_2\delta_3 \sigma_{3,t-1}^2 + \gamma_1\delta_2\delta_3 \varepsilon_{3,t-1}^2. \tag{2.22}$$

If we think of these relations as those defining a VEC model (see [Boll 88]), stationarity can be ensured by imposing that the spectral radius of the matrix A given by

$$A := \begin{pmatrix} 0 & 0 & \delta_1 + \gamma_1 \\ 0 & 0 & \delta_2(\delta_1 + \gamma_1) \\ 0 & 0 & \delta_2\delta_3(\delta_1 + \gamma_1) \end{pmatrix} \tag{2.23}$$

is smaller than one [Gour 97]. It is easy to verify that this results in the inequality

$$\delta_2\delta_3(\delta_1 + \gamma_1) < 1, \tag{2.24}$$

which we treat later on as a nonlinear parameter constraint imposed at the time of the model estimation.

Model 2 for the conditional variances in the nonlinear state-space model. Based on the same arguments that we used for Model 1, we consider another GARCH-type variant for the functions $\sigma_{i,t}(\boldsymbol{e}_{t-1})$, $i \in \{1, 2, 3\}$, that determine the nonlinear state-space model (2.7a)-(2.7b) by allowing this time the possibility of autoregressive behavior in the volatilities and in the components of the GST. We set:

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \varepsilon_{3,t-1}^2 + \rho_1 \sigma_{1,t-1}^2 + \tau_1 \varepsilon_{1,t-1}^2, \tag{2.25}$$
$$\sigma_{2,t}^2 = a_2 + \delta_2 \sigma_{1,t}^2 + \rho_2 \sigma_{2,t-1}^2 + \tau_2 \varepsilon_{2,t-1}^2, \tag{2.26}$$
$$\sigma_{3,t}^2 = a_3 + \delta_3 \sigma_{2,t}^2 + \rho_3 \sigma_{3,t-1}^2 + \tau_3 \varepsilon_{3,t-1}^2, \tag{2.27}$$

where, again, we ensure positivity by requiring that $\gamma_1 \geq 0$ and $a_i > 0$, $\delta_i, \rho_i, \tau_i \geq 0$, for all $i \in \{1, 2, 3\}$. In order to find sufficient stationarity conditions we proceed by rewriting Model 2 as

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \varepsilon_{3,t-1}^2 + \rho_1 \sigma_{1,t-1}^2 + \tau_1 \varepsilon_{1,t-1}^2, \tag{2.28}$$
$$\sigma_{2,t}^2 = (a_2 + a_1\delta_2) + \delta_1\delta_2 \sigma_{3,t-1}^2 + \gamma_1\delta_2 \varepsilon_{3,t-1}^2 + \rho_1\delta_2 \sigma_{1,t-1}^2 + \tau_1\delta_2 \varepsilon_{1,t-1}^2 + \rho_2 \sigma_{2,t-1}^2 + \tau_2 \varepsilon_{2,t-1}^2, \tag{2.29}$$
$$\begin{aligned}\sigma_{3,t}^2 =& (a_3 + a_2\delta_3 + a_1\delta_2\delta_3) + (\delta_1\delta_2\delta_3 + \rho_3)\sigma_{3,t-1}^2 + (\gamma_1\delta_2\delta_3 + \tau_3)\varepsilon_{3,t-1}^2 + \rho_1\delta_2\delta_3 \sigma_{1,t-1}^2 \\ &+ \tau_1\delta_2\delta_3 \varepsilon_{1,t-1}^2 + \rho_2\delta_3 \sigma_{2,t-1}^2 + \tau_2\delta_3 \varepsilon_{2,t-1}^2.\end{aligned} \tag{2.30}$$

As for Model 1 we ensure stationarity by requiring that the spectral radius $\rho(A)$ of the matrix A defined by:

$$A := \begin{pmatrix} \rho_1 + \tau_1 & 0 & \delta_1 + \gamma_1 \\ \delta_2(\rho_1 + \tau_1) & \rho_2 + \tau_2 & \delta_2(\delta_1 + \gamma_1) \\ \delta_2\delta_3(\rho_1 + \tau_1) & \delta_3(\rho_2 + \tau_2) & \delta_2\delta_3(\delta_1 + \gamma_1) + \tau_3 + \rho_3 \end{pmatrix}, \tag{2.31}$$

is smaller than one. Since in this case the general expression of the eigenvalues of A is very convoluted, we take advantage of the fact that for any matrix norm $||\cdot||$ the inequality $\rho(A) \leq ||A||$ is satisfied and hence it suffices to require that $||A|| < 1$ to ensure that $\rho(A) < 1$. We implement this condition by using the so called maximum column and row sum norms (see [Horn 13]). In the case of the maximum column sum norm, the inequality $||A|| < 1$ amounts to the following three conditions

$$\begin{cases} (\rho_1 + \tau_1)(1 + \delta_2(1 + \delta_3)) < 1, & \text{(2.32a)} \\ (\rho_2 + \tau_2)(1 + \delta_2(1 + \delta_3)) < 1, & \text{(2.32b)} \\ (\delta_1 + \gamma_1)(1 + \delta_2(1 + \delta_3)) + \tau_3 + \rho_3 < 1, & \text{(2.32c)} \end{cases}$$

while the use of the maximum row sum norm results in three other different conditions, namely,

$$\begin{cases} \delta_1 + \gamma_1 + \rho_1 + \tau_1 < 1, & \text{(2.33a)} \\ \delta_2(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \rho_2 + \tau_2 < 1, & \text{(2.33b)} \\ \delta_2\delta_3(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \delta_3(\rho_2 + \tau_2) + \tau_3 + \rho_3 < 1. & \text{(2.33c)} \end{cases}$$

Any of these two sets of inequalities may be used as parameter constraints at the time of model estimation in order to ensure stationarity. Nevertheless, imposing different parameter constraints may obviously produce different estimation results.

We conclude by stating a result, whose proof can be obtained by mimicking that of Proposition 2.1, that provides conditions that ensure the proper identification of the nonlinear state-space model.

Proposition 2.2 *The nonlinear state-space model* (2.7a)-(2.7b) *is well identified with respect to the natural invariance properties spelled out in the Appendix A.1 if one of the elements of the vector* $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ *that define the matrix* B *in* (2.4) *is set equal to a constant or, alternatively, when one of the components of the vector* $\boldsymbol{a} := (a_1, a_2, a_3)^\top$ *that determine the models 1 or 2 is set equal to a positive constant.*

3 Volatility forecasting using the global stochastic trend

The objective of this section is producing one-step ahead forecasts for the covariance matrices of the observed asset returns. The state-space models that we introduced in the previous section have such a forecast naturally associated. Indeed, by expression (2.11), the conditional covariances $H_t := \mathrm{E}[(\mathbf{r}_t - \boldsymbol{\alpha})(\mathbf{r}_t - \boldsymbol{\alpha})^\top \mid \mathcal{F}_{t-1}]$ are given by $H_t = BP_tB^\top + \Sigma_u$ and, moreover, expression (2.15) provides the conditional covariance among the different components of the GST.

However, the values of the conditional covariance provided by these formulas are not compatible with the empirically observed properties of estimates of the global trend $\{\boldsymbol{e}_t\}$ or of the innovations $\{\mathbf{u}_t\}$ obtained with both the linear (2.5a)-(2.5b) and the nonlinear (2.7a)-(2.7b) state-space models. These processes are systematically heteroscedastic and exhibit time-varying correlation among its different components. However, regarding the innovations, both models capture by construction only the unconditional covariance Σ_u. As to the time-varying correlation in the GST, we analyze separately the linear and the nonlinear state-space models.

First, the linear model specification (2.5a)-(2.5b) is intrinsically homoscedastic, that is, the conditional covariances $H_t := \mathrm{E}[(\mathbf{r}_t - \boldsymbol{\alpha})(\mathbf{r}_t - \boldsymbol{\alpha})^\top \mid \mathcal{F}_{t-1}]$ and $P_t := \mathrm{E}\left[\boldsymbol{e}_t\boldsymbol{e}_t^\top \mid \mathcal{F}_{t-1}\right]$ of both the returns $\{\mathbf{r}_t\}$ and the GST associated to it are asymptotically constant in time. Indeed, the time independence of the matrices R and Σ_u implies that after a certain number of iterations, the model reaches a steady state solution in which P_t converges to the constant matrix $\overline{P}$ determined by the matrix equation (see [Durb 12, page 86]):

$$\overline{P} = T\overline{P}T^\top - T\overline{P}B^\top\overline{H}^{-1}B\overline{P}T^\top + RR^\top, \quad \text{with} \quad \overline{H} = B\overline{P}B^\top + \Sigma_u.$$

Second, in the nonlinear case (2.7a)-(2.7b), the matrix Q_t has a nontrivial time evolution and hence so does P_t. Nevertheless, a straightforward computation shows that the dynamics prescribed by either Model 1 (2.17)-(2.19) or Model 2 (3.11)-(3.13) induces a non-trivial dynamical behavior of the conditional variances of the components of the GST but makes zero the correlation between them.

These observations entail that the description of the GST associated to the models (2.5a)-(2.5b) or (2.7a)-(2.7b) is not complete enough to be used for volatility forecasting. We solve this limitation by using appropriate multivariate heteroscedastic models on both the filtered estimates $\{\boldsymbol{\epsilon}_{t|t}\}$ of the GST $\{\boldsymbol{\varepsilon}_t\}$ and on the associated residuals $\{\mathbf{u}_t\}$. This strategy comes down to using the state-space models and the Kalman filters that go with them as a first step that provides an estimation of the GST and the residuals out of the observed returns; this allows us to construct a larger filtration whose elements $\mathcal{F}_t^*$ are the pseudo-information sets generated by both the observed returns and the filtered values $\widehat{\varepsilon}_t := \boldsymbol{\epsilon}_{t|t}$ of the GST up to time t, that is,

$$\mathcal{F}_t^* := \sigma\left(\{\mathbf{r}_1, \ldots, \mathbf{r}_t\} \cup \{\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_t\}\right).$$

We refer to $\{\mathcal{F}_t^*\}$ as the **extended information set**. We subsequently estimate multivariate volatility models that are, first, designed to take into account the asynchronicity between the components of the GST and, second, are predictable with respect to $\{\mathcal{F}_t^*\}$. More specifically, we will use adapted models that will produce predictable matrix processes $\{P_t^*\}$ and $\{\Sigma_{u,t}^*\}$ with respect to $\{\mathcal{F}_t^*\}$ such that:

$$\widehat{\boldsymbol{e}}_t|_{\mathcal{F}_{t-1}^*} \sim \mathrm{WN}(\mathbf{0}_5, P_t^*) \quad \text{and} \quad \mathbf{u}_t|_{\mathcal{F}_{t-1}^*} \sim \mathrm{WN}(\mathbf{0}_3, \Sigma_{u,t}^*).$$

Combining these ingredients together with the expression (2.3), we can easily produce a forecast for the covariance H_t^* of the returns based on the information set $\mathcal{F}_{t-1}^*$:

$$H_t^* := \mathrm{E}\left[(\mathbf{r}_t - \boldsymbol{\alpha})(\mathbf{r}_t - \boldsymbol{\alpha})^\top \mid \mathcal{F}_{t-1}^*\right] = B P_t^* B^\top + \Sigma_{u,t}^*. \tag{3.1}$$

In the following paragraphs we provide the implementation details of this forecasting scheme for the linear and the nonlinear state-space models. Its empirical performance is evaluated later on in Section 4.

3.1 GST-based volatility forecasting using the nonlinear state-space model

The nonlinear state-space model (2.7a)-(2.7b) prescribes a conditionally heteroscedastic behavior on the components of the GST with respect to the extended information set $\{\mathcal{F}_t^*\}$. Indeed, the relation (2.7b) implies that

$$\mathrm{E}\left[\widehat{\varepsilon_{i,t}}^2 \mid \mathcal{F}_{t-1}^*\right] = \sigma_{i,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right), \quad \mathrm{E}\left[\widehat{\varepsilon_{i,t}}\widehat{\varepsilon_{j,t}} \mid \mathcal{F}_{t-1}^*\right] = 0, \quad \text{for any} \quad i,j \in \{1,2,3\}. \tag{3.2}$$

where the functional prescription $\sigma_{i,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right)$ is given by one of the models (2.17)-(2.19) or (3.11)-(3.13) under consideration. The second identity in (3.2) shows that this model neglects the conditional correlation between the components of the GST that is nevertheless empirically observed [Durd 14, Pere 15]. This leads us to refine the description by introducing a dynamic conditional correlation (DCC) model for the filtered GST values $\{\widehat{\varepsilon}_t\}$ constructed out of GST returns that have been standardized using the conditional covariances $\{\sigma_{i,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right)\}$, $i \in \{1,2,3\}$. This strategy introduces time-varying correlation between the components of the GST while preserving the conditional variance (3.2) captured by the non-linear state-space model.

More specifically, construct the standardized return vector $\boldsymbol{\zeta}_t \in \mathbb{R}^n$ via the component-wise assignment $\zeta_{i,t} := \widehat{\varepsilon_{i,t}}/\sigma_{i,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right)$ and let $D_t := \mathrm{diag}\left(\sigma_{1,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right), \ldots, \sigma_{n,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right)\right)$ denote the corresponding diagonal matrix of conditional standard deviations. We now specify the dynamics of the conditional correlation matrix R_t of the standardized returns $\boldsymbol{\zeta}_t$ as:

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \quad Q_t^* := \mathbb{I}_3 \odot Q_t, \tag{3.3}$$

$$Q_t = (\mathbf{i}_3\mathbf{i}_3^\top - A - B) \odot Q + A \odot (\boldsymbol{\zeta}_{t-1}\boldsymbol{\zeta}_{t-1}^\top) + B \odot Q_{t-1}, \tag{3.4}$$

where $\odot$ denotes the Hadamard (or component wise) matrix product, the parameter matrices A and B are symmetric of order 3, Q is a positive semidefinite parameter matrix of order three, and $\mathbf{i}_3$ is the column vector of three elements all equal to one. Equation (3.4) is the most general DCC prescription proposed by [Engl 02]; we call it the **Hadamard DCC model**. A simplified and much more parsimonious version of this model is the **scalar subfamily** in which all the elements of A are considered identical and likewise those of B; in that case the expression (3.4) is replaced by

$$Q_t = (1-a-b)Q + a(\boldsymbol{\zeta}_{t-1}\boldsymbol{\zeta}_{t-1}^{\top}) + bQ_{t-1}, \tag{3.5}$$

with $a, b \in \mathbb{R}^+$ such that $a + b < 1$. The matrix Q is obtained following an approximate targeting procedure that consists in assuming that $Q = \mathrm{E}\left[\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t^{\top}\right]$ and can thus be estimated by $\hat{Q} := \sum_{t=1}^{T} \boldsymbol{\zeta}_t\boldsymbol{\zeta}_t^{\top}/T$ prior to estimating the model parameters. Despite the fact that Q is not equal to the second moment matrix of $\{\boldsymbol{\zeta}_t\}$ and, as a consequence, $\hat{Q}$ is not a consistent estimator of Q (see [Aiel 13]), this targeting procedure is used in almost all applications of the DCC model which, according to simulation results in [Aiel 13], does not lead to strong biases in practice. The use of this model yields a conditional covariance matrix

$$\Sigma^*_{\widehat{\varepsilon},t} := \mathrm{E}\left[\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^{\top} \mid \mathcal{F}^*_{t-1}\right] = D_t R_t D_t, \tag{3.6}$$

and hence

$$P^*_t := \mathrm{E}\left[\widehat{\boldsymbol{e}}_t\widehat{\boldsymbol{e}}_t^{\top} \mid \mathcal{F}^*_{t-1}\right] = \left(\begin{array}{c|cc} \Sigma^*_{\widehat{\varepsilon},t} & \multicolumn{2}{c}{0} \\ \hline 0 & \widehat{\varepsilon}^2_{2,t-1} & \widehat{\varepsilon}_{2,t-1}\widehat{\varepsilon}_{3,t-1} \\ & \widehat{\varepsilon}_{2,t-1}\widehat{\varepsilon}_{3,t-1} & \widehat{\varepsilon}^2_{3,t-1} \end{array}\right) \tag{3.7}$$

Finally, we proceed analogously by modeling the conditional covariance $\{\Sigma^*_{u,t}\}$ of the residuals $\{\mathbf{u}_t\}$ by using another DCC model also based on standardized returns constructed using the conditional variances associated to GARCH models of the type 1 or 2. The use of the resulting conditional covariance matrices $\Sigma^*_{u,t}$ and P^*_t in (3.7) in the expression (3.1) leads to the required forecast for the covariance of the returns

$$H^*_{t|t-1} := \mathrm{E}[(\mathbf{r}_t - \boldsymbol{\alpha})(\mathbf{r}_t - \boldsymbol{\alpha})^{\top} \mid \mathcal{F}^*_{t-1}] = BP^*_t B^{\top} + \Sigma^*_{u,t}.$$

based on the information set $\{\mathcal{F}^*_t\}$.

3.2 GST-based volatility forecasting using the linear state-space model

As we already pointed out, the linear state-space model is intrinsically homoscedastic. We will hence proceed again by estimating appropriate DCC models on the filtered estimates $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ of the GST and on the associated residuals $\{\mathbf{u}_t\}$ but, this time, the state space model does not produce any conditional variance on the components of the GST that needs to be preserved in the correlation modeling stage.

Models for the conditional variances of the GST and the residuals. We will carry out this construction for the estimates $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ of the GST but keeping in mind that the same approach is applicable to the residuals. We proceed by using a DCC model constructed using a strategy that is reminiscent of the one used for the nonlinear state-space model case, that is, we will use standardized returns in the construction of the dynamical correlation model obtained out of conditional variances whose parametric prescription respects the chronology with which the different components $(\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$ of the GST $\boldsymbol{\varepsilon}_t$ are disclosed. An important difference with respect to the approach taken in the nonlinear state-space model is that, this time, we are not obliged to respect functional prescriptions for the conditional variances of the form $\sigma_{i,t}(\widehat{\boldsymbol{\varepsilon}_{t-1}})$ that were imposed by the use of the extended Kalman filter. In particular, we can use intraday dependences that will allow us to update more frequently the information set and hence to improve the forecasts. More specifically, we use two parameter families of conditional variance dynamics that generalize the models 1 and 2 that we used in the context of the nonlinear state-space model, namely:

- **Model 1 for the conditional variances.**

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2, \tag{3.8}$$
$$\sigma_{2,t}^2 = a_2 + \delta_2 \sigma_{1,t}^2 + \gamma_2 \widehat{\varepsilon}_{1,t}^2, \tag{3.9}$$
$$\sigma_{3,t}^2 = a_3 + \delta_3 \sigma_{2,t}^2 + \gamma_3 \widehat{\varepsilon}_{2,t}^2. \tag{3.10}$$

- **Model 2 for the conditional variances.**

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2 + \rho_1 \sigma_{1,t-1}^2 + \tau_1 \widehat{\varepsilon}_{1,t-1}^2, \tag{3.11}$$
$$\sigma_{2,t}^2 = a_2 + \delta_2 \sigma_{1,t}^2 + \gamma_2 \widehat{\varepsilon}_{1,t}^2 + \rho_2 \sigma_{2,t-1}^2 + \tau_2 \widehat{\varepsilon}_{2,t-1}^2, \tag{3.12}$$
$$\sigma_{3,t}^2 = a_3 + \delta_3 \sigma_{2,t}^2 + \gamma_3 \widehat{\varepsilon}_{2,t}^2 + \rho_3 \sigma_{3,t-1}^2 + \tau_3 \widehat{\varepsilon}_{3,t-1}^2. \tag{3.13}$$

Now, for each trading date t we single out three different **intraday extended information sets** $\mathcal{F}_{t_1}^*$, $\mathcal{F}_{t_2}^*$, and $\mathcal{F}_{t_3}^*$ defined by

$$\mathcal{F}_{t_1}^* := \sigma\left(\{\mathbf{r}_1, \ldots, \mathbf{r}_t\} \cup \{\widehat{\boldsymbol{\varepsilon}}_1, \ldots, \widehat{\boldsymbol{\varepsilon}}_{t-1}, \widehat{\varepsilon}_{t_1}\}\right), \ \mathcal{F}_{t_2}^* := \mathcal{F}_{t_1}^* \cup \sigma\left(\widehat{\varepsilon}_{t_2}\right), \ \mathcal{F}_{t_3}^* := \mathcal{F}_t^*. \tag{3.14}$$

The importance of the filtrations determined by these information sets in the context of the models that we just introduced lays in the fact that the conditional variances that they determine are such that $\sigma_{1,t}^2 = \sigma_{t_1}^2$ is $\mathcal{F}_{t_1}^*$-predictable, $\sigma_{2,t}^2 = \sigma_{t_2}^2$ is $\mathcal{F}_{t_2}^*$-predictable, and $\sigma_{3,t}^2 = \sigma_{t_3}^2$ is $\mathcal{F}_{t_3}^*$-predictable. This observation will prompt us, at the time of carrying out forecasting using these intraday information sets, to use subfamilies of the models 1 and 2 for which the entire volatility vectors $(\sigma_{1,t}, \sigma_{2,t}, \sigma_{3,t})$ are $\mathcal{F}_{t_3-1}^*$, $\mathcal{F}_{t_1}^*$, or $\mathcal{F}_{t_2}^*$-predictable.

Positivity and stationarity of the conditional variance models. Before we proceed with the implementation of a forecasting scheme using these models, we study the conditions that need to be imposed in their parameters in order to ensure that the conditional variances that they produce are positive and that they exhibit second order stationary solutions. A way to approach this question consists of thinking of the three dimensional time series $\{\boldsymbol{\varepsilon}_t\}$ as the one-dimensional process $\{\varepsilon_{t_i}\}$ obtained by ordering the components of each element $\boldsymbol{\varepsilon}_t$ according to the intraday time at which they have been disclosed. Using this point of view, the models 1 and 2 become one-dimensional GARCH models with time varying (periodic in this case) coefficients that are usually designated with the acronym tvGARCH (see [Dahl 06, Cize 09, Roha 13], and references therein). More specifically, they can be considered as tvGARCH(1,1) and tvGARCH(3,3) models, respectively, if we rewrite them as:

$$\sigma_{t_i}^2 = a_{t_i} + \delta_{t_i} \sigma_{t_i-1}^2 + \gamma_{t_i} \widehat{\varepsilon}_{t_i-1}^2, \quad \text{and} \quad \sigma_{t_i}^2 = a_{t_i} + \delta_{t_i} \sigma_{t_i-1}^2 + \gamma_{t_i} \widehat{\varepsilon}_{t_i-1}^2 + \rho_{t_i} \sigma_{t_i-3}^2 + \tau_{t_i} \widehat{\varepsilon}_{t_i-3}^2, \tag{3.15}$$

with $i \in \{1,2,3\}$, $a_{t_i} := a_i$, $\delta_{t_i} := \delta_i$, $\gamma_{t_i} := \gamma_i$, $\rho_{t_i} := \rho_i$, $\tau_{t_i} := \tau_i$, and where $t_i - 1$ and $t_i - 3$ are defined by using recursively the convention

$$t_i - 1 := \begin{cases} (t-1)_3 & \text{when} \quad i = 1, \\ t_{i-1} & \text{when} \quad i \in \{2,3\}. \end{cases}$$

The positivity of the conditional variances implied by these models can be obtained by using only positive coefficients in the expressions that define them. Regarding stationarity, a sufficient condition of widespread use in the tvGARCH context (see for example [Roha 13]) is that $\delta_{t_i} + \gamma_{t_i} < 1$ for the model 1, and that $\delta_{t_i} + \gamma_{t_i} + \rho_{t_i} + \tau_{t_i} < 1$ for the model 2, with $i \in \{1,2,3\}$. Numerical experiments show that, in our particular situation, these conditions lack sharpness and produce mediocre estimation results. In the following proposition, whose proof is provided in Appendix A.2, we establish less restrictive stationarity solutions that take advantage of the periodicity of the GARCH coefficients. In order to formulate them, we need to introduce the matrices A_{t_i} associated to the Markov representations of the recursions in (3.15) corresponding to the second model, as well as their expectations $A_i := \mathrm{E}\left[A_{t_i}\right]$ (see Section 2.2.2 in [Fran 10] for the details). Let $\{\mathbf{v}_t\} \sim \mathrm{WN}(\mathbf{0}_3, \mathbb{I}_3)$ be the innovations introduced in the definition of the state-space model (2.5b). Then:

$$A_{t_i} := \begin{pmatrix} \gamma_{t_i} v_{t_i}^2 & 0 & \tau_{t_i} v_{t_i}^2 & \delta_{t_i} v_{t_i}^2 & 0 & \rho_{t_i} v_{t_i}^2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \gamma_{t_i} & 0 & \tau_{t_i} & \delta_{t_i} & 0 & \rho_{t_i} \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad A_i := \mathrm{E}\left[A_{t_i}\right] = \begin{pmatrix} \gamma_i & 0 & \tau_i & \delta_i & 0 & \rho_i \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \gamma_i & 0 & \tau_i & \delta_i & 0 & \rho_i \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Proposition 3.1 *Consider the GARCH models with time-varying coefficients defined by the recursions in the expression (3.15). If the innovations $\{\mathbf{v}_t\}$ that drive them are independent, then the following conditions imply the existence of a unique periodic (with period equal to three) stationary solution:*

(i) *For Model 1:* $(\delta_1 + \gamma_1)(\delta_2 + \gamma_2)(\delta_3 + \gamma_3) < 1$.

(ii) *For Model 2:* $\rho(A_3 A_2 A_1) < 1$, *where* $\rho(\cdot)$ *denotes the spectral radius.*

The stationarity condition $\rho(A_3 A_2 A_1) < 1$ for Model 2 cannot be implemented as such at the time of estimation due to the convoluted analytic expression of the spectrum of $A_3 A_2 A_1$. Indeed, a straightforward computation shows that:

$$A_3 A_2 A_1 = \begin{pmatrix} \gamma_1\zeta_2\zeta_3 + \tau_3 & \tau_2\zeta_3 & \tau_1\zeta_2\zeta_3 & \delta_1\zeta_2\zeta_3 + \rho_3 & \rho_2\zeta_3 & \rho_1\zeta_2\zeta_3 \\ \gamma_1\zeta_2 & \tau_2 & \tau_1\zeta_2 & \delta_1\zeta_2 & \rho_2 & \rho_1\zeta_2 \\ \gamma_1 & 0 & \tau_1 & \delta_1 & 0 & \rho_1 \\ \gamma_1\zeta_2\zeta_3 + \tau_3 & \tau_2\zeta_3 & \tau_1\zeta_2\zeta_3 & \delta_1\zeta_2\zeta_3 + \rho_3 & \rho_2\zeta_3 & \rho_1\zeta_2\zeta_3 \\ \gamma_1\zeta_2 & \tau_2 & \tau_1\zeta_2 & \delta_1\zeta_2 & \rho_2 & \rho_1\zeta_2 \\ \gamma_1 & 0 & \tau_1 & \delta_1 & 0 & \rho_1 \end{pmatrix},$$

with $\zeta_i := \delta_i + \gamma_i$, $i = \{2, 3\}$. Therefore, as we already did for (2.31), we take advantage of the fact that for any matrix norm $||\cdot||$ the inequality $\rho(A_3 A_2 A_1) \leq ||A_3 A_2 A_1||$ is satisfied and hence it suffices to require that $||A_3 A_2 A_1|| < 1$ to ensure that $\rho(A_3 A_2 A_1) < 1$. We implement this condition by using, for example, the maximum row sum norm (see [Horn 13]), in which case the inequality $||A|| < 1$ amounts to the following three conditions:

$$\begin{cases} \delta_1 + \gamma_1 + \rho_1 + \tau_1 < 1, & \text{(3.16a)} \\ (\delta_2 + \gamma_2)(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \rho_2 + \tau_2 < 1, & \text{(3.16b)} \\ (\delta_3 + \gamma_3)((\delta_2 + \gamma_2)(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \rho_2 + \tau_2) + \tau_3 + \rho_3 < 1. & \text{(3.16c)} \end{cases}$$

Forecasting using the intraday extended information sets. In order to forecast using the different intraday information sets, we start by estimating heteroscedastic models of the type 1 or 2 on the filtered GST $\{\widehat{\varepsilon}_t\}$ and on the corresponding residuals $\{\mathbf{u}_t\}$, that yield predictable variance processes with respect to them. More specifically, we distinguish three cases:

(i) Forecasting with respect to $\mathcal{F}^*_{t_2}$: when we carry this out, we assume that the index returns $r_{1,t}$ and $r_{2,t}$ quoted at the instants t_1 and t_2 of day t have already been observed, that the corresponding GST returns $\widehat{\varepsilon}_{1,t}$, and $\widehat{\varepsilon}_{2,t}$ have been filtered, and that the corresponding residuals $u_{1,t}$ and $u_{2,t}$ are hence available. If we construct models of the type 1 or 2 for $\{\widehat{\varepsilon}_t\}$ and $\{\mathbf{u}_t\}$ we obtain $\mathcal{F}^*_{t_3}$-predictable conditional variance processes $\{\boldsymbol{\sigma}^{\widehat{\boldsymbol{\varepsilon}}}_t\}$ and $\{\boldsymbol{\sigma}^{\mathbf{u}}_t\}$, that is, for any fixed day t, the components of $\boldsymbol{\sigma}^{\widehat{\boldsymbol{\varepsilon}}}_t$ and $\boldsymbol{\sigma}^{\mathbf{u}}_t$ are $\mathcal{F}^*_{t_3-1} = \mathcal{F}^*_{t_2}$-measurable and

$$\mathrm{var}\left(\widehat{\boldsymbol{\varepsilon}}_t \mid \mathcal{F}^*_{t_2}\right) = \mathrm{diag}\left(\widehat{\varepsilon}^2_{1,t}, \widehat{\varepsilon}^2_{2,t}, \sigma^{\widehat{\boldsymbol{\varepsilon}}\,2}_{3,t}\right) \quad \text{and} \quad \mathrm{var}\left(\mathbf{u}_t \mid \mathcal{F}^*_{t_2}\right) = \mathrm{diag}\left(u^2_{1,t}, u^2_{2,t}, \sigma^{\mathbf{u}\,2}_{3,t}\right).$$

(ii) Forecasting with respect to $\mathcal{F}^*_{t_1}$: in this case we use models of the type 1 or 2 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ and $\{\mathbf{u}_t\}$ but we fix the coefficient $\gamma_3 = 0$. These restrictions produce $\mathcal{F}^*_{t_2}$-predictable conditional variance processes $\{\boldsymbol{\sigma}^{\widehat{\boldsymbol{\varepsilon}}}_t\}$ and $\{\boldsymbol{\sigma}^{\mathbf{u}}_t\}$ for which

$$\mathrm{var}\left(\widehat{\boldsymbol{\varepsilon}}_t \mid \mathcal{F}^*_{t_1}\right) = \mathrm{diag}\left(\widehat{\varepsilon}^2_{1,t}, \sigma^{\widehat{\boldsymbol{\varepsilon}}\,2}_{2,t}, \sigma^{\widehat{\boldsymbol{\varepsilon}}\,2}_{3,t}\right) \quad \text{and} \quad \mathrm{var}\left(\mathbf{u}_t \mid \mathcal{F}^*_{t_1}\right) = \mathrm{diag}\left(u^2_{1,t}, \sigma^{\mathbf{u}\,2}_{2,t}, \sigma^{\mathbf{u}\,2}_{3,t}\right).$$

(iii) Forecasting with respect to $\mathcal{F}^*_{(t-1)_3} = \mathcal{F}^*_{t-1}$: in this case we use models of the type 1 or 2 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ and $\{\mathbf{u}_t\}$ but we fix the coefficients $\gamma_3 = \gamma_2 = 0$. These restrictions produce $\mathcal{F}^*_{t_1}$-predictable conditional variance processes $\{\boldsymbol{\sigma}^{\widehat{\boldsymbol{\varepsilon}}}_t\}$ and $\{\boldsymbol{\sigma}^{\mathbf{u}}_t\}$ for which

$$\mathrm{var}\left(\widehat{\boldsymbol{\varepsilon}}_t \mid \mathcal{F}^*_{(t-1)_3}\right) = \mathrm{diag}\left(\sigma^{\widehat{\boldsymbol{\varepsilon}}\,2}_{1,t}, \sigma^{\widehat{\boldsymbol{\varepsilon}}\,2}_{2,t}, \sigma^{\widehat{\boldsymbol{\varepsilon}}\,2}_{3,t}\right) \quad \text{and} \quad \mathrm{var}\left(\mathbf{u}_t \mid \mathcal{F}^*_{(t-1)_3}\right) = \mathrm{diag}\left(\sigma^{\mathbf{u}\,2}_{1,t}, \sigma^{\mathbf{u}\,2}_{2,t}, \sigma^{\mathbf{u}\,2}_{3,t}\right).$$

We subsequently use these conditional variance processes to construct standardized vectors out of which we specify DCC models as in (3.3)-(3.4) or in (3.5). This procedure yields conditional covariance matrices $\{H^{\widehat{\boldsymbol{\varepsilon}}}_t\}$ and $\{H^{\mathbf{u}}_t\}$ that have the same predictability properties as the conditional variance processes used to construct them. More specifically, they are in general $\mathcal{F}^*_{t_2}$-measurable, if $\gamma_3 = 0$ in the conditional variance models then they are $\mathcal{F}^*_{t_1}$-measurable, and if $\gamma_2 = \gamma_3 = 0$, then they are $\mathcal{F}^*_{(t-1)_3}$-measurable. Additionally, these covariance matrices are such that

$$\widehat{\boldsymbol{\varepsilon}}_t = L^{\widehat{\boldsymbol{\varepsilon}}}_t \boldsymbol{\xi}^{\widehat{\boldsymbol{\varepsilon}}}_t \quad \text{and} \quad \mathbf{u}_t = L^{\mathbf{u}}_t \boldsymbol{\xi}^{\mathbf{u}}_t, \quad \text{with} \quad \{\boldsymbol{\xi}^{\widehat{\boldsymbol{\varepsilon}}}_t\}, \{\boldsymbol{\xi}^{\mathbf{u}}_t\} \sim \mathrm{IN}\left(\mathbf{0}, \mathbb{I}_3\right),$$

and $L^{\widehat{\boldsymbol{\varepsilon}}}_t$ (respectively, $L^{\mathbf{u}}_t$) the lower triangular Cholesky factor of $H^{\widehat{\boldsymbol{\varepsilon}}}_t$ (respectively $H^{\mathbf{u}}_t$), that is, $H^{\widehat{\boldsymbol{\varepsilon}}}_t = L^{\widehat{\boldsymbol{\varepsilon}}}_t L^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t$ (respectively, $H^{\mathbf{u}}_t = L^{\mathbf{u}}_t L^{\mathbf{u}\,\top}_t$). The lower triangularity of $L^{\widehat{\boldsymbol{\varepsilon}}}_t$ and $L^{\mathbf{u}}_t$ imply that the extended intraday information sets generated by the components of $\widehat{\boldsymbol{\varepsilon}}_t$ are identical to those spanned by the corresponding entries of $\boldsymbol{\xi}^{\widehat{\boldsymbol{\varepsilon}}}_t$. This important observation allows us to explicitly write down in the following paragraphs covariance forecasting formulas with respect to the different intraday extended information sets.

(i) Covariance forecasting with respect to $\mathcal{F}^*_{t_2}$ using models 1 or 2 with no restrictions. In that case:

$$\begin{aligned} \mathrm{E}\left[\widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top \mid \mathcal{F}^*_{t_2}\right] &= \mathrm{E}\left[L^{\widehat{\boldsymbol{\varepsilon}}}_t \boldsymbol{\xi}^{\widehat{\boldsymbol{\varepsilon}}}_t \boldsymbol{\xi}^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t L^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t \mid \mathcal{F}^*_{t_2}\right] = L^{\widehat{\boldsymbol{\varepsilon}}}_t \mathrm{E}\left[\boldsymbol{\xi}^{\widehat{\boldsymbol{\varepsilon}}}_t \boldsymbol{\xi}^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t \mid \mathcal{F}^*_{t_2}\right] L^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t \\ &= L^{\widehat{\boldsymbol{\varepsilon}}}_t \begin{pmatrix} \xi^{\widehat{\boldsymbol{\varepsilon}}\,2}_{1,t} & \xi^{\widehat{\boldsymbol{\varepsilon}}}_{1,t}\xi^{\widehat{\boldsymbol{\varepsilon}}}_{2,t} & 0 \\ \xi^{\widehat{\boldsymbol{\varepsilon}}}_{1,t}\xi^{\widehat{\boldsymbol{\varepsilon}}}_{2,t} & \xi^{\widehat{\boldsymbol{\varepsilon}}\,2}_{2,t} & 0 \\ 0 & 0 & 1 \end{pmatrix} L^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t =: \Sigma^*_{\widehat{\boldsymbol{\varepsilon}},t|t_2}. \end{aligned} \tag{3.17}$$

Analogously,

$$\mathrm{E}\left[\mathbf{u}_t \mathbf{u}_t^\top \mid \mathcal{F}^*_{t_2}\right] = L^{\mathbf{u}}_t \begin{pmatrix} \xi^{\mathbf{u}\,2}_{1,t} & \xi^{\mathbf{u}}_{1,t}\xi^{\mathbf{u}}_{2,t} & 0 \\ \xi^{\mathbf{u}}_{1,t}\xi^{\mathbf{u}}_{2,t} & \xi^{\mathbf{u}\,2}_{2,t} & 0 \\ 0 & 0 & 1 \end{pmatrix} L^{\mathbf{u}\,\top}_t =: \Sigma^*_{\mathbf{u},t|t_2}. \tag{3.18}$$

(ii) Covariance forecasting with respect to $\mathcal{F}^*_{t_1}$ using models 1 or 2 with the restriction $\gamma_3 = 0$. In that case:

$$\mathrm{E}\left[\widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top \mid \mathcal{F}^*_{t_1}\right] = L^{\widehat{\boldsymbol{\varepsilon}}}_t \begin{pmatrix} \xi^{\widehat{\boldsymbol{\varepsilon}}\,2}_{1,t} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} L^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t =: \Sigma^*_{\widehat{\boldsymbol{\varepsilon}},t|t_1}, \tag{3.19}$$

$$\mathrm{E}\left[\mathbf{u}_t \mathbf{u}_t^\top \mid \mathcal{F}^*_{t_1}\right] = L^{\mathbf{u}}_t \begin{pmatrix} \xi^{\mathbf{u}\,2}_{1,t} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} L^{\mathbf{u}\,\top}_t =: \Sigma^*_{\mathbf{u},t|t_1}. \tag{3.20}$$

(iii) Covariance forecasting with respect to $\mathcal{F}^*_{(t-1)_3}$ using models 1 or 2 with the restrictions $\gamma_2 = \gamma_3 = 0$. In that case:

$$\mathrm{E}\left[\widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top \mid \mathcal{F}^*_{(t-1)_3}\right] = L^{\widehat{\boldsymbol{\varepsilon}}}_t L^{\widehat{\boldsymbol{\varepsilon}}\,\top}_t = H^{\widehat{\boldsymbol{\varepsilon}}}_t =: \Sigma^*_{\widehat{\boldsymbol{\varepsilon}},t|(t-1)_3}, \tag{3.21}$$

$$\mathrm{E}\left[\mathbf{u}_t \mathbf{u}_t^\top \mid \mathcal{F}^*_{(t-1)_3}\right] = L^{\mathbf{u}}_t L^{\mathbf{u}\,\top}_t = H^{\mathbf{u}}_t =: \Sigma^*_{\mathbf{u},t|(t-1)_3}. \tag{3.22}$$

A straightforward computation shows that in all three cases, these forecasts of the covariance matrices of the processes $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ and $\{\mathbf{u}_t\}$ yield the following forecasts for the covariance matrices $H^*_{t|t_i}$ of the returns based on the different intraday extended information sets:

$$H^*_{t|t_i} := \mathrm{E}[(\mathbf{r}_t - \boldsymbol{\alpha})(\mathbf{r}_t - \boldsymbol{\alpha})^\top \mid \mathcal{F}^*_{t_i}] = BP^*_{t|t_i}B^\top + \Sigma^*_{u,t|t_i}, \tag{3.23}$$

where

$$P^*_{t|t_i} := \mathrm{E}\left[\widehat{\boldsymbol{e}}_t \widehat{\boldsymbol{e}}_t^{\,\top} \mid \mathcal{F}^*_{t_i}\right] = \left(\begin{array}{c|cc} \Sigma^*_{\widehat{\varepsilon},t|t_i} & \multicolumn{2}{c}{0} \\ \hline 0 & \widehat{\varepsilon}^2_{2,t-1} & \widehat{\varepsilon}_{2,t-1}\widehat{\varepsilon}_{3,t-1} \\ & \widehat{\varepsilon}_{2,t-1}\widehat{\varepsilon}_{3,t-1} & \widehat{\varepsilon}^2_{3,t-1} \end{array}\right), \tag{3.24}$$

$\Sigma^*_{\widehat{\varepsilon},t|t_i}$ is given by (3.17), (3.19), or (3.21), and $\Sigma^*_{u,t|t_i}$ is provided in (3.18), (3.20), and (3.22).

4 Empirical performance of the GST-based volatility forecasting schemes

In this section we carry out an empirical study in order to evaluate the one-day ahead volatility forecasting performances of the proposed linear and nonlinear state-space models concerning the log-returns of three major market indices with non-synchronous closing times.

4.1 Dataset and competing models

Dataset. We use as dataset the daily closing values of three major stock market indices, namely, NIKKEI 225, FTSE, and S&P500[5]. These markets are geographically located in different time zones and have asynchronous closing times: NIKKEI 225 is an index based on the quotes of the Tokyo Stock Exchange that closes at 6:00 UTC. FTSE and S&P500 are based on the quotes of the London and the New York stock exchanges that close at 16:30 and 21:00 UTC, respectively. The closing values are adjusted for dividend payments and stock splits and the resulting data is synchronized by taking into account all the holidays of the different markets. The daily log-returns for the three indices are computed between January 5, 1996 and April 1, 2015 which yields a dataset with $T := 4581$ observations. The

[5]Data were downloaded from the Yahoo Finance database

whole log-returns sample is demeaned and it is divided into two parts. The first one corresponds to the period between January 5, 1996 and April 1, 2013; it has length $T_{\text{est}} := 4095$ and it is reserved for estimation purposes. The remaining $T_{\text{out}} := 486$ observations from April 2, 2013 to April 1, 2015 are reserved for an out-of-sample study consisting on one-day ahead volatility forecasting.

In order to illustrate the robustness of the results obtained in this empirical study, we have included in the technical appendix in Section A.3 a similar analysis based on the same dataset but using a shorter estimation period (January 5, 1996 – December 4, 2006) that does not contain the volatility events in the Fall 2008. The out-of-sample study in that case comprises the entire Great Recession (December 5, 2006 – April 1, 2015).

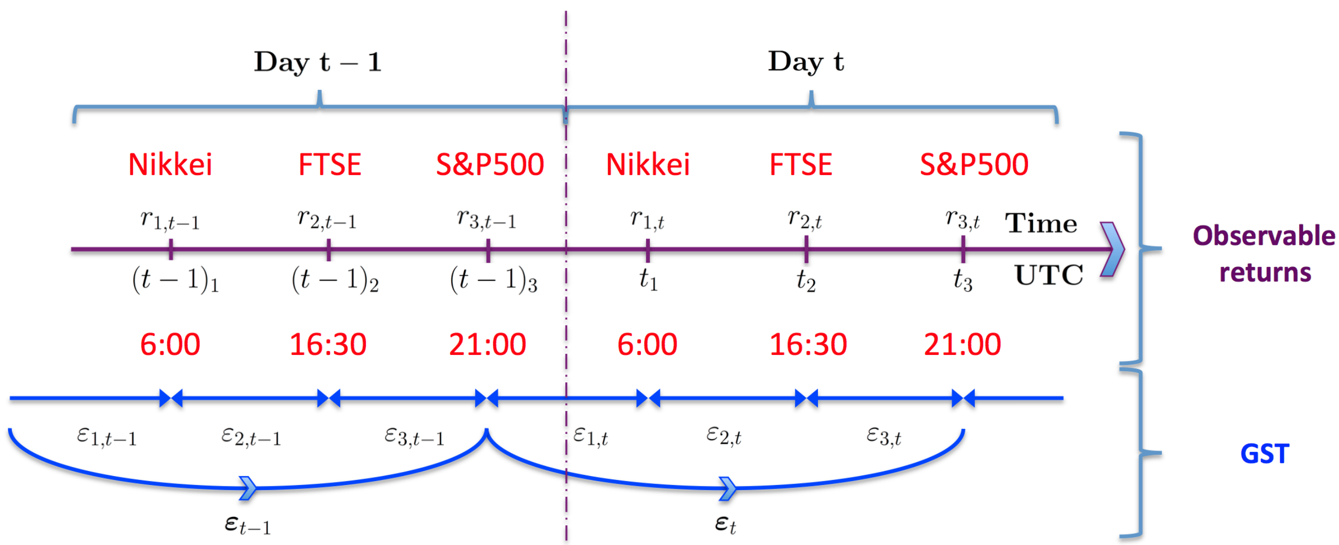

Figure 2: Diagram representing the different variables, time labels, and chronology corresponding to the indices used in the empirical study.

Competing models. We consider three groups of models whose comparative one-day ahead volatility forecasting performance will be assessed, namely:

(i) Linear state space model combined with Models 1 and 2 (LSS Model 1/Model 2): We use the linear state-space setup (2.5a)-(2.5b) and we estimate the components $\{\widehat{\varepsilon}_t\}$ of the GST via the Kalman recursions (2.10)-(2.15), together with the model parameters by minimizing minus the log-likelihood function provided in (2.16) associated with the considered T_{est} in-sample observations. The model has been properly identified using Proposition 2.1 by setting $\beta_1 = 1$. We proceed by estimating Hadamard DCC models on the filtered estimates $\{\widehat{\varepsilon}_t\}$ of the GST and on the associated linear state-space model residuals $\{\mathbf{u}_t\}$. In order to construct the standardized returns for the estimates $\{\widehat{\varepsilon}_t\}$ that are needed in the first step of the DCC estimation, we use either Model 1 in (3.8)-(3.10) or Model 2 in (3.11)-(3.13). These dynamical prescriptions model the conditional variances of the GST components taking into account the particular temporal dependance between them due to the chronology of the market closings they originate from. In this particular empirical experiment, the conditional variances of the residuals $\{\mathbf{u}_t\}$ are modeled using standard individual GARCH(1,1) models even though an approach analogous to the one followed for the GST estimates based on Model 1 or Model 2 could be adopted. Hadamard DCC models of the type (3.4) are used at the time of modeling the conditional variances of both for the

GST components $\{\widehat{\varepsilon}_t\}$ and of the residuals $\{\mathbf{u}_t\}$. These are subsequently used in the construction the one-day ahead forecast of the covariance matrix (3.1) of the indices log-returns.

(ii) Nonlinear state space model combined with Models 1 (NSS Model 1): A procedure identical to the one presented in the previous point is followed but, in this case, using the nonlinear state-space model (2.7a)-(2.7b) for the GST components $\{\widehat{\varepsilon}_t\}$ in which the Model 1 (2.17)-(2.19) prescribes their conditional variances using a GARCH-type functional dependence adapted to their chronology. The model has been properly identified using Proposition 2.2 by setting $\beta_1 = 1$. As in the previous case, two Hadamard DCC models are then used for both the filtered $\{\widehat{\varepsilon}_t\}$ and the residuals $\{\mathbf{u}_t\}$. The deGARCHing of the GST components is performed by using the conditional deviations implied by the Model 1 in (2.17)-(2.19). Again, as in the linear state-space model case, a standard GARCH(1,1) model prescription is used to determine the conditional variances of the residuals $\{\mathbf{u}_t\}$ that is subsequently used to standardize them.

(iii) Scalar and Hadamard DCC models: These families of models are designed and widely used for volatility forecasting and we hence choose them to serve as a benchmark for the forecasting tasks that we perform in this empirical section. We proceed in a standard way by constructing and estimating both scalar and Hadamard three dimensional models that use exclusively the daily quoted information on the closing values of the indices under consideration and ignore their non-synchronicity. The first stage of the model construction is common for both the scalar and the Hadamard setups. The deGARCHing of the returns is accomplished with the conditional deviations provided by standard one-dimensional GARCH(1,1) models that are estimated on the individual daily returns. The estimation procedure in the case of the scalar DCC (3.5) is straightforward (see for instance [Engl 09]), while the Hadamard prescription (3.4) presents some complications due to the presence of positivity constraints to which the model parameters are subjected and that we handle using the tools presented in [Bauw 15].

4.2 Volatility forecasting study

We now carry out volatility forecasting using the T_{out} observations reserved for the out-of-sample study following the indications provided in Section 3. The one-step ahead volatility forecast is constructed in a different way for each groups of models. More specifically,

(i) Volatility forecasting with the scalar/Hadamard DCC models: the value of the one-day ahead forecast of the conditional covariance matrix H_t of the returns $(r_{1,t}, r_{2,t}, r_{3,t})$, $t \in \{T_{\text{est}} + 1, \ldots, T_{\text{est}} + T_{\text{out}}\}$, with respect to the information set $\mathcal{F}_{t-1}$ is computed by setting $H_t := D_t R_t D_t$, with R_t given by (3.3)-(3.4) and $D_t := \text{diag}(\sigma_{1,t}, \sigma_{2,t}, \sigma_{3,t})$ a diagonal matrix containing the conditional standard deviations obtained out of the GARCH(1,1) model that has been previously fit to the log-returns during the first stage of the DCC model construction.

(ii) GST based volatility forecasting with the nonlinear state-space models: the one-day ahead forecast for the conditional covariance H_t^* of the returns $(r_{1,t}, r_{2,t}, r_{3,t})$ with respect to the extended information set $\mathcal{F}_{t-1}^*$ is obtained out of the relation (3.1), namely $H_{t|t-1}^* := BP_t^*B^\top + \Sigma_{u,t}^*$. In this expression B is the parameter matrix of the observation equation (2.7a) in the nonlinear state-space model, P_t^* is the forecast with respect to $\mathcal{F}_{t-1}^*$ of the covariance matrix of the state variables provided in (3.7), and $\Sigma_{u,t}^*$ is the forecast of the covariance matrix of the nonlinear state-space model residuals $(u_{1,t}, u_{2,t}, u_{3,t})$ with respect to the same information set. The covariance matrix P_t^* in (3.7) is computed using the relation (3.6) and taking into account the specific prescription imposed by Model 1 (2.20)-(2.22) on the elements of the diagonal matrix $D_t := \text{diag}\left(\sigma_{1,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right), \sigma_{2,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right), \sigma_{3,t}\left(\widehat{\boldsymbol{e}_{t-1}}\right)\right)$.

(iii) GST based volatility forecasting with the linear state-space models: Since in the linear setup the extended information sets can be updated more frequently, we use this feature to improve the forecasts. As it has been discussed earlier, we distinguish three separate cases in the context of the LSS Model 1/Model 2 depending on the extended intraday information sets involved:

1. **Forecasting with respect to extended information set** $\mathcal{F}^*_{t_2}$: Estimation of models 1 and 2 subjected to no parametric restrictions. The forecast $H^*_{t|t_2}$ is computed using the expression (3.23) together with (3.17) and (3.18).
2. **Forecasting with respect to extended information set** $\mathcal{F}^*_{t_1}$: Estimation of models 1 and 2 is subjected to the parametric restriction $\gamma_3 = 0$. The forecast $H^*_{t|t_1}$ is computed using the expression (3.23) together with (3.19) and (3.20).
3. **Forecasting with respect to extended information set** $\mathcal{F}^*_{(t-1)_3} = \mathcal{F}^*_{t-1}$: Estimation of models 1 and 2 is subjected to the parametric restrictions $\gamma_2 = \gamma_3 = 0$. The forecast $H^*_{t|t-1}$ is computed using the expression (3.23) together with (3.21) and (3.22).

4.3 Model confidence sets based on covariance and KLIC loss functions

The different models are compared using the model confidence set (MCS) approach introduced in [Hans 03, Hans 11] with loss functions that involve the daily log-returns of the three indices under consideration and the forecasts of the conditional covariance matrices associated to each of the models under consideration.

Covariance loss functions. We will use three different covariance loss functions in the implementation of the MCS approach depending on the specific intraday extended information set used at the time of forecasting, namely:

$$d^{\text{cov}}_{\mathcal{F}^*_{t_2}} := \left(r^2_{3,t} - h_{33,t}\right)^2, \tag{4.1}$$

$$d^{\text{cov}}_{\mathcal{F}^*_{t_1}} := \frac{1}{3} \sum_{i \leq j = 2,3} \left(r_{i,t} r_{j,t} - h_{ij,t}\right)^2, \tag{4.2}$$

$$d^{\text{cov}}_{\mathcal{F}^*_{(t-1)_3}} := \frac{1}{6} \sum_{i \leq j = 1,2,3} \left(r_{i,t} r_{j,t} - h_{ij,t}\right)^2, \tag{4.3}$$

where $t \in \{T_{\text{est}} + 1, \ldots, T_{\text{est}} + T_{\text{out}}\}$ and $h_{ij,t}$ are the (i, j)-entries of the corresponding model dependent forecasts for the conditional covariance matrices at t. More specifically, when using the scalar/Hadamard DCC model, we will consider the conditional covariance matrix H_t at t with respect to the information set $\mathcal{F}_{t-1}$. In the nonlinear state-space model case, we will consider $H^*_{t|t-1}$ associated to $\mathcal{F}^*_{t-1}$ and determined by (3.1). Finally, when dealing with the linear state-space model case, we will use the forecasts $H^*_{t|t_i}$ determined by (3.23) and based on the different intraday extended information sets.

The MCS approach identifies, from a set of competing models, the subset of models that are equivalent in terms of out-of-sample conditional covariance predictive ability and which outperform all the other models at a considered significance level α for the so called equivalence test. We set this significance level at 10% and 25%, and use 100 000 block bootstrap replicates with block length two in order to obtain the distribution of the relevant test statistic under the null of the equal predictive ability.

Tables 4.1 and 4.2 contain the MCS results associated to the values of the covariance loss functions obtained in 36 different out-of-sample time intervals of the form $\{T_{\text{est}} + 1, \ldots, T_{\text{est}} + 136 + 10k\}$ with , $k = \{0, 1, \ldots, 35\}$. The first 136 elements in the out-of-sample period are included in all these intervals in order to ensure that there are enough values available for the bootstrapping process that is necessary in the estimation of the distribution of the model equivalence test statistic. The date corresponding to the end of this offset interval is October 18, 2013.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3 = 0$)	Model 1 ($\gamma_3, \gamma_2 = 0$)	Model 2	Model 2 ($\gamma_3 = 0$)	Model 2 ($\gamma_3, \gamma_2 = 0$)	Model 1
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_2}}$	4	4	**36**	8	4	36	8	4	0
Sum p-values	1.0384	1.0384	**36.0000**	1.8603	1.2836	16.6883	1.7096	1.2841	0
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_1}}$	4	4	—	32	8	—	22	**36**	0
Sum p-values	1.4948	1.4948	—	29.5786	2.2696	—	9.8610	**31.1728**	0.0013
MCS $d^{\text{cov}}_{\mathcal{F}^*_{(t-1)_3}}$	34	34	—	—	34	—	—	**36**	2
Sum p-values	16.3389	17.2042	—	—	11.9633	—	—	**35.2726**	0.4470

Table 4.1: Model confidence sets (MCS) constructed using the covariance based loss functions (4.1), (4.2), and (4.3), respectively, for 36 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\text{est}} + 136 + 10k$, $k \in \{0, 1, \ldots, 35\}$. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at a 90% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 36 tests. The best performing models for the considered information set are marked in bold red.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3 = 0$)	Model 1 ($\gamma_3, \gamma_2 = 0$)	Model 2	Model 2 ($\gamma_3 = 0$)	Model 2 ($\gamma_3, \gamma_2 = 0$)	Model 1
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_2}}$	0	0	**36**	0	0	16	0	0	0
Sum p-values	1.0579	1.0579	**36.0000**	1.8893	1.3104	9.5121	1.7355	1.3109	0
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_1}}$	0	0	—	29	0	—	16	**36**	0
Sum p-values	1.4963	1.4963	—	29.5771	2.2736	—	9.8608	**31.1885**	0.0013
MCS $d^{\text{cov}}_{\mathcal{F}^*_{(t-1)_3}}$	22	22	—	—	22	—	—	**36**	1
Sum p-values	16.3370	17.1966	—	—	11.9473	—	—	**35.2788**	0.4492

Table 4.2: Model confidence sets (MCS) constructed using the covariance based loss functions (4.1), (4.2), and (4.3), respectively, for 36 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\text{est}} + 136 + 10k$, $k \in \{0, 1, \ldots, 35\}$. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at a 75% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 36 tests. The best performing models for the considered information set are marked in bold red.

These tables report, for each model, the number of times that it is included in the model confidence set with the a significance level of 10% or 25%. The second figure represents the total sum of the 36 obtained MCS p-values corresponding to each model. Figures 3-5 depict the evolution of the MCS p-values when the number of the out-of-sample observations considered grows with a step equal to ten observations. These results show that:

(i) The group of Kalman-based models significantly outperforms the standard DCC models regardless the information sets involved.

(ii) The linear state-space models significantly outperform their nonlinear state-space counterparts regardless the information sets involved.

The appendix A.3 contains the results of an analogous experiment with a shorter estimation period that does not contain the high volatility events of the Fall 2008 period. In that situation the empirical study shows that: first, the conclusion **(i)** above holds and, second, **the nonlinear state-space model based approach outperforms the linear one when the information set involved is the largest available**.

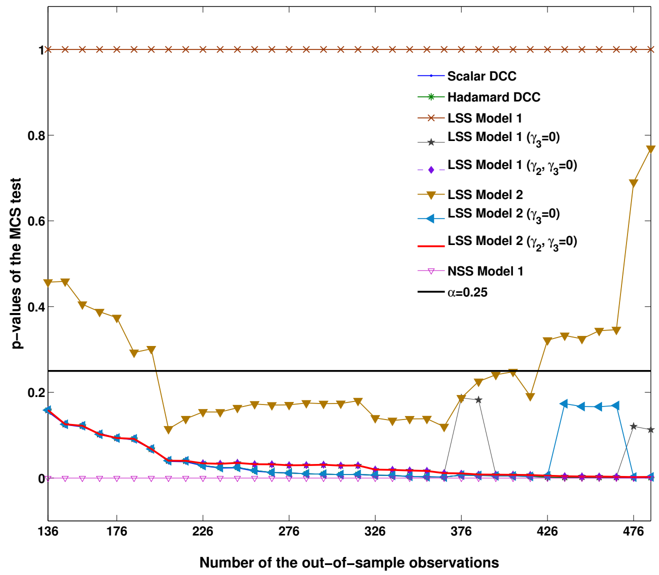

Figure 3: Evolution of the p-values of the MCS test based on the covariance loss function $d^{\text{cov}}_{\mathcal{F}^*_{t_2}}$ in (4.1) in terms of the out-of-sample length. The significance level of the MCS test is $\alpha = 0.25$.

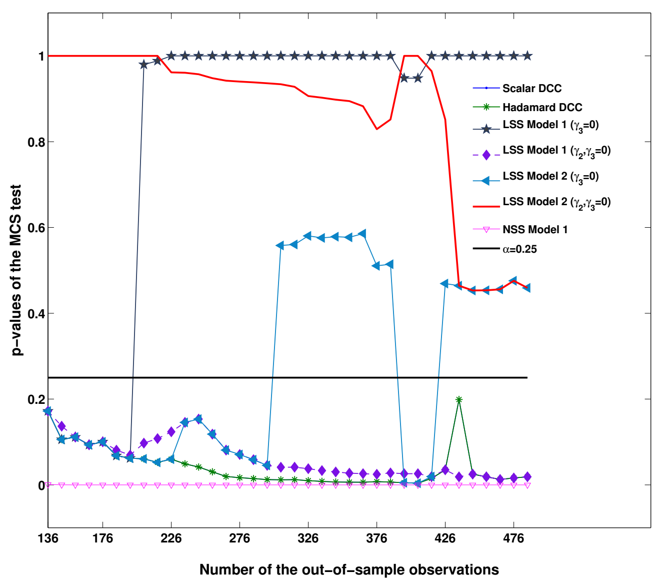

Figure 4: Evolution of the p-values of the MCS test based on the covariance loss function $d^{\text{cov}}_{\mathcal{F}^*_{t_1}}$ in (4.2) in terms of the out-of-sample length. The significance level of the MCS test is $\alpha = 0.25$.

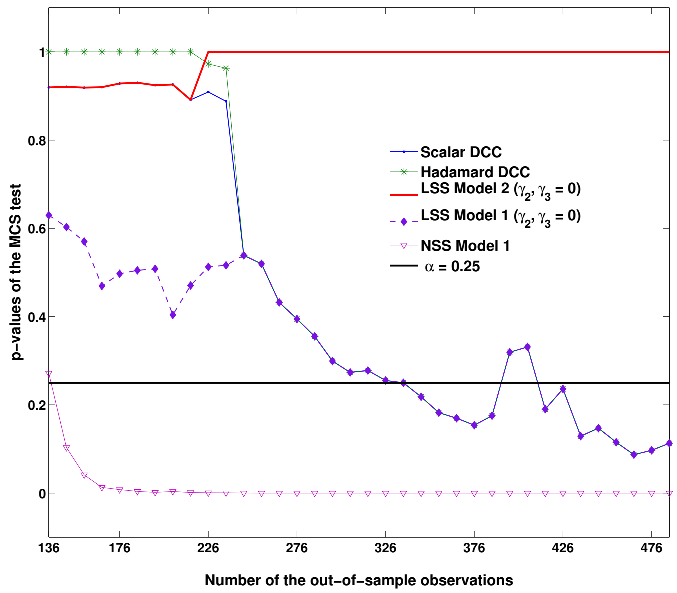

Figure 5: Evolution of the p-values of the MCS test based on the covariance loss function $d^{\text{cov}}_{\mathcal{F}^*_{(t-1)_3}}$ in (4.3) in terms of the out-of-sample length. The significance level of the MCS test is $\alpha = 0.25$.

KLIC loss functions. We also implement the MCS approach using loss functions based on the Kullback-Leibler Information Criterion (KLIC) [Kull 51]. We recall that the KLIC divergence $D_{T_{\text{est}},t}(\phi\|\psi)$ of a density ψ that depends on the parameters $\boldsymbol{\theta}$ from another density ϕ, is defined as:

$$D_{T_{\text{est}},t}(\phi\|\psi) = \frac{1}{t - T_{\text{est}}} \sum_{i=T_{\text{est}}+1}^{t} \ln\left[\frac{\phi_i(\mathbf{r}_i)}{\psi_i(\mathbf{r}_i;\boldsymbol{\theta})}\right], \quad t \in \{T_{\text{est}}+1, \ldots, T_{\text{est}}+T_{\text{out}}\} \tag{4.4}$$

where $\phi_i(\mathbf{r}_i)$ is the real underlying conditional density associated to the data under consideration and $\psi_i(\mathbf{r}_i;\boldsymbol{\theta})$ is the one coresponding to the competing model of interest.

We use this information criterion in order to construct a loss function to evaluate the out-of-sample density forecasting abilities of the considered models, that we subsequently use in the MCS context (see [Bao 06] and [Banu 15]). Since the terms having to do with the real density $\phi(\mathbf{r})$ are common to all the models and appear as an additive constant, we then disregard the numerator in (4.4) at the time of constructing the KLIC loss functions. Additionally, in order to account for the specific intraday extended information sets used at the time of forecasting, we use again three different KLIC loss functions adapted to these different filtrations, namely:

$$d^{\text{KLIC}}_{\mathcal{F}^*_{t_2}} := -\frac{1}{t - T_{\text{est}}} \sum_{i=T_{\text{est}}+1}^{t} \ln\left[\frac{1}{\sqrt{2\pi h_{33,i}}} \exp\left(-\frac{r^2_{3,i}}{2h_{33,i}}\right)\right], \tag{4.5}$$

$$d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_1}} := -\frac{1}{t-T_{\mathrm{est}}}\sum_{i=T_{\mathrm{est}}+1}^{t} \ln\left[\frac{1}{2\pi}\left[\det\begin{pmatrix} h_{22,i} & h_{23,i}\\ h_{32,i} & h_{33,i}\end{pmatrix}\right]^{-\frac{1}{2}}\times \right.$$
$$\left. \times\exp\left(-\frac{1}{2}\begin{pmatrix} r_{2,i}\\ r_{3,i}\end{pmatrix}^{\top}\begin{pmatrix} h_{22,i} & h_{23,i}\\ h_{32,i} & h_{33,i}\end{pmatrix}^{-1}\begin{pmatrix} r_{2,i}\\ r_{3,i}\end{pmatrix}\right)\right], \tag{4.6}$$

$$d^{\mathrm{KLIC}}_{\mathcal{F}^*_{(t-1)_3}} := -\frac{1}{t-T_{\mathrm{est}}}\sum_{i=T_{\mathrm{est}}+1}^{t} \ln\left[\frac{1}{(2\pi)^{3/2}\det(H_i)^{1/2}}\exp\left(-\frac{1}{2}\mathbf{r}_i^{\top}H_i^{-1}\mathbf{r}_i\right)\right], \tag{4.7}$$

where $t \in \{T_{\mathrm{est}}+1,\ldots,T_{\mathrm{est}}+T_{\mathrm{out}}\}$ and $h_{ij,t}$ are the (i,j)-entries of the model dependent forecasts for the conditional covariance matrices at t explained above in the context of the covariance loss functions (4.1)-(4.3).

Tables 4.3 and 4.4 contain the MCS results at significance levels 10% and 25% associated to the values of the covariance loss functions obtained in 36 different out-of-sample time intervals of the form $\{T_{\mathrm{est}}+1,\ldots,T_{\mathrm{est}}+136+10k\}$ with $k=\{0,1,\ldots,35\}$.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3=0$)	Model 1 ($\gamma_3,\gamma_2=0$)	Model 2	Model 2 ($\gamma_3=0$)	Model 2 ($\gamma_3,\gamma_2=0$)	Model 1
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_2}}$	0	0	0	0	0	**36**	0	0	0
Sum p-values	0	0	0	0	0	**36.0000**	0	0	0
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_1}}$	0	0	—	**36**	0	—	0	0	0
Sum p-values	0	0	—	**36.0000**	0	—	0	0	0
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$	0	0	—	—	0	—	—	**36**	0
Sum p-values	0	0	—	—	0	—	—	**36.0000**	0

Table 4.3: Model confidence sets (MCS) constructed using the KLIC loss functions (4.5), (4.6), and (4.7), for 36 different out-of-sample lengths $l(k)$, namely for $l(k)=T_{\mathrm{est}}+136+10k$, $k\in\{0,1,\ldots,35\}$. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at a 90% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 36 tests. The best performing models for the considered information set are marked in bold red.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3=0$)	Model 1 ($\gamma_3,\gamma_2=0$)	Model 2	Model 2 ($\gamma_3=0$)	Model 2 ($\gamma_3,\gamma_2=0$)	Model 1
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_2}}$	0	0	0	0	0	**36**	0	0	0
Sum p-values	0	0	0	0	0	**36.0000**	0	0	0
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_1}}$	0	0	—	**36**	0	—	11	2	0
Sum p-values	0	0	—	**34.4943**	0	—	9.9140	1.2591	0
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$	0	0	—	—	0	—	—	**36**	0
Sum p-values	0	0	—	—	0	—	—	**36.0000**	0

Table 4.4: Model confidence sets (MCS) constructed using the KLIC loss functions (4.5), (4.6), and (4.7), for 36 different out-of-sample lengths $l(k)$, namely for $l(k)=T_{\mathrm{est}}+136+10k$, $k\in\{0,1,\ldots,35\}$. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at a 75% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 36 tests. The best performing models for the considered information set are marked in bold red.

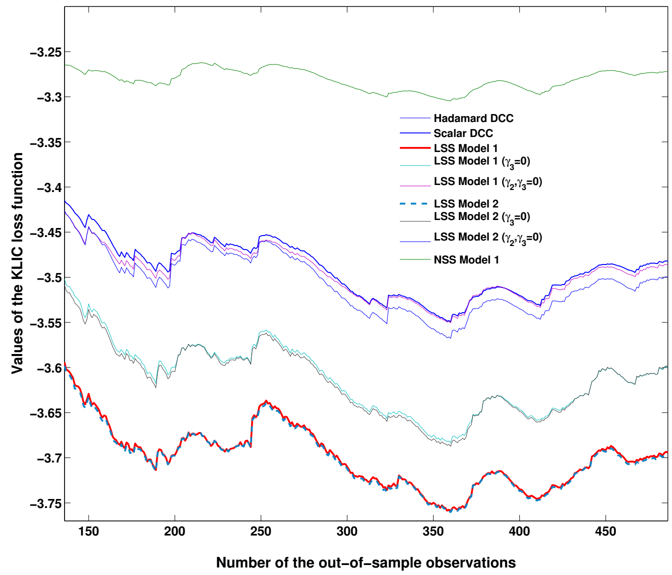

Figure 6: The values of the KLIC loss function $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_2}}$ in (4.5) for the competing models in terms of the out-of-sample length.

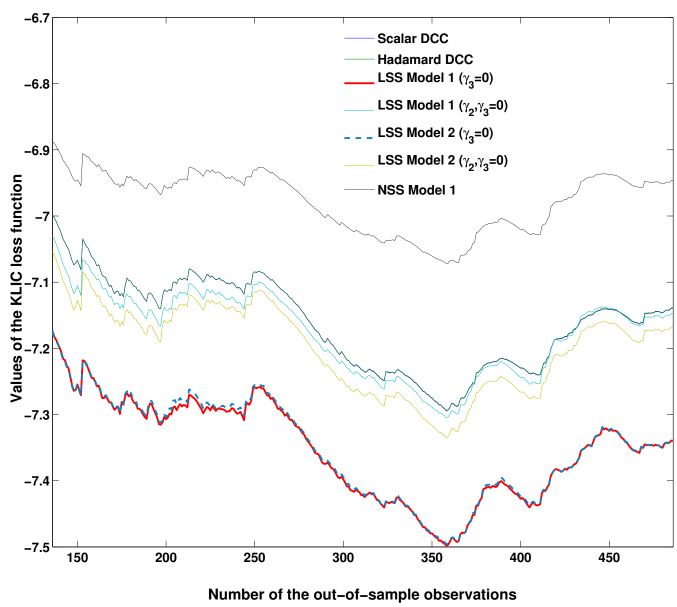

Figure 7: The values of the KLIC loss function $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_1}}$ in (4.6) for the competing models in terms of the out-of-sample length.

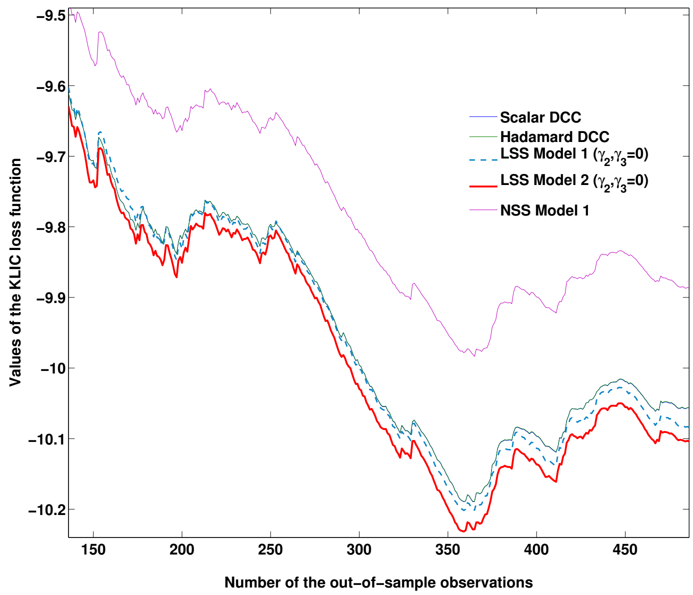

Figure 8: The values of the KLIC loss function $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$ in (4.7) for the competing models in terms of the out-of-sample length.

Figures 6-8 depict the evolution of the loss function values (4.5)-(4.7) with 10 observations stepwise increase in the out-of-sample length. The lowest values correspond to the best performing models. The results are robust with respect to the number of the out-of-sample observations and as for the previous MCS experiment we can conclude that:

(i) The group of Kalman-based models significantly outperforms the standard DCC models regardless the information sets involved.

(ii) The linear state-space models significantly outperform their nonlinear state-space counterparts regardless the information sets involved.

The appendix A.3 contains the results of an analogous experiment with a shorter estimation period that does not contain the high volatility events of the Fall 2008 period. In that situation the empirical study shows that: first, the conclusion **(i)** above holds and, second, **the nonlinear state-space model based approach significantly outperforms the linear one when the information set involved is the largest available**.

5 Conclusions

In this work we have used linear and nonlinear state-space models that extract global stochastic financial trends out of asynchronous daily data. These models are specifically constructed to take advantage of the intraday arrival of closing information coming from different international markets located in lagged time zones in order to enhance volatility and correlation forecasting performance.

The state-space models considered incorporate nonlinearities at various levels capable of capturing the heteroscedasticity that global trends empirically exhibit. This feature is of much importance since correlation forecasting is the main application developed. The identification of these models, as well as the constraints that their parameters need to satisfy in order to exhibit stationary solutions and positive semidefinite conditional correlation matrices, are carefully studied.

A volatility forecasting empirical study using the adjusted closing values of three major indices (NIKKEI, FTSE, and S&P500) has been conducted using the models introduced in the theoretical part and two different estimation periods. In this experiment, we use the model confidence set (MCS) approach of [Hans 03, Hans 11] implemented with loss functions constructed with the conditional covariance matrices implied by the different models under consideration. The results show that **the proposed Kalman-based forecasting scheme exhibits excellent and statistically significant performance improvements** when compared to the use of standard multivariate parametric correlation models (scalar and non-scalar DCC).

References

[Aiel 13] G. P. Aielli. "Dynamic Conditional Correlation: on Properties and Estimation". *Journal of Business & Economic Statistics*, Vol. 31, No. 3, pp. 282–299, July 2013.

[Ambr 08] D. Ambroževičit and A. Klivečka. "On the tvGARCH(1,1) model: Existence, CLT, and tail index". *Lithuanian Mathematical Journal*, Vol. 48, No. 1, pp. 1–16, Jan. 2008.

[Bae 11] B. Y. Bae and D. H. Kim. "Global and Regional Yield Curve Dynamics and Interactions: The Case of Some Asian Countries". *International Economic Journal*, Vol. 25, No. 4, pp. 717–738, Dec. 2011.

[Banu 15] G.-D. Banulescu, P. R. Hansen, and Z. Huang. "Volatility during the financial crisis through the lens of high frequency data: a realized GARCH approach". 2015.

[Bao 06] Y. Bao, T.-H. Lee, and B. Saltoglu. "Comparing density forecast models". 2006.

[Bauw 15] L. Bauwens, L. Grigoryeva, and J.-P. Ortega. "Estimation and empirical performance of non-scalar dynamic conditional correlation models". *To appear in Computational Statistics and Data Analysis*, 2015.

[Bent 15] S. R. Bentes. "On the integration of financial markets: How strong is the evidence from five international stock markets?". *Physica A: Statistical Mechanics and its Applications*, Vol. 429, pp. 205–214, July 2015.

[Boll 88] T. Bollerslev, R. F. Engle, and J. M. Wooldridge. "A capital asset pricing model with time varying covariances". *Journal of Political Economy*, Vol. 96, pp. 116–131, 1988.

[Cart 11] A. Cartea and D. Karyampas. "Volatility and covariation of financial assets: A high-frequency analysis". *Journal of Banking & Finance*, Vol. 35, No. 12, pp. 3319–3334, Dec. 2011.

[Chan 09] Y. Chang, J. Isaac Miller, and J. Y. Park. "Extracting a common stochastic trend: Theory with some applications". *Journal of Econometrics*, Vol. 150, No. 2, pp. 231–247, June 2009.

[Chou 07] T. Choudhry, L. Lu, and K. Peng. "Common stochastic trends among Far East stock prices: Effects of the Asian financial crisis". *International Review of Financial Analysis*, Vol. 16, No. 3, pp. 242–261, Jan. 2007.

[Chre 14] S. Chrétien and J.-P. Ortega. "Multivariate GARCH estimation via a Bregman-proximal trust-region method". *Computational Statistics and Data Analysis*, Vol. 76, pp. 210–236, 2014.

[Chun 94] P. J. Chung and D. J. Liu. "Common stochastic trends in pacific rim stock markets". *The Quarterly Review of Economics and Finance*, Vol. 34, No. 3, pp. 241–259, Sep. 1994.

[Cize 09] P. Cizek and V. Spokoiny. "Varying coefficient GARCH models". In: T. Mikosch, J.-P. Kreiß, R. A. Davis, and T. G. Andersen, Eds., *Handbook of Financial Time Series*, pp. 169–185, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[Dahl 06] R. Dahlhaus and S. S. Rao. "Statistical inference for time-varying ARCH processes". *The Annals of Statistics*, Vol. 34, No. 3, pp. 1075–1114, June 2006.

[Dung 01] M. Dungey, V. L. Martin, and A. R. Pagan. "A multivariate latent factor decomposition of international bond yield spreads". *Journal of Applied Econometrics*, Vol. 15, No. 6, pp. 697–715, 2001.

[Durb 12] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2012.

[Durd 14] R. I. Durdyev and A. Peresetsky. "Autocorrelation in the global stochastic trend". *Applied Econometrics*, Vol. 35, No. 3, pp. 81–100, 2014.

[Engl 02] R. F. Engle. "Dynamic conditional correlation -a simple class of multivariate GARCH models". *Journal of Business and Economic Statistics*, Vol. 20, pp. 339–350, 2002.

[Engl 09] R. Engle. *Anticipating Correlations*. Princeton University Press, Princeton, NJ, 2009.

[Feli 12] G. Felices and T. Wieladek. "Are emerging market indicators of vulnerability to financial crises decoupling from global factors?". *Journal of Banking & Finance*, Vol. 36, No. 2, pp. 321–331, Feb. 2012.

[Fran 10] C. Francq and J.-M. Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, 2010.

[Gour 97] C. Gouriéroux. *ARCH models and financial applications*. *Springer Series in Statistics*, Springer-Verlag, New York, 1997.

[Hans 03] P. R. Hansen, A. Lunde, and J. M. Nason. "Choosing the best volatility models: the model confidence set approach". *Oxford Bulletin of Economics and Statistics*, Vol. 65, No. s1, pp. 839–861, Dec. 2003.

[Hans 11] P. R. Hansen, A. Lunde, and J. M. Nason. "The model confidence set". *Econometrica*, Vol. 79, No. 2, pp. 453–497, 2011.

[Horn 13] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, second Ed., 2013.

[Jeon 91] B. N. Jeon and T. C. Chiang. "A system of stock prices in world stock exchanges: Common stochastic trends for 19751990". *Journal of Economics and Business*, Vol. 43, No. 4, pp. 329–338, Nov. 1991.

[Kasa 92] K. Kasa. "Common stochastic trends in international stock markets". *Journal of Monetary Economics*, Vol. 29, No. 1, pp. 95–124, Feb. 1992.

[Korh 13] I. Korhonen and A. Peresetsky. "Extracting global stochastic trend from non-synchronous data". *Bank of Finland, BOFIT Discussion Papers*, Vol. 15/2013, 2013.

[Kull 51] L. Kullback and R. A. Leibler. "On information and sufficiency". *Annals of Mathematical Statistics*, Vol. 22, pp. 79–86, 1951.

[Luce 11] B. M. Lucey and C. Muckley. "Robust global stock market interdependencies". *International Review of Financial Analysis*, Vol. 20, No. 4, pp. 215–224, Aug. 2011.

[Pere 15] A. Peresetsky and R. Yakubov. "Autocorrelation in an unobservable global trend: does it help to forecast market returns?". *To appear in International Journal of Computational Economics and Econometrics*, 2015.

[Phen 04] C. Phengpis and V. P. Apilado. "Economic interdependence and common stochastic trends: A comparative analysis between EMU and non-EMU stock markets". *International Review of Financial Analysis*, Vol. 13, No. 3, pp. 245–263, Sep. 2004.

[Rang 01] J. Rangvid. "Increasing convergence among European stock markets?". *Economics Letters*, Vol. 71, No. 3, pp. 383–389, June 2001.

[Rang 02] J. Rangvid and C. Sorensen. "Convergence in the ERM and Declining Numbers of Common Stochastic Trends". *Journal of Emerging Market Finance*, Vol. 1, No. 2, pp. 183–213, Sep. 2002.

[Roha 13] N. Rohan and T. V. Ramanathan. "Nonparametric estimation of a time-varying GARCH model". *Journal of Nonparametric Statistics*, Vol. 25, No. 1, pp. 33–52, March 2013.

[Sikl 01] P. L. Siklos and P. Ng. "Integration Among Asia-Pacific and International Stock Markets: Common Stochastic Trends and Regime Shifts". *Pacific Economic Review*, Vol. 6, No. 1, pp. 89–110, Feb. 2001.

[Tse 02] Y. K. Tse and A. K. C. Tsui. "A multivariate GARCH with time-varying correlations". *Journal of Business and Economic Statistics*, Vol. 20, pp. 351–362, 2002.

A Appendices and supplementary material

A.1 Proof of Proposition 2.1

Let A be an element of the general linear group of order five, that is, $A \in GL_5(\mathbb{R})$. Consider the model prescription (2.5a)-(2.5b) and transform it according to the following prescription. First, replace the parameter B by $BA^{-1}A$ in (2.5a) and second, apply A to both sides of (2.5b). This yields:

$$\begin{cases} \mathbf{r}_t = \boldsymbol{\alpha} + BA^{-1}A\boldsymbol{e}_t + \mathbf{u}_t, & \text{(A.1a)} \\ A\boldsymbol{e}_t = AT\boldsymbol{e}_{t-1} + AR\mathbf{v}_{t-1}. & \text{(A.1b)} \end{cases}$$

The model (2.5a)-(2.5b) remains invariant under this transformation if the following conditions hold:

(i) BA^{-1} has the same entries structure as B.

(ii) The matrices A and T commute, that is, $AT = TA$.

(iii) $\bar{Q} := AQA^\top$ with $Q := RR^\top$ is a matrix of the same entries structure as Q, namely, $\bar{Q} = \text{diag}(\bar{\sigma}^2_{v,1}, \bar{\sigma}^2_{v,2}, \bar{\sigma}^2_{v,3}, 0, 0)$, for some $\bar{\sigma}^2_{v,1}, \bar{\sigma}^2_{v,2}, \bar{\sigma}^2_{v,3} \in \mathbb{R}^+$.

Indeed, under these hypotheses, the transformed equations (A.1a)-(A.1b) become:

$$\begin{cases} \mathbf{r}_t = \boldsymbol{\alpha} + (BA^{-1})(A\boldsymbol{e}_t) + \mathbf{u}_t, & \text{(A.2a)} \\ (A\boldsymbol{e}_t) = T(A\boldsymbol{e}_{t-1}) + AR\mathbf{v}_{t-1}. & \text{(A.2b)} \end{cases}$$

It is hence easy to see that the model (A.2a)-(A.2b) has the same structure as the original model (2.5a)-(2.5b) with the variables $\boldsymbol{e}_t$ replaced by $(A\boldsymbol{e}_t)$, provided that BA^{-1} has the same entries structure as B and that the covariance matrix $\Sigma(\bar{\mathbf{v}}_t)$ of $\bar{\mathbf{v}}_t := AR\mathbf{v}_t$ is of the form $\bar{Q} = \text{diag}(\bar{\sigma}_{v,1}^2, \bar{\sigma}_{v,2}^2, \bar{\sigma}_{v,3}^2, 0, 0)$, with some $\bar{\sigma}_{v,1}, \bar{\sigma}_{v,2}, \bar{\sigma}_{v,3} \in \mathbb{R}^+$. This covariance matrix equals

$$\Sigma(\bar{\mathbf{v}}_t) := \mathrm{E}\left[\bar{\mathbf{v}}_t \bar{\mathbf{v}}_t^\top\right] = \mathrm{E}\left[AR\mathbf{v}_t\mathbf{v}_t^\top R^\top A^\top\right] = ARR^\top A^\top = AQA^\top \tag{A.3}$$

with $Q := RR^\top$.

We first study what the implications that conditions **(i)-(iii)** have in the structure of $A \in GL_5(\mathbb{R})$. First, by point **(iii)** suppose that $A \in GL_5(\mathbb{R})$ is such that for any $\sigma_{v,1}, \sigma_{v,2}, \sigma_{v,3} \in \mathbb{R}^+$ there exist $\bar{\sigma}_{v,1}, \bar{\sigma}_{v,2}, \bar{\sigma}_{v,3} \in \mathbb{R}^+$ such that

$$A \cdot \begin{pmatrix} \sigma_{v,1}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{v,2}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{v,3}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot A^\top = \begin{pmatrix} \bar{\sigma}_{v,1}^2 & 0 & 0 & 0 & 0 \\ 0 & \bar{\sigma}_{v,2}^2 & 0 & 0 & 0 \\ 0 & 0 & \bar{\sigma}_{v,3}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{A.4}$$

Define now $\Sigma_v := \begin{pmatrix} \sigma_{v,1}^2 & 0 & 0 \\ 0 & \sigma_{v,2}^2 & 0 \\ 0 & 0 & \sigma_{v,3}^2 \end{pmatrix}$, $\bar{\Sigma}_v := \begin{pmatrix} \bar{\sigma}_{v,1}^2 & 0 & 0 \\ 0 & \bar{\sigma}_{v,2}^2 & 0 \\ 0 & 0 & \bar{\sigma}_{v,3}^2 \end{pmatrix}$, and let $K, P \in \mathbb{M}_3$, $C, W \in \mathbb{M}_{3,2}$, $D, X \in \mathbb{M}_{2,3}$, $E, U \in \mathbb{M}_2$ be such that $A = \left(\begin{array}{c|c} K & C \\ \hline D & E \end{array}\right)$, $A^\top = \left(\begin{array}{c|c} K^\top & D^\top \\ \hline C^\top & E^\top \end{array}\right)$, and $A^{-1} = \left(\begin{array}{c|c} P & W \\ \hline X & U \end{array}\right)$. Condition (A.4), namely, $AQA^\top = \bar{Q}$ is equivalent to $A^{-1}AQA^\top = A^{-1}\bar{Q}$ or to $QA^\top = A^{-1}\bar{Q}$ which in the notation that we just introduced amounts to

$$\left(\begin{array}{c|c} \Sigma_v & 0 \\ \hline 0 & 0 \end{array}\right) \left(\begin{array}{c|c} K^\top & D^\top \\ \hline C^\top & E^\top \end{array}\right) = \left(\begin{array}{c|c} P & W \\ \hline X & U \end{array}\right) \left(\begin{array}{c|c} \bar{\Sigma}_v & 0 \\ \hline 0 & 0 \end{array}\right). \tag{A.5}$$

Expression (A.5) is equivalent to the following three conditions:

$$\Sigma_v K^\top = P\bar{\Sigma}_v, \tag{A.6}$$

$$\Sigma_v D^\top = 0, \tag{A.7}$$

$$X\bar{\Sigma}_v = 0. \tag{A.8}$$

We continue by noticing that since Σ_v and $\bar{\Sigma}_v$ are invertible, the expressions (A.7) and (A.8) amount to $D^\top = 0$ and $X = 0$, respectively. This shows that $A = \left(\begin{array}{c|c} K & C \\ \hline 0 & E \end{array}\right)$ and $A^{-1} = \left(\begin{array}{c|c} P & W \\ \hline 0 & U \end{array}\right)$. We now impose the condition **(ii)**, that is, $AT = TA$:

$$\left(\begin{array}{c|c} K & C \\ \hline 0 & E \end{array}\right) \left(\begin{array}{c|c} 0 & 0 \\ \hline M & 0 \end{array}\right) = \left(\begin{array}{c|c} 0 & 0 \\ \hline M & 0 \end{array}\right) \left(\begin{array}{c|c} K & C \\ \hline 0 & E \end{array}\right) \tag{A.9}$$

with $M := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. This relation implies that

$$CM = 0, \tag{A.10}$$

$$EM = MK, \tag{A.11}$$

$$0 = MC. \tag{A.12}$$

The expressions (A.10) and (A.12) imply that $C = 0$, which yields that

$$A = \left(\begin{array}{c|c} K & 0 \\ \hline 0 & E \end{array}\right). \tag{A.13}$$

As A is by definition invertible, in view of (A.13) so are the submatrices K and E, and hence in the block structure of A^{-1} we can set $W = 0$, $P = K^{-1}$, and $U = E^{-1}$, respectively, that is,

$$A^{-1} = \left(\begin{array}{c|c} K^{-1} & 0 \\ \hline 0 & E^{-1} \end{array}\right). \tag{A.14}$$

At the same time it is easy to verify that the relation (A.11) implies that

$$K = \left(\begin{array}{c|cc} k_{11} & k_{12} & k_{13} \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{E}} \\ 0 & & \end{array}\right). \tag{A.15}$$

Let now denote by k^*_{ij} and by e^*_{ij} with $i, j \in \{1, 2, 3\}$ the generic entries of the matrices K^{-1} and E^{-1}, respectively. We may hence write by (A.15) that

$$K^{-1} = \left(\begin{array}{c|cc} \frac{1}{k_{11}} & k^*_{12} & k^*_{13} \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{E^{-1}}} \\ 0 & & \end{array}\right) = \begin{pmatrix} \frac{1}{k_{11}} & k^*_{12} & k^*_{13} \\ 0 & e^*_{11} & e^*_{12} \\ 0 & e^*_{21} & e^*_{22} \end{pmatrix}. \tag{A.16}$$

We now use the fact that condition **(i)** requires that the matrix BA^{-1} has the same structure as B, that is, there exist some $\bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3 \in \mathbb{R}$ such that

$$BA^{-1} = \begin{pmatrix} \bar{\beta}_1 & 0 & 0 & \bar{\beta}_1 & \bar{\beta}_1 \\ \bar{\beta}_2 & \bar{\beta}_2 & 0 & 0 & \bar{\beta}_2 \\ \bar{\beta}_3 & \bar{\beta}_3 & \bar{\beta}_3 & 0 & 0 \end{pmatrix}. \tag{A.17}$$

We first partition the matrix B and write it as $B := (B_1 | B_2)$, with

$$B_1 = \begin{pmatrix} \beta_1 & 0 & 0 \\ \beta_2 & \beta_2 & 0 \\ \beta_3 & \beta_3 & \beta_3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} \beta_1 & \beta_1 \\ 0 & \beta_2 \\ 0 & 0 \end{pmatrix}. \tag{A.18}$$

We now use (A.14), (A.18) and write

$$BA^{-1} = (B_1 | B_2) \cdot \left(\begin{array}{c|c} K^{-1} & 0 \\ \hline 0 & E^{-1} \end{array}\right) = (B_1 K^{-1} | B_2 E^{-1})$$

which by (A.17) requires both

$$\begin{pmatrix} \beta_1 & 0 & 0 \\ \beta_2 & \beta_2 & 0 \\ \beta_3 & \beta_3 & \beta_3 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{k_{11}} & k^*_{12} & k^*_{13} \\ 0 & e^*_{11} & e^*_{12} \\ 0 & e^*_{21} & e^*_{22} \end{pmatrix} = \begin{pmatrix} \bar{\beta}_1 & 0 & 0 \\ \bar{\beta}_2 & \bar{\beta}_2 & 0 \\ \bar{\beta}_3 & \bar{\beta}_3 & \bar{\beta}_3 \end{pmatrix} \tag{A.19}$$

and

$$\begin{pmatrix} \beta_1 & \beta_1 \\ 0 & \beta_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} e^*_{11} & e^*_{12} \\ e^*_{21} & e^*_{22} \end{pmatrix} = \begin{pmatrix} \bar{\beta}_1 & \bar{\beta}_1 \\ 0 & \bar{\beta}_2 \\ 0 & 0 \end{pmatrix}. \tag{A.20}$$

The relations implied by the matrix equation (A.19) for the entries (1,2), (1,3), and (2,3) yield that $k^*_{12} = 0$, $k^*_{13} = 0$, and $e^*_{12} = 0$, respectively. At the same time, the relation for the component (2,1) of the matrix equation (A.20) yields that $e^*_{21} = 0$. Consequently, both K^{-1} and E^{-1} are diagonal matrices. Finally, the relation (A.19) computed for the corresponding diagonal elements of K^{-1} implies that that $e^*_{11} = \frac{1}{k_{11}}$, $e^*_{22} = \frac{1}{k_{11}}$ and we can hence write that

$$K = \lambda \mathbb{I}_3, \ \ \lambda \in \mathbb{R}.$$

Consequently, by (A.15)

$$E = \lambda \mathbb{I}_2,$$

which automatically guarantees by (A.13) that $A = \lambda \mathbb{I}_5$, necessarily. This implies that the matrix B in the model (2.5a)-(2.5b) is defined up to multiplication by a homothety and hence the model is well identified provided that one of the elements β_i or $\sigma^2_{v,i}$, $i \in \{1, 2, 3\}$ is set to a constant (positive in the case of $\sigma^2_{v,i}$). ■

A.2 Proof of Proposition 3.1

Proof of part (i). The recursion that defines the tvGARCH(1,1) model in (3.15) implies that (see for example formula (2.2) in [Ambr 08]):

$$\sigma^2_{t_i} = \sum_{j=0}^{\infty} a_{t_i - j} \prod_{k=1}^{j} \left(\gamma_{t_i-k+1} v^2_{t_i-k} + \delta_{t_i-k+1}\right) = \sum_{j=0}^{\infty} a_{t_i-j} \left[b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1}\right], \tag{A.21}$$

where $b_{t_i-k+1} := \left(\gamma_{t_i-k+1} v^2_{t_i-k} + \delta_{t_i-k+1}\right)$. We notice that the process $\{b_{t_i}\}$ is made of positive independent random variables. Moreover, by the Cauchy rule for series with non-negative terms, expression (A.21) converges if

$$\lambda := \lim_{j\to\infty} \left[b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1}\right]^{1/j} < 1.$$

We therefore compute:

$$\begin{aligned} \lim_{j\to\infty} \left[b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1}\right]^{1/j} &= \lim_{j\to\infty} \exp\left[\frac{1}{j}\sum_{k=1}^{j} \log\left(b_{t_i-k+1}\right)\right] = \exp \lim_{j\to\infty} \left[\frac{1}{j}\sum_{k=1}^{j} \log\left(b_{t_i-k+1}\right)\right] \\ &= \exp \frac{1}{3}\sum_{l=1}^{3} \mathrm{E}\left[\log\left(\gamma_l v_t^2 + \delta_l\right)\right] \leq \exp \frac{1}{3}\sum_{l=1}^{3} \log\left(\mathrm{E}\left[\left(\gamma_l v_t^2 + \delta_l\right)\right]\right) \quad \text{(A.22)} \\ &= \left[(\delta_1+\gamma_1)(\delta_2+\gamma_2)(\delta_3+\gamma_3)\right]^{1/3}, \quad \text{(A.23)} \end{aligned}$$

where the first equality in (A.22) follows from the strong law of large numbers and the relation that follows it is a consequence of Jensen's inequality. The inequality in the statement implies hence by (A.23) that $\lambda := \lim_{j\to\infty} \left[b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1}\right]^{1/j} < 1$. A strategy mimicking, for example, the proof of Theorem 2.1 in [Fran 10], shows that in that situation model 1 has a unique stationary solution.

Proof of part (ii). By Theorem 2.4 in [Fran 10], it suffices to show that the top Lyapunov exponent γ of the sequence $\{A_{t_i}\}$ is smaller than zero. By Theorem 2.3 in [Fran 10]:

$$\gamma = \lim_{t_i \to \infty} \frac{1}{t_i} \mathrm{E}\left[\log \|A_{t_i} A_{t_i-1} \cdots A_1\|\right],$$

with $\|\cdot\|$ any matrix norm. We now use the norm $\|A\| = \sum_{i,j} |a_{ij}|$ and notice that if all the elements of A are positive then

$$\mathrm{E}\left[\|A\|\right] = \|\mathrm{E}\left[A\right]\|. \tag{A.24}$$

Consequently,

$$\begin{aligned}
\gamma &= \lim_{t_i \to \infty} \frac{1}{t_i} \mathrm{E}\left[\log \|A_{t_i} A_{t_i-1} \cdots A_1\|\right] \leq \lim_{t_i \to \infty} \frac{1}{t_i} \log\left(\mathrm{E}\left[\|A_{t_i} A_{t_i-1} \cdots A_1\|\right]\right) && \text{(A.25)} \\
&= \lim_{t_i \to \infty} \frac{1}{t_i} \log\left(\|\mathrm{E}\left[A_{t_i} A_{t_i-1} \cdots A_1\right]\|\right) = \lim_{t_i \to \infty} \frac{1}{t_i} \log\left(\|\mathrm{E}\left[A_{t_i}\right] \mathrm{E}\left[A_{t_i-1}\right] \cdots \mathrm{E}\left[A_1\right]\|\right) && \text{(A.26)} \\
&= \frac{1}{3} \lim_{t \to \infty} \frac{1}{t} \log\left(\|A_3 A_2 A_1\|^t\right) = \frac{1}{3} \log\left(\rho(A_3 A_2 A_1)\right) = \log\left(\rho(A_3 A_2 A_1)^{1/3}\right). && \text{(A.27)}
\end{aligned}$$

The relation in (A.25) follows from Jensen's inequality, the first equality in (A.26) is a consequence of (A.24) and the second one of the independence of the elements in the process $\{A_{t_i}\}$. Finally, in (A.27) we use Gelfand's formula for the characterization of the spectral radius of a matrix. The inequality $\gamma \leq \log\left(\rho(A_3 A_2 A_1)^{1/3}\right)$ that we just proved guarantees that the condition in the statement ensures that $\log\left(\rho(A_3 A_2 A_1)^{1/3}\right) < 0$ and hence $\gamma < 0$, as required.

In both cases, the arguments that we provided show that the unconditional variance $\mathrm{E}\left[\sigma_{t_i}^2\right]$ depends only on i and hence establishes the periodic stationarity claimed in the statement. ■

A.3 Empirical performance of the GST-based volatility forecasting schemes using a smaller estimation sample

In this appendix we illustrate the robustness of the results obtained in the empirical study in Section 4 by performing a similar analysis based on the same dataset but using a shorter estimation period (January 5, 1996 – December 4, 2006) that does not contain the volatility events in the Fall 2008. The out-of-sample study in that case comprises the entire Great Recession (December 5, 2006 – April 1, 2015). This choice yields a dataset with a length of $T := 4581$ observations and for which $T_{\mathrm{est}} := 2600$ and $T_{\mathrm{out}} := 1981$.

The study follows the same scheme as the one in Section 4. In particular, we consider the same competing models and the same loss functions at the time of implementing the MCS strategy. In the case of the covariance loss functions (4.1)-(4.3), the results of the corresponding MCS comparison are contained in the Tables A.5 and A.6. These results correspond to the values of the covariance loss functions obtained in 185 different out-of-sample time intervals of the form $\{T_{\mathrm{est}}+1, \ldots, T_{\mathrm{est}}+141+10k\}$ with , $k = \{0, 1, \ldots, 184\}$. The first 141 elements in the out-of-sample period are included in all these intervals in order to ensure that there are enough values available for the bootstrapping process that is necessary in the estimation of the distribution of the model equivalence test statistic. The date corresponding to the end of this offset interval is July 10, 2007.

Figures 9-11 depict the evolution of the MCS p-values when the number of the out-of-sample observations considered grows with a step equal to 10 observations and the significance level is set to $\alpha = 0.25$. Notice that only the figures corresponding to the DCC benchmarks and to the models that exhibit the best performances for each of the intraday extended information sets are represented in the relevant figures.

It is worth to be noted that the most visible changes in the behavior of the models illustrated in Figures 9-11 takes place in the interval given by the observations 431 and 481, which corresponds to the period of time between September 25, 2008 and December 10, 2008 in which well-known market events created high levels of volatility and noticeable dynamical changes.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3=0$)	Model 1 ($\gamma_3,\gamma_2=0$)	Model 2	Model 2 ($\gamma_3=0$)	Model 2 ($\gamma_3,\gamma_2=0$)	Model 1
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_2}}$	21	21	**185**	39	15	40	39	12	15
Sum p-values	13.6606	13.6606	**177.9308**	22.1921	11.5341	29.0059	20.4807	11.2417	12.7414
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_1}}$	16	16	—	**185**	18	—	184	19	27
Sum p-values	14.3331	14.3888	—	**184.2558**	14.0385	—	45.4232	15.9279	16.1003
MCS $d^{\text{cov}}_{\mathcal{F}^*_{(t-1)_3}}$	165	166	—	—	22	—	—	176	**185**
Sum p-values	120.2549	120.3713	—	—	12.5710	—	—	157.3145	**174.6149**

Table A.5: Model confidence sets (MCS) constructed using the covariance based loss functions (4.1), (4.2), and (4.3), respectively, for 185 different out-of-sample lengths $l(k)$, namely for $l(k)=T_{\text{est}}+141+10k$, $k\in\{0,1,\ldots,184\}$. For each model and information set under consideration, the corresponding value indicates the number of times that that model has been included in the MCS at a 90% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 185 tests. The best performing models for the considered information set are marked in bold red.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3=0$)	Model 1 ($\gamma_3,\gamma_2=0$)	Model 2	Model 2 ($\gamma_3=0$)	Model 2 ($\gamma_3,\gamma_2=0$)	Model 1
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_2}}$	7	7	**183**	29	1	33	28	2	6
Sum p-values	13.5594	13.5594	**177.9236**	22.0953	11.4554	28.9149	20.3553	11.1596	12.7184
MCS $d^{\text{cov}}_{\mathcal{F}^*_{t_1}}$	5	5	—	**185**	4	—	62	10	11
Sum p-values	14.3411	14.4063	—	**184.2549**	14.0456	—	45.4273	15.9334	16.0872
MCS $d^{\text{cov}}_{\mathcal{F}^*_{(t-1)_3}}$	155	155	—	—	1	—	—	156	**185**
Sum p-values	120.2530	120.3687	—	—	12.5903	—	—	157.3078	**174.6051**

Table A.6: Model confidence sets (MCS) constructed using the covariance based loss functions (4.1), (4.2), and (4.3), respectively, for 185 different out-of-sample lengths $l(k)$, namely for $l(k)=T_{\text{est}}+141+10k$, $k\in\{0,1,\ldots,184\}$. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at a 75% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 185 tests. The best performing models for the considered information set are marked in bold red.

The results corresponding to the KLIC loss functions (4.5)-(4.7) are contained in Tables A.7 and A.8 at a significance levels 10% and 25%, respectively. These values are obtained out of 185 different out-of-sample time intervals of the form $\{T_{\text{est}}+1,\ldots,T_{\text{est}}+141+10k\}$ with $k=\{0,1,\ldots,184\}$. As the previous MCS experiment constructed using covariance based loss functions already showed, **the forecasting approaches based on the linear state-space Model 1/Model 2 and nonlinear state-space Model 1 setups significantly outperform the standard DCC models in this context**. Figures 12-8 depict the evolution of the loss function values (4.5)-(4.7) using a 10 observations stepwise increase in the out-of-sample length. The lowest values correspond to the best performing models. The results that we just obtained show the robustness of the study conducted in Section 4.3 with respect to the choice of estimation period and the number of the out-of-sample observations.

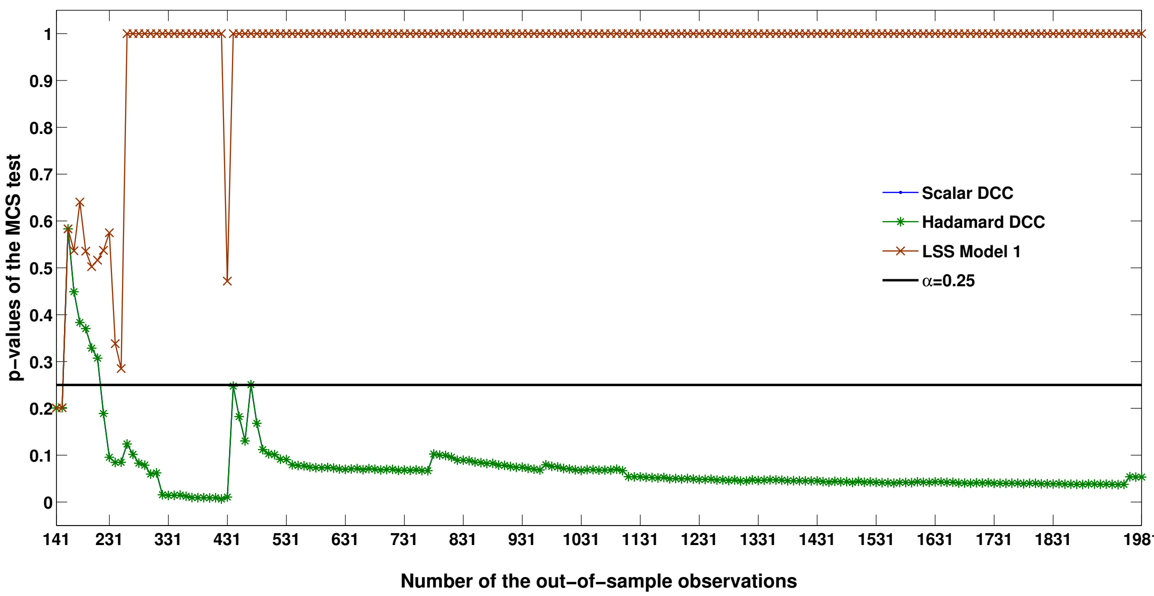

Figure 9: Evolution of the p-values of the MCS test based on the covariance loss function $d^{\text{cov}}_{\mathcal{F}^*_{t_2}}$ in (4.1) in terms of the out-of-sample length. The significance level of the MCS test is $\alpha = 0.25$.

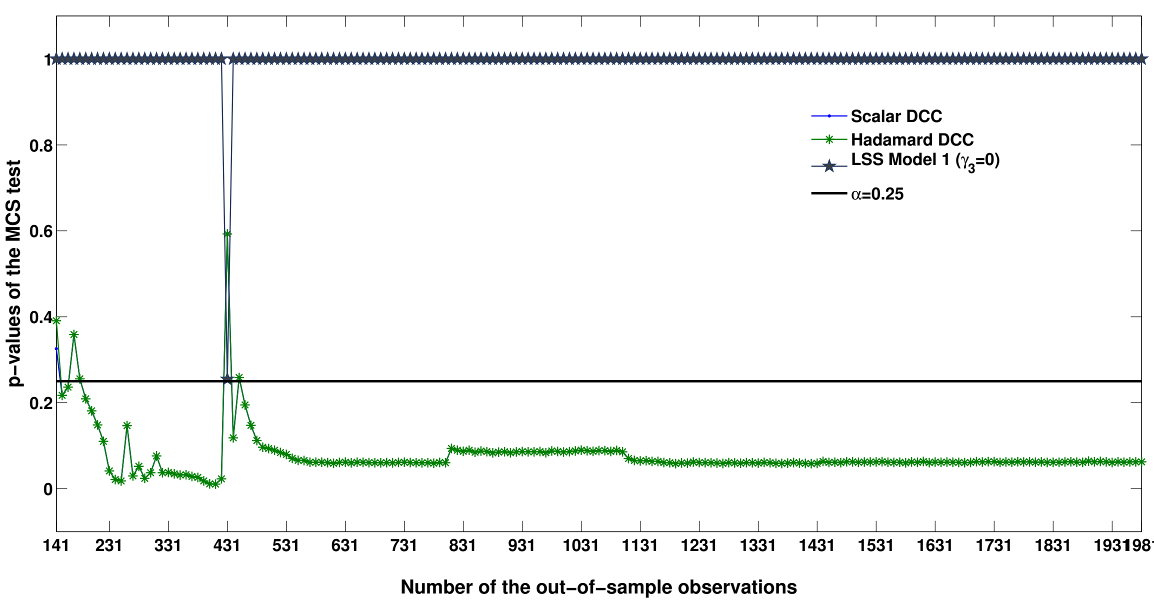

Figure 10: Evolution of the p-values of the MCS test based on the covariance loss function $d^{\text{cov}}_{\mathcal{F}^*_{t_1}}$ in (4.2) in terms of the out-of-sample length. The significance level of the MCS test is $\alpha = 0.25$.

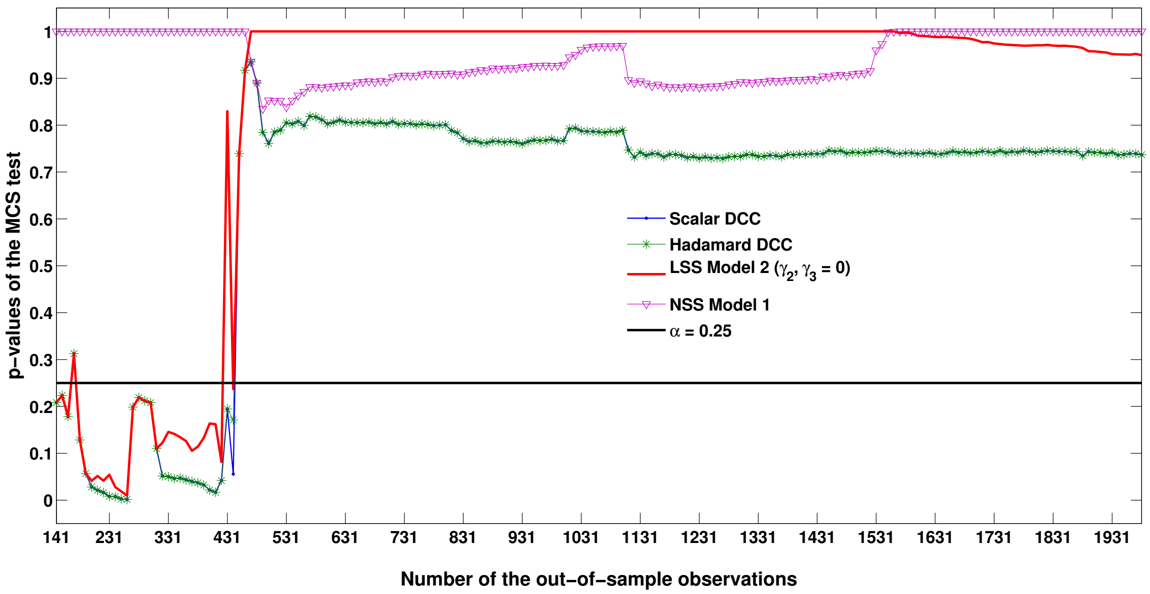

Figure 11: Evolution of the p-values of the MCS test based on the covariance loss function $d^{\text{cov}}_{\mathcal{F}^*_{(t-1)_3}}$ in (4.3) in terms of the out-of-sample length. The significance level of the MCS test is $\alpha = 0.25$.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3 = 0$)	Model 1 ($\gamma_3, \gamma_2 = 0$)	Model 2	Model 2 ($\gamma_3 = 0$)	Model 2 ($\gamma_3, \gamma_2 = 0$)	Model 1
MCS $d^{\text{KLIC}}_{\mathcal{F}^*_{t_2}}$	22	23	0	0	0	**167**	0	0	0
Sum p-values	9.8074	22.7378	0	0	0	**165.9413**	0	0	0
MCS $d^{\text{KLIC}}_{\mathcal{F}^*_{t_1}}$	0	3	—	0	0	—	**185**	0	0
Sum p-values	0	1.6090	—	0	0	—	**185.0000**	0	0
MCS $d^{\text{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$	0	16	—	—	0	—	—	0	**175**
Sum p-values	0	13.3927	—	—	0	—	—	0	**175.0000**

Table A.7: Model confidence sets (MCS) constructed using the KLIC loss functions (4.5), (4.6), and (4.7), for 185 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\text{est}} + 141 + 10k$, $k \in \{0, 1, \ldots, 184\}$. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at the 90% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 185 tests. The best performing models for the considered information set are marked in bold red.

	DCC models		Linear state-space models						Nonlinear state-space model
	Scalar	Hadamard	Model 1	Model 1 ($\gamma_3 = 0$)	Model 1 ($\gamma_3, \gamma_2 = 0$)	Model 2	Model 2 ($\gamma_3 = 0$)	Model 2 ($\gamma_3, \gamma_2 = 0$)	Model 1
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_2}}$	18	23	0	0	0	**167**	0	0	0
Sum p-values	9.8125	22.7399	0	0	0	**165.9375**	0	0	0
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_1}}$	0	3	—	0	0	—	**185**	0	0
Sum p-values	0	1.6073	—	0	0	—	**185.0000**	0	0
MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$	0	15	—	—	0	—	—	0	**175**
Sum p-values	0	13.3896	—	—	0	—	—	0	**175.0000**

Table A.8: Model confidence sets (MCS) constructed using the KLIC loss functions (4.5), (4.6), and (4.7), for 185 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\mathrm{est}} + 141 + 10k$, $k \in \{0, 1, \ldots, 184\}$. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at the 75% confidence level; the value underneath indicates the sum of all the MCS p-values obtained by a given model in the 185 tests. The best performing models for the considered information set are marked in bold red.

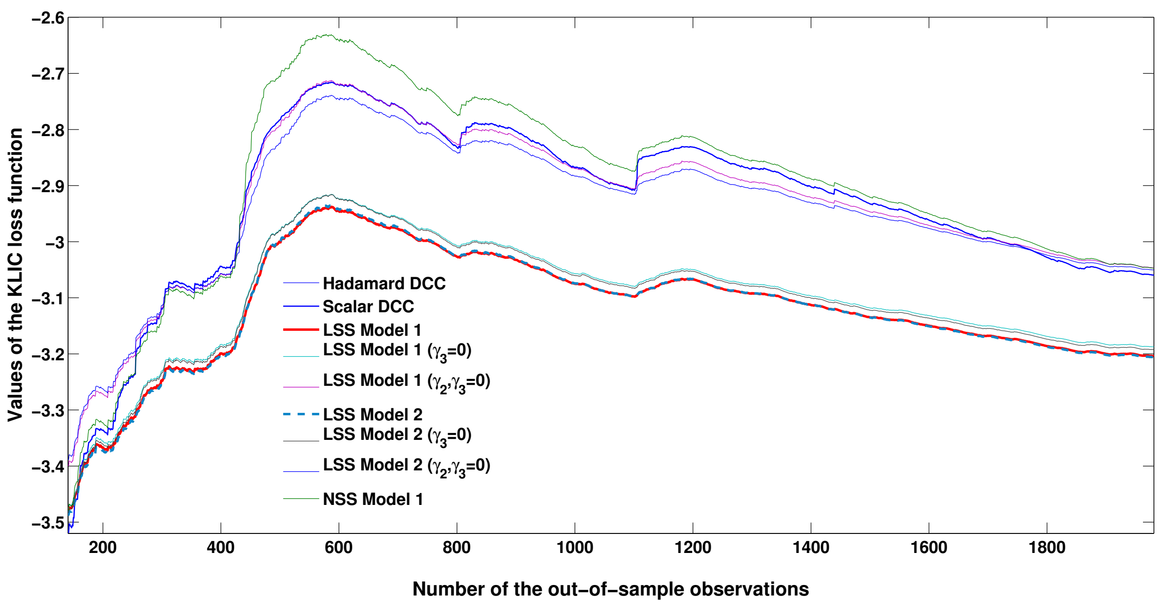

Figure 12: The values of the KLIC loss function $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_2}}$ in (4.5) for the competing models in terms of the out-of-sample length.

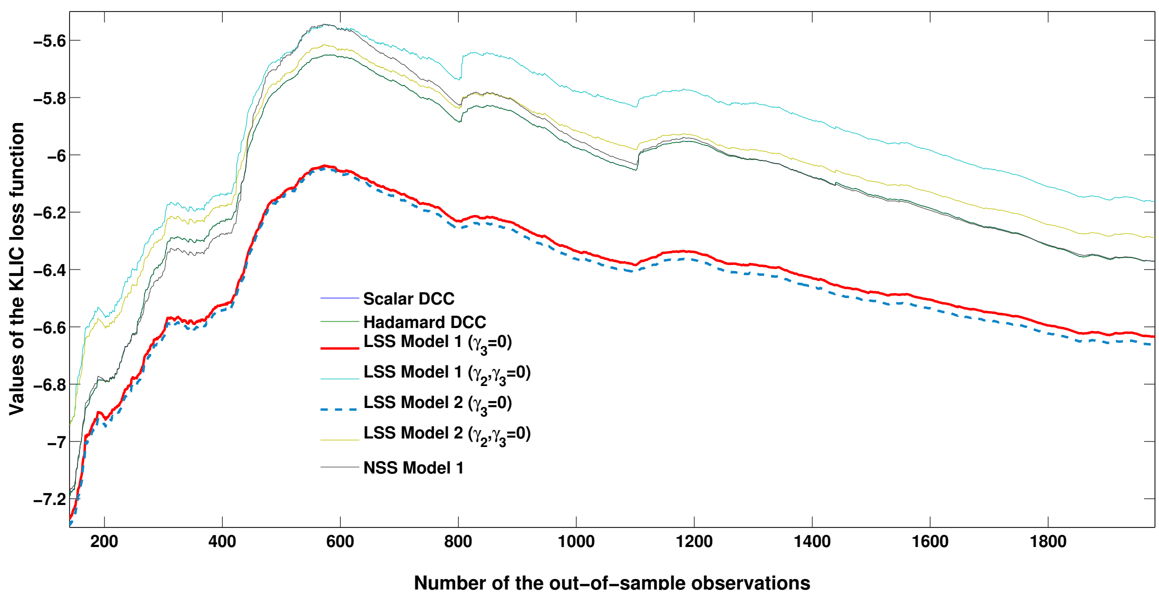

Figure 13: The values of the KLIC loss function $d^{\text{KLIC}}_{\mathcal{F}^*_{t_1}}$ in (4.6) for the competing models in terms of the out-of-sample length.

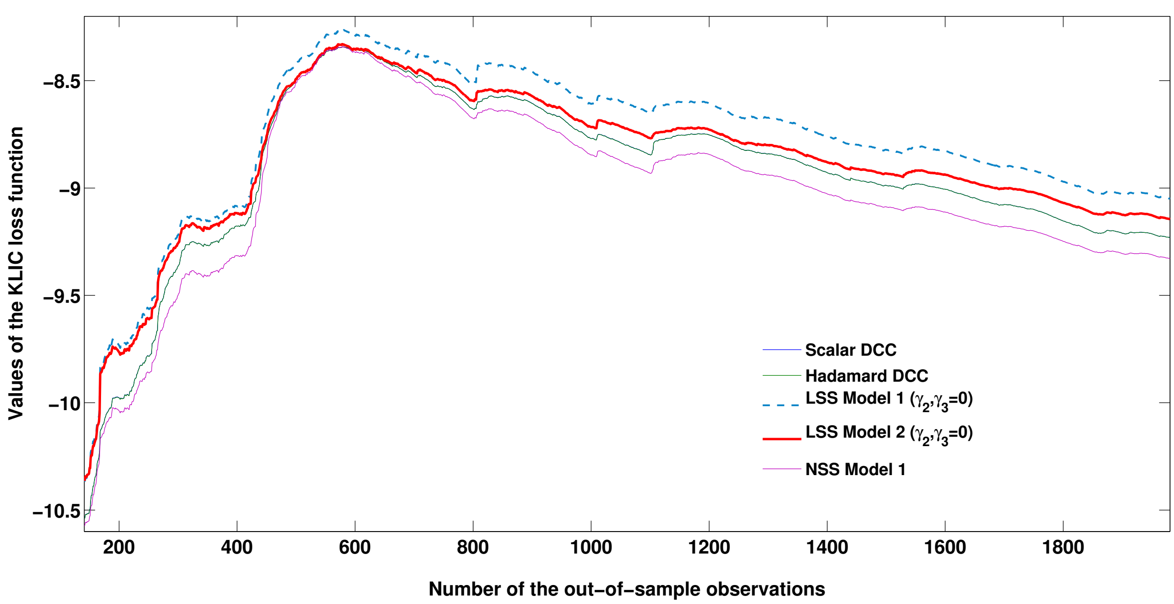

Figure 14: The values of the KLIC loss function $d^{\text{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$ in (4.7) for the competing models in terms of the out-of-sample length.